

QUALITY & TECHNOLOGY
FACULTY OF LIFE SCIENCES
UNIVERSITY OF COPENHAGEN



*Molecular Modeling and NMR
Studies of Starch Structural Motifs*

Hydration, Phosphorylation, and Nano-crystallites

Molecular Modeling and NMR Studies of Starch Structural Motifs - Hydration, Phosphorylation and Nano-crystallites

PhD thesis by
Peter Ibsen Hansen, 2008

Supervisors:

Professor Søren Balling Engelsen
Quality and Technology Department of Food Science
Plant Food Science and Spectroscopy and Chemometrics
and

Associate Professor Andreas Blennow
Molecular Plant Biology Department
of Plant Biology and Biotechnology

Faculty of Life Sciences -
University of Copenhagen

Title: Molecular Modeling and NMR Studies of Starch Structural Motifs -
Hydration, Phosphorylation and Nano-crystallites
2008 © Peter Ibsen Hansen
ISBN: 978-87-7611-241-7
Printed by Samfundslitteratur Grafik, Frederiksberg C, Denmark

Preface

This PhD thesis represents the conclusion of my efforts to explore the basic structural units of starch by means of combining molecular modeling methods and NMR spectroscopy. The study was funded through a full PhD scholarship from what was then named The Royal Veterinary and Agricultural University, and now is part of Copenhagen University.

The work presented here has mostly been conducted at the Faculty of Life Sciences, Copenhagen University, in a collaboration between the Quality & Technology group (Q&T), Department of Food science, and the Plant Biochemistry Laboratory, Department of Plant Biology. The project has been supervised by Professor Søren Balling Engelsen and Associate Professor Andreas Blennow, whom I would like to thank for the opportunity to enter the world of carbohydrate research. Their commitment to my work has been invaluable.

I would like to thank Serge Pérez, Jean-Luc Putaux, Andreas Fuchs, Christoffer Oberg, and many other people who were in Grenoble at that time for making my stay at CERMAV memorable.

I am also thankful to all of the people in the Q&T group who made the department a joyful place to be. Special mentions goes to Flemming for many good discussions on NMR, Thomas and Karin (the debate club), and my office-mates over the years, Franz, Nanna and specially Hanne who has become a good friend.

To my family, many friends, and in particular Marie. Thank you for all your patience and help in the past years. I will not forget that.

Peter I. Hansen
May, 2008

Summary

Starch is one of the most ubiquitous bio-polymers in nature, and is of indisputable importance for the life of both plants and all animals. This molecule has also become a widely used food ingredient, because of its versatile functional properties as a texturizer and its dietary benefits. The arrival of modern biochemical methods to genetically alter plants to produce starch with designed structural features has created a renewed interest in obtaining a better understanding of the governing principles which determines the relationship between the structural elements in the carbohydrate polymer, at the molecular level, and its macroscopic physical properties that we wish to control.

In this PhD project, models of the most important structural motifs of starch are explored by a combined approach using NMR spectroscopy and molecular dynamics simulations to measure and describe the effects of hydration and phosphorylation on α -glucan fragments.

NMR spectroscopy is today the primary experimental method for measuring structure at atomic resolution of molecules in the solution state. Combining this technique with molecular dynamics modeling, give an opportunity to extract both the dynamic and structural information at the atomic level on a femtosecond timescale.

A very basic version of molecular modeling was used with success to explain structural features of starch crystallites observed by TEM microscopy (Paper III), by building three dimensional computer models of the possible arrangements of the double helical amylopectin segments that constitutes the elements of the crystal. This study gave rise to a new principle for building the A-type crystalline starch domain, and added support for the traditional structure of B-type starch.

A series of trisaccharide models, that spans all the possible glucosidic linkage combinations in starch was examined in detail by molecular dynamics simulations in a second generation carbohydrate force field including explicit water molecules for solvation (Paper II). This way the preferred conformational space of the saccharides was mapped and number of water

binding sites involved in stabilizing the carbohydrate structures was identified using analysis of two dimensional pair distributions. This gave a full characterization of the structural elements in starch that can be described by trisaccharides, showing some combinations of glucosidic linkages result in stable conformations, mainly those involved in the amylopectin branch point.

Using a similar approach (Paper IV), the possible α -glucan phosphorylation structures was modeled, and the agreement between these chemical modifications and the accepted amylopectin double helix model was evaluated. This study indicated that of the two known phosphorylation sites in natural starch, only the 6-phosphorylation fits into the requirements of the double helical structure, while the 3-phosphorylation put to much structural strain on the $\alpha(1\rightarrow4)$ glucosidic linkage.

The results from both molecular dynamics simulation studies, was validated using homo- and heteronuclear scalar coupling constants and NOEs obtained by NMR spectroscopic measurements, which included a full assignment of all proton and carbon resonances in the examined carbohydrates.

As a minor side project on NMR method development a couple of designed series of mixtures containing solutes known to have overlapping resonances was measured using the DOSY experiment. The spectra was shown to be decomposable by chemometric analysis using the PARAFAC method (Paper VI), and the methods was demonstrated as a powerful technique for quantitatively separating lipoprotein fractions human blood plasma (Paper V).

The combined work presented in this thesis demonstrates the successful application of how detailed analysis of molecular dynamics data, supported by spectroscopic methods, can contribute to an improved understanding of α -glucan polymer structure and functionality.

Sammendrag

Stivelse er en af de hyppigst naturligt forekommende bio-polymerer, og spiller en afgørende rolle for existensen af alle planter og dyr. Stivelsesmolekylet er desuden blevet en udbredt ingrediens i fødevarer på grund af dets forskellige funktionelle egenskaber, heriblandt evnen til at kontrollere konsistens og visse ernæringsmæssige kvaliteter. Udviklingen af biokemiske metoder gør det muligt at ændre planter genetisk, således at de kan producere stivelse med en forudbestemt strukturel sammensætning. Dette har fornyet interessen for at opnå en forbedret indsigt i de styrende principer bag sammenhængen mellem strukturelle elementer i kulhydratstrukturen på molekylært niveau, og de makroskopiske fysiske egenskaber ved stivelse som vi ønsker at kontrollere.

Formålet med dette PhD projekt var at undersøge modeller af de vigtigste strukturelle motiver i stivelsesstrukturen, ved brug af NMR spektroskopi og simulationer af molekylær dynamik til at måle og beskrive hvilken effekt hydrering og phosphorylering har på fragmenter af α -glukaner.

NMR spektroskopi er i dag den primære eksperimentelle metode til opklaring af molekylære strukturer i væskefasen. Ved kombination af denne teknik og modellering af molekylær dynamik, har man en unik mulighed for at kunne opnå både dynamisk og strukturel information på atomart niveau samt på en femtosekund tidsskala.

I en del af projektet blev en simplificeret form for molekylær modellering succesfuldt anvendt til at forklare strukturelle egenskaber af stivelseskrystaller, som var blevet observeret ved hjælp af TEM mikroskopi (Paper II). Dette blev gjort ved at bygge tredimensionale computermodeller af en række mulige måder at ordne de amylopektin dobbelthelices, som udgør elementerne af denne type krystaller. Resultaterne af denne metode var et nyt princip for opbygningen af A-type krystallinsk stivelse samtidig med at den traditionelle struktur af B-type stivelse.

En række trisaccharidmodeller blev designet til at beskrive alle de mulige kombinationer af glukosidiske bindinger, som kan forekomme i stivelse.

Disse blev undersøgt detaljeret ved hjælp af molekylær dynamik simulationer i et andengenerations kulhydrat kraftfelt inklusiv explicitte vandmolekyler som solvent (Paper III).

På denne måde blev de foretrukne konformationer kortlagt, og et antal vandbindingspositioner, som er involveret i stabilisering af kulhydratstrukturen, blev identificeret ud fra analyse af todimensionale pardistributioner. Dette gav en fuld beskrivelse af alle de strukturelle elementer i stivelse som består af trisaccharider, og viste at især dem som er involveret i forgreningspunktet viser sig at antage stabile konformationer.

På tilsvarende vis (Paper IV), blev de mulige phosphorylede α -glukan strukturer modelleret. Overensstemmelsen mellem disse kemiske modifikationer og den accepterede model af amylopektin dobbelthelixen blev evalueret. Resultaterne af dette studie tyder på at af de to kendte phosphoryleringspositioner i naturlig stivelse er det kun 6-phosphoryleringen som passer til dobbelthelix strukturen. Derimod ses det at 3-phosphoryleringen inducere for store strukturelle ændringer omkring $\alpha(1\rightarrow4)$ bindingen.

Resultaterne fra begge de studier hvor der blev gjort brug af simulationer af molekylær dynamik blev valideret ved hjælp af homo- og heteronukleare skalare koblingskonstanter målt med NMR spektroskopi, efter forudgående tilordning af samtlige proton og carbon resonanser i de undersøgte kulhydrater.

Som et mindre sideprojekt i NMR metodeudvikling blev der designet flere rækker af blandinger indeholdende solutter, som vides at have overlappende resonanser. Disse blev målt ved brug af DOSY eksperimentet. Det blev vist at de resulterende spektre kunne adskilles med den kemometriske multivejs metode PARAFAC (Paper VI), og denne fremgangsmåde blev også vist at være et godt redskab til kvantitativt at adskille lipoprotein fraktioner i human blodplasma (Paper V).

Det samlede arbejde, som er præsenteret i denne afhandling viser hvorledes anvendelsen af simulationer af molekylær dynamik, understøttet af spektroskopiske metoder, kan bidrage til en forbedret forståelse af α -glukanpolymerers struktur og funktion.

Table of contents

	Preface	i
	Summary	ii
	Sammendrag	iv
	List of publications	ix
	Abbreviations	xi
	Notation	xii
1	Introduction	1
2	Microscopic starch structure	5
2.1	Starch, The Bigger Picture	5
2.2	α -D-glucopyranose	6
2.3	Linkage structure of the starch polymer	7
2.4	The starch granule	9
2.5	Starch crystallinity	10
2.5.1	Crystalline domains	12
3	Modeling carbohydrate dynamics	17
3.1	Molecular dynamics	17
3.1.1	Force fields	18
3.1.2	Dynamics	19
3.1.2.1	Physical conditions	21
3.1.2.2	Ewald summation	21
3.1.3	Water models	22
3.1.4	Adiabatic mapping	22
3.1.5	Fast generation of conformational maps	24
3.2	Trajectory Analysis	25
3.2.1	Diffusional properties	26
3.3	Hydration	27
3.3.1	Radial pair distributions	28
3.3.2	Finding hydrated structures	30
4	Measurement of carbohydrate structure	33
4.1	Single crystal structures	33
4.2	NMR spectroscopy	37
4.2.1	Assignment of NMR spectra	38
4.2.2	Measurement of hetero-nuclear scalar couplings	41

4.2.2.1	The Tvaroska parametrization	42
4.2.3	Homo-nuclear scalar couplings	44
4.2.3.1	The Haasnoot-Altona parametrization	44
4.2.4	Nuclear Overhauser Effects	45
4.2.5	Diffusion edited spectroscopy	47
5	Starch building blocks	51
5.1	The five starch trisaccharides	52
5.2	Trajectory analysis	54
5.2.1	Conformational distribution	54
5.2.2	Molecular flexibility	58
5.2.3	Structural water molecules	59
5.3	Experimental results	62
6	Starch phosphorylation	67
6.1	Phosphorylation models	69
6.2	Molecular dynamics	70
6.2.1	Force-field parameters	71
6.3	Conformational analysis	72
6.4	Hydrational analysis	75
6.5	Experimental data	76
7	Thesis out-look and perspectives	79
7.1	Experimental	80
7.1.1	Crystal structures	80
7.1.2	NMR structures	80
7.2	Computational	81
7.2.1	Karplus parametrization	82
7.2.2	Force field parametrization	82
8	Conclusions	85
9	References	87

PAPERS I-V

List of publications

Paper I

P. I. Hansen and S. B. Engelsen, *Modeling, Simulation and Visualization of the Hydration of Starch Glucans*. Starch - Biology, Structure and Functionality, Renewable Resources, Wiley and Sons Ltd. 2008. Editors: Anton Huber, Werner Praznik, ISBN-10: 0470014628, ISBN-13: 978-047001462.

In press

Paper II

P. I. Hansen, F. H. Larsen, S. M. Motawia, A. Blennow, M. Spraul, P. Dvortsak and S. B. Engelsen, *Structure and hydration of the amylopectin trisaccharide building blocks – Synthesis, NMR and molecular dynamics*.

ChemBioChem, **Submitted**

Paper III

P. I. Hansen, J-L. Putaux, A. Blennow and S. B. Engelsen, *Nano-crystal platelets of A- and B-type starches – A preliminary study*. Selected Investigations in the Field of Starch Science, Nova Science Publishers, Inc. 2008. Editors: Vladimir Yuryev, Piotr Tomasik, Andreas Blennow, Lyubov Wasserman, G.E. Zaikov

In press

Paper IV

P. I. Hansen, M. Spraul, P. Dvortsak, F. H. Larsen, A. Blennow, M. S. Motawia and S. B. Engelsen, *Starch phosphorylation - maltosidic restrains upon 3'- and 6'-phosphorylation investigated by chemical synthesis, molecular dynamics modeling and NMR spectroscopy*.

Journal of the American Chemical Society, **Submitted**

Paper V

R. Bro, N. Viereck, M. Toft, H. Toft, P. I. Hansen and S. B. Engelsen, *Solving the nuclear magnetic cocktail party effect in systems biology* Angewandte Chemie International Edition, **Submitted**

Paper VI

R. Bro, P.I. Hansen, N. Viereck, M. Dyrby, H.T. Pedersen, and S.B. Engelsen, *A new principle for unique spectral decomposition of 2D NMR data*, 195-203 The RSC proceeding for the 7th International Conference on Application of Magnetic Resonance in Food Science. 2004

Additional publication

N. Viereck, P.I. Hansen, S.B. Engelsen, *Undersøgelser af funktionelle faktorer i mejeriprodukter ved hjælp af NMR spektroskopi og kemometri*, Mælkeritidende. -rg. 116, nr. 25/26 (2003). - S. 616-620
(*Investigations of functional factors in dairy products by NMR spectroscopy and chemometrics*, Danish dairy journal)

Abbreviations

κ	J-HMBC scaling factor
\underline{a}	vector a
K	Temperature in kelvin
s	seconds
IUPAC	International Union of Pure and Applied Chemistry
cm^{-1}	Wavenumber
fs	femtosecond 10^{-15} seconds
ps	picosecond 10^{-12} seconds
ns	nanosecond 10^{-9} seconds
DFT	Density Functional Theory
DOSY	Diffusion Ordered SpectroscopY
HMBC	Heteronuclear Multiple Bond Correlation
HSQC	Heteronuclear Single Quantum Correlation
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
TEM	Transmission Electron Microscopy
TOCSY	TOtal Correlation SpectroscopY
SAXS	Small Angle X-ray Scattering
SEM	Scanning Electron Microscopy
Å	10^{-10} meter
η	viscosity
ϕ	dihedral angle 'phi'
ψ	dihedral angle 'psi'
ω	dihedral angle 'omega'
$J(\theta)$	scalar coupling as a function of θ
${}^nJ_{XY}$	n -bond X-Y scalar coupling
ρ_{water}	density of water
K_a	acid dissociation constant
pK_a	$-\log_{10} K_a$
y_a	fraction of acid
χ_n	Exocyclic dihedral n
ϵ_0	Vacuum permittivity
ϵ_r	relative dielectric coefficient
r_{ij}	distance from atom i to atom j
$f^{(n)}(x)$	n 'th derivative of the function f in x .
Δt	$t_2 - t_1$

Notation

The names ‘iso-panose’, ‘ α -panose’ (or just panose), ‘maltotriose’, ‘iso-maltotriose’, and ‘ α -forkose’ (or forkose) will be used as short hand for the trisaccharides described in Paper II and Chapter 5, in stead of ‘methyl 6-O-(α -maltosyl)- α -D-glucopyranoside’, ‘methyl 6’- α -D-glucopyranosyl- α -maltoside’, ‘methyl α -D-glucopyranosyl-(1 \rightarrow 4)- α -D-glucopyranosyl-(1 \rightarrow 4)- α -D-glucopyranoside’, ‘methyl α -D-glucopyranosyl-(1 \rightarrow 6)- α -D-glucopyranosyl-(1 \rightarrow 6)- α -D-glucopyranoside’, ‘methyl 4,6-di-O-(α -D-glucopyranosyl)- α -D-glucopyranoside’ respectively.

For the two phosphorylated maltoses studied in Paper IV and Chapter 6, ‘maltose-3’-O-phosphate’ and ‘maltose-6’-O-phosphate’ the names used in this text will be maltose-3-phosphate and maltose-6-phosphate or simply ‘the 3-phosphorylated maltose’ or equivalent.

Footnotes will be marked by superscript numbers like this⁹⁹⁹.

References to literature is given by numbers in square brackets like this[999].

1 Introduction

Carbohydrates are a major food component with a wide variety of nutritional and functional properties. The recent progress of biotechnology, in particular genetics, has opened up a whole new area of interest by giving scientists the tools to modify, or even design specific genes in plants, which are then able to synthesize starch with desirable properties. If we are going to exploit these new opportunities in an optimal way, it calls for a better understanding of the basic structure-function relationships of the starch polymer [1].

It is believed that most functional properties of carbohydrates, is dependent on their aqueous solvation. However, the detailed mechanisms behind this intermolecular relationship is still not very well described, due to the lack of fast time resolution of experimental techniques.

Historically, carbohydrate structure has mostly been studied by use of X-ray diffraction techniques, which do give high resolution structural information, but is only a static image of the molecular system, and do not provide any information on the dynamics interaction between the solute and solvent.

NMR spectroscopy is the only experimental method that is able to give structural information at the atomic level and on the dynamics of the system at the same time, which makes it the method of choice for studying biologically important molecules in solution. However, many of the intermolecular mechanisms that we would like to study happens on a much faster time scale than it is possible to monitor by NMR.

Therefore, the experimental techniques must be combined with theoretical methods, such as molecular dynamics simulation, that give infinite resolution on both the atomic scale and in time. Thus, by validating molecular dynamics simulations with spectroscopic measurements, the interactions with water can be described in more detail than by other approaches [2]. Only by further developing molecular modeling methods in the area of carbohydrate research, the understanding of carbohydrate hydration can be fully elucidated.

This Ph.D thesis is a compilation of the efforts to establish a systematic approach to investigate the structural properties of starch by thoroughly

examination of the basic α -glucan building blocks and their interaction with the surrounding water molecules.

The starch elements investigated in this thesis is, a set of phosphorylated di-saccharides functioning as models for the starch phosphorylations known to exist in vivo (Chapter 6), and a set of tri-saccharides that comprise all possible combinations of linkages in starch (Chapter 5). With both sets of model compounds, the methodology and exploration of the possibilities in combining MD simulation data with multi dimensional NMR techniques will be similar.

The present thesis is based on 6 manuscripts that has been submitted for publication, in either book-form or as peer-reviewed articles. Additional work was done providing the basis for these results, and some of those bits and pieces will now serve as some of the glue binding together the parts of this thesis.

Chapter 2, is meant to serve as a basic introduction to the current knowledge on starch structure, and the basic chemical structures that are needed to be able to discuss the subjects in following chapters. In addition, the results of the macroscopic modeling in Paper III is presented.

Chapter 3, is about the concepts of molecular dynamics simulations of carbohydrates, which is also the topic of Paper I, and will outline why semi empirical approaches is a useful tool for conformational analysis, and the study of hydrational features.

In **Chapter 4**, an overview is given of the possible experimental techniques available for studying carbohydrate structure, and thereby qualify the choice of methods used in Paper II, Paper IV, and Paper V.

In **Chapter 5**, the design and analysis of the five tri-saccharides of Paper II that constitutes the basis for building any starch polymer is described, including a full description of all covalent configurations in amylose and amylopectin.

Chapter 6 introduces the concept of phosphorylation of the starch polymer that is explored in Paper IV. An analysis of the structure and hydration of

two phosphorylated maltose's as models of the possible substitution sites is presented and their influence on the conformation of the saccharides.

Finally, in **Chapter 7** and **Chapter 8** I have collected my thoughts on what can be done to further explore the area of research on carbohydrate structure, and the conclusions that can be drawn from this PhD. study is summarized.

2 Microscopic starch structure

In this chapter a brief introduction to the structure of starch will be given. Starting from glucose - the smallest repetitive unit of the starch molecule - and the way it is linked in the polymer, the microscopic starch granule will be presented. Moving towards a smaller scale, the most important structural elements that occur in starch is introduced, leading to the structure of nanometer scale amylopectin crystal platelets, which is the result of Paper III.

2.1 Starch, The Bigger Picture

In the green plant seeds and tubers, starch granules function as the reserve for long time storage of energy. Starch granules which are densely packed and not soluble in water, is however still available to the plants metabolism when needed. Besides the functionality provided by starch in the plant, it has become one of the main constituents in the human diet. In addition to being a source of energy, starch has also become a widely used ingredient in various foods because of its high diversity in physical properties which are used to control parameters in a product such as the texture, its ability to absorb moisture and general consistency controlled by thickening and gelling [3]. Other benefits of starch is that it is a biodegradable, renewable resource that is inexpensive to produce.

As an example, high amylose starches has been produced by special milling methods [4]. Its improved gelling ability has been used to replace gelatin in food products, and the high amount of crystalline matter - which is not easily available to digestion enzymes has been shown to have properties in common with dietary fibers [5–6].

The functional properties of starch is in most cases closely related to the hydration of the molecule [1]. It is therefore imperative to understand the carbohydrate-water interaction mechanisms at the atomic level, to be able to make suggestions for the design of improved or new structures with desired properties.

The starch ‘molecule’ is a polymer primarily constructed using a single unique unit, the glucose ring, which will have to be introduced in more

detail to provide a basis for the following discussion of some important structural elements in the microscopic domain.

2.2 α -D-glucopyranose

The fundamental building block of starch is the α -D-glucopyranose molecule, which is often just referred to by the shorter trivial name α -glucose and sometimes simply *glucose*.¹ In other words, starch is a homo-polymer of α -glucose. Glucose exists in two diastereomers called ‘anomers’. In aqueous solution glucose reversibly changes between the α and β forms shown to yield a 36:64 mixture of the anomers at 25° [7]. The more stable anomer, β -glucose, is the single monomer of another ubiquitous glucose polymer; cellulose. Figure 2.1 illustrates the chemical structure of α - and β -glucose, and the methylated equivalent of the α form, the latter is often used in synthetic model carbohydrates. The chemical composition of the sugar rings is commonly represented by a Haworth projection, giving a three-dimensional view of the ring-structure. Differing from normal conventions, hydrogen’s are explicitly depicted to increase clarity.

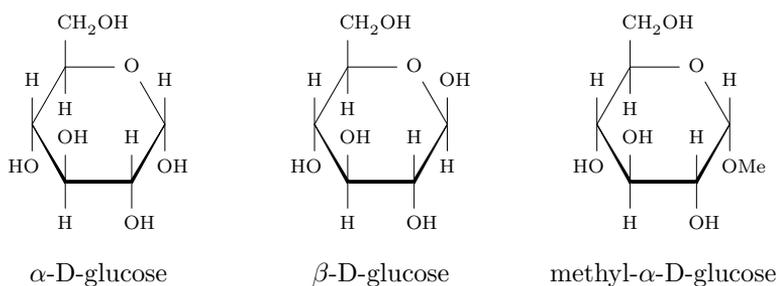


Figure 2.1 Haworth projections of α -glucose, β -glucose, and the methyl- α -D-glucose that is used in model compounds to block the mutarotation between the two former structures.

When referring to individual atoms in the saccharides, the appropriate IUPAC numbering system[8] will be used, as defined by **Figure 2.2**. This notation is similar to the one used in the definition of most molecular

¹ Technically, glucose is the correct name for the open chain aldehyde form, but it has become a common term also used for the cyclic hemiacetal.

dynamics force field parameter files, and will be used in Chapter 5 and Chapter 6.

The α -glucose ring structure is known to be most stable in the 4C_1 chair conformation [9–10] as shown in **Figure 2.2**, and this conformation will be implied in rest of this thesis.

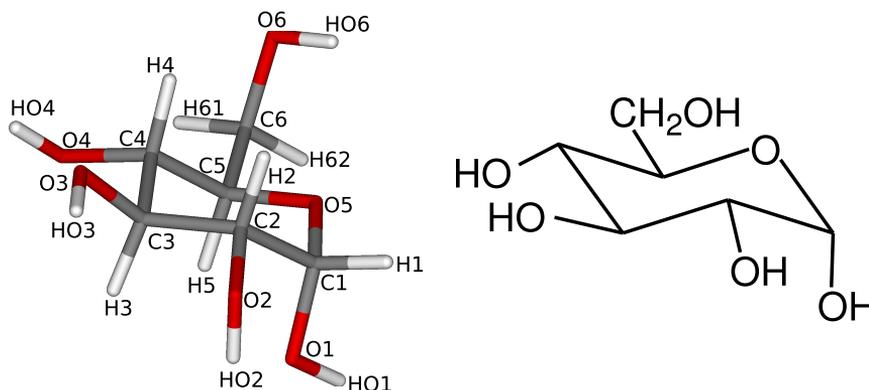


Figure 2.2 Chair conformation of the 6-membered ring-structure, and labeling scheme for specifying all individual atoms in the glucose ring. Notice that 'HO1' only designates a single atom, namely the proton bound to 'O1'.

2.3 Linkage structure of the starch polymer

The structure of the starch polymer is divided into two groups of polymers based on occurrence of the two only types of linkages that are found to connect the glucose units. The first structural principle, amylose is a linear polymer constructed using the $\alpha(1\rightarrow4)$ linkage as illustrated by the α -maltose structure in **Figure 2.3**. Here, it is also shown how the conformation around such a bond can be described by two dihedral angles Φ and Ψ . In amylose, the occurrence of branching is less than 1% making it almost completely linear. When glucose units are linked, the numbering scheme of **Figure 2.2** still applies, but atoms in the unit linked to the non-reducing end will be denoted with a prime, and the next with double prime².

² This notation is reverse of what is standard in organic chemistry, but is used to conform with the nomenclature of the literature on modeling of disaccharides.

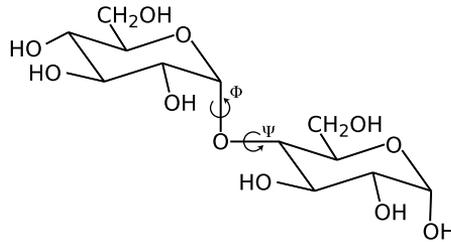


Figure 2.3 The $\alpha(1\rightarrow4)$ glycosidic linkage illustrated by α -maltose, including definitions of the dihedral angles that determine the overall structure of the molecule.

The second molecular principle of the starch structure, is amylopectin that is characterized by its basic $\alpha(1\rightarrow4)$ structure being branched by additional $\alpha(1\rightarrow6)$ bonds. In the amylopectin parts the branching bonds make up around 5-6% of the total [11]. The $\alpha(1\rightarrow6)$ bond is exemplified by the *iso*-maltose molecule in **Figure 2.4** and it is seen that a third dihedral angle, ω , is needed to describe the geometry. The definitions of the three glycosidic linkage dihedrals that will be used throughout this thesis are: Φ is the angle between the vector defined by the O5-C1 bond atoms and the one defined by the O1-C4' bond. Ψ is similarly defined by the angle between the vectors C1-O1 and C4'-C5'. ω is defined by O1-C6' and C5'-O5'.

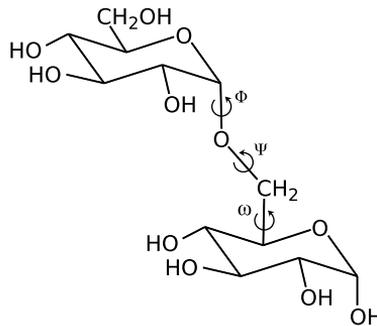


Figure 2.4 The *iso*-maltose molecule used as an example of the $\alpha(1\rightarrow6)$ bond. The geometry of the linkage is given by the three dihedrals Φ , Ψ , and ω .

Linkage structure of the starch polymer

As will be described later in **Section 4.1** the inherent structure of the $\alpha(1\rightarrow4)$ bond is a turn that in a polymer is predisposed to induce helical structures that tend to pack into crystalline layers between the amorphous regions systematically created by the $\alpha(1\rightarrow6)$ branching [12].

2.4 The starch granule

The starch granules can have different shapes, often spherical or elliptical, in which the starch molecules are densely packed as a semi crystalline solid. **Figure 2.5** is an example of a scanning electron microscopy image of a potato starch sample, which is useful to give a general impression of the overall size and shape of the granules, as well as the rather large variation of these from a single plant source. For detailed information on the biosynthesis of starch see the review from Buléon et al. [11].

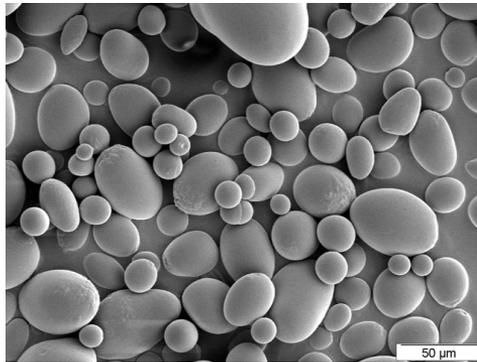


Figure 2.5 Scanning electron microscopy image of potato starch granules. (Photo: Jean-Luc Putaux).

The granule size has a wide distribution both between and within species and is in the range of 0.1-100 μm [13]. For example, in potato starch, the variation of granules has been observed from approximately 5 μm to 100 μm , and the functionality of the starch has been shown to depend on the size of the granules [14], partly because of the easier access of digestion enzymes on smaller surfaces, but also because of size dependent differences with the content of organic phosphate which will be further explored in Chapter 6.

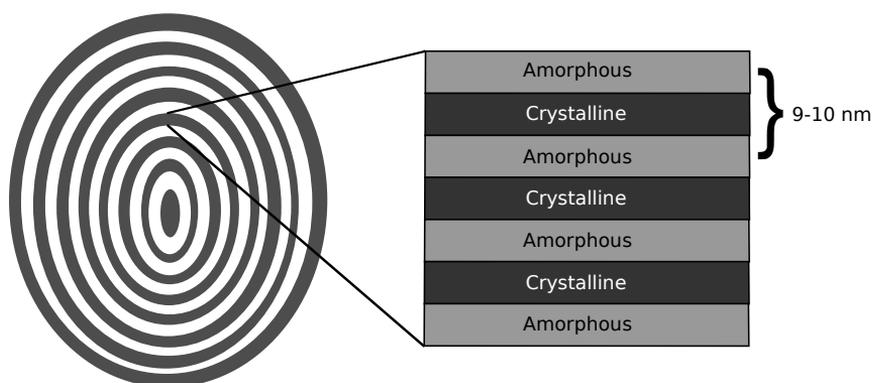


Figure 2.6 Conceptual drawing of growth rings in the elliptical starch granule, and how they are subdivided into alternating amorphous and crystalline layers. Adapted from Jenkins et al. [15]

Looking closer into the starch granule, reveals that it is sub-organized as layered structures, as illustrated in **Figure 2.6**, in analogy to what is often observed in plants on a macroscopic level. Examples of this are the layers in onions and the growth rings in trees. In starch the thickness of these layers has been found to be in the 100-400 nm range [13], and they are again built by alternating amorphous and crystalline organizations of the polymer, as in **Figure 2.7**, where a repeat layer has been measured to be approximately 9-10 nm in thickness [16–19]. A super-helical model for this type of alternating layer structure has been proposed, originating from Oostergetel et al. [20–22] and recently a modified structure of this has been suggested by Bertoft [23]. However, the platelet model described here will suffice for the topics covered by this text.

The organized part of these repeats is double helical amylopectin [20] crystallized into either an A- or B-type crystal structure, depending on the plant origin. The detailed geometry of the different starch types on starch crystals, which is essential to the subject of Paper III, will be further introduced in **Section 2.5**.

2.5 Starch crystallinity

Early on, powder diffraction studies of starch granules clearly showed that it contains some parts which are organized in a crystalline fashion, and that

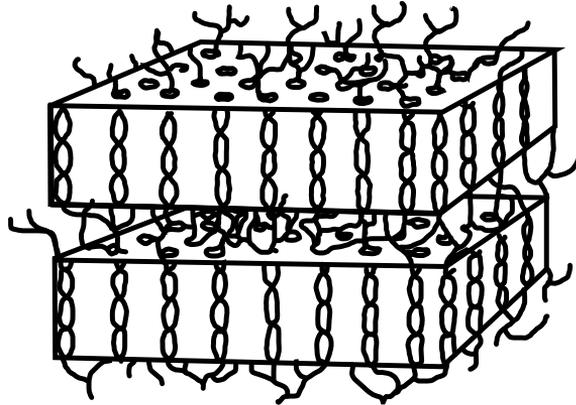


Figure 2.7 Alternating crystalline and amorphous layers. Adapted from Gallant et al. [13]

starches of different origin can be categorized into groups with different polymorphism. Traditionally, starch is categorized in three types: A-, B-, and C-type. Starch from a specific plant origin is most often found as only one of these types. The diffraction patterns obtained with granules from cereal and grain starches has been classified in numerous studies as A-type starch [24–26], the starch from tubers and roots as B-type starch [24–25, 27–30], and a rare third kind of granules from bananas, beans, and peas has been labeled as being C-type [24, 29, 31]. The C-type allomorph was later found, by Wu et al. [29] to be a mixture of A- and B-type starch. It must be emphasized that this classification is a simplification as exceptions are known, such as kudzu root starch, which is found in all three forms dependent on the geographical origin.

Detailed models for the A- and B-type crystal was proposed from refinement of fiber X-ray diffraction data by Sarko and Wu [24] and later improvements have been published by Imberty et al. [26, 30], which are the currently accepted structures.

The crystalline forms of starch, is described by two different schemes for packing amylose strands that are curled into double helical structures. A double helix conformation was first suggested by Kreger [27], and further refined by several groups [24, 32–33], until a final structure was demonstrated by Imberty et al. [26, 30], that showed that the helices are left

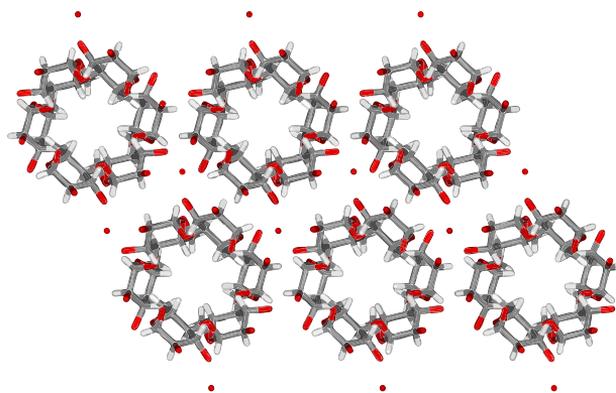


Figure 2.8 The monoclinic A-type starch crystal packing, using the coordinates published by Imberty et al. [26]. Water molecules are represented by a dot at the position of their oxygen atom

handed and parallel stranded with 12 glucose units in the unit cell of the crystal, and that the A- and B-helices are nearly the same in molecular conformation. The A-type starch is crystallized into a monoclinic lattice³ that contains only four water molecules. The B-type structure is packed into a unit cell with a geometry that resembles the ‘honeycomb’ pattern, which is not a real lattice, but corresponds to a hexagonal lattice where all the centers of the hexagons are removed. This gives a less dense structure than observed for the A-type, with room for 36 water molecules in the unit cell. **Figure 2.8** and **Figure 2.9** are illustrations of the A- and B-type structure respectively, both generated from the original coordinates.

2.5.1 Crystalline domains

The crystalline part of starch granule layers, such as the one illustrated in **Figure 2.7**, can be separated from the amorphous parts by using mild acid hydrolysis, with the remaining parts being the so called ‘lintners’. This procedure was investigated by Robin et al. [34], who also showed by X-ray diffraction that the hydrolysis of native starches follows two different kinetic mechanisms, a fast and a slow, and that the acid resistant fraction is mainly

³ And is almost hexagonal.

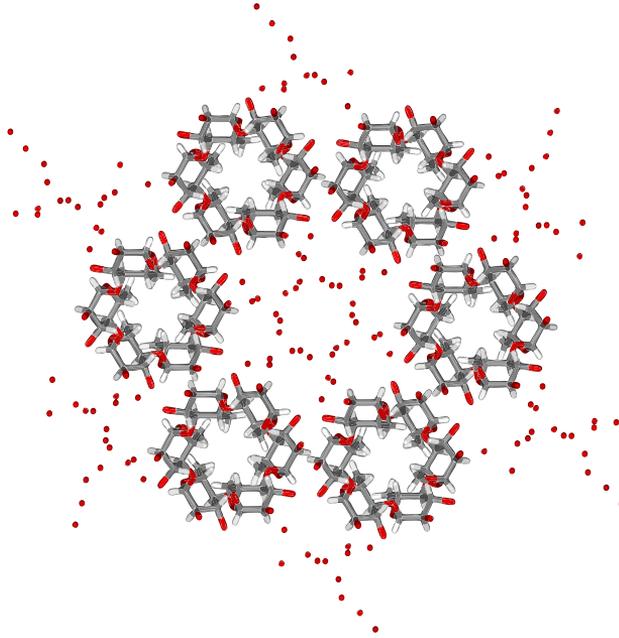


Figure 2.9 The honeycomb B-type starch pseudo-crystal packing, using the coordinates published by Imberty et al. [30]. Water molecules are represented by a dot at the position of their oxygen atom

crystalline. Lately these crystalline domains have been shown to work as a reinforcing agent when added to polystyrene material [35].

Recent activity in this field has been encouraged by progress in quality of electron microscopy technology that makes it possible to observe suspensions the platelet structures lying on a flat surface with a reasonably low degree of overlapping as in **Figure 2.10**. Images obtained in this manner, of hydrolyzed A-type starches, reveal a collection of parallelogram geometries that are closely related to the results from crystallographic studies [24]. Similar (unpublished) results have also been obtained for B-type starches. From the results of Putaux et al. [37] it is seen that the average acute angle of the A-type crystallites of waxy maize starch was approximately 60° . That is evidently larger than the 56.5° angle that was previously found from diffraction data [26]. These observations lead to the proposition that the crystalline buildup of the lintners might involve an inclination between

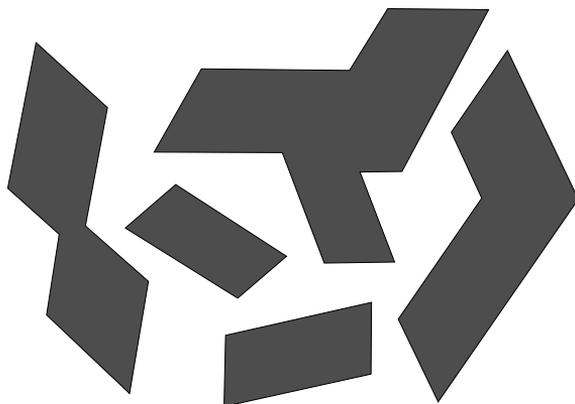


Figure 2.10 Conceptual drawing of 2D crystals, as they look on TEM images when suspended in a solvent and allowed to settle on a horizontal surface. Adapted from Pohn et al. [36]

helical units, which would alter the geometry of the crystal when observed perpendicular to the newly generated surface plane. Another source of inspiration for this model was the results from Pohn et al. [36] where *in vitro* amylose crystals were grown and lead to observations of ‘feather-like’ structures from a viewpoint perpendicular to the c -axis.

This novel idea is the topic of Paper III in which the idea was examined, by building molecular models using the suggested principles. Here the basic structural element was chosen to be a single amylose double helix based on the results of Imberty et al. [26, 30] for the A- and B-type crystallite. In this study the coordinates of the helix structure was propagated according to the dimensions of the pseudo-hexagonal packing required in the (a, b) -plane of the crystal repeating unit. In the original study by Pohn et al. [36], a flat structure was proposed with an up-down-up-down repeat which is in agreement with the experimental data that indicated a $c/2$ translation symmetry of the crystal. This translation also has an energetic importance since it matches wide and narrow parts of the adjacent helices in a way that allows for dense packing. In our *in silico* experiment a series of single layer crystallites was generated where in every new structure there is an additional increase in the inter-helical displacement along the helical axis. Since we know the complete coordinates of the structure the geometry can

easily be calculated as a function of either the displacement distance or alternatively as a function of the inclination angle.

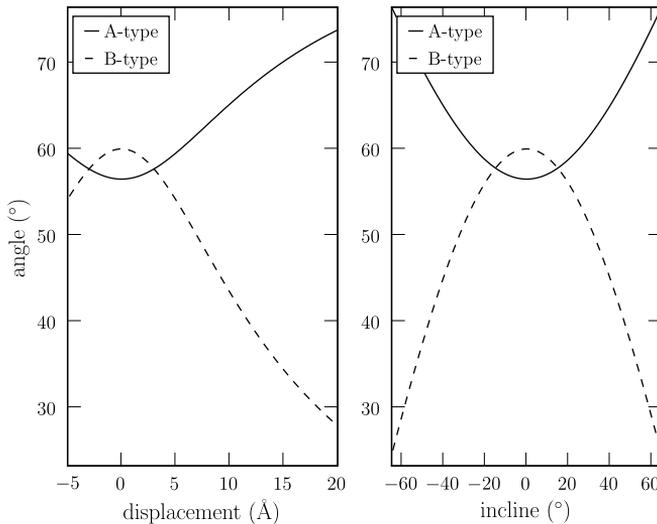


Figure 2.11 The acute angle of a constructed crystal platelet of both A- and B-type packing as a function of the inter-helical displacement distance (left), or the slope of the buildup with respect to a horizontal structure.

The results from the study is summarized by the plots in **figure 2.11**, which shows that for the A-type crystal to reproduce the experimental data, the requirement is a inter-helical displacement of 5.0 Å equivalent to an inclination of 26.7°. This result is in excellent agreement with the expected and energetically favored $c/2$ translation in the unit cell, which is 5.35 Å for A-type starch, and it supports the previously mentioned observations by Pohn et al.

For the B-type packing structure, all observations also indicate an acute angle in the platelets of about 60° but, contrary to the case of the A-type crystals, this angle is characteristic for the hexagonal-like honeycomb organization of this allomorph. Previous studies of B-type starches [20, 22], using SAXS, have also suggested tilted structures, similar to the A-type but this time with a 17.25° angle. In this case, the modeling data

does not agree with a tilt model, but supports the original ‘flat’ structure of Imberty and Pérez [30], since all displacements of the helices will produce structures with smaller acute angles.

3 Modeling carbohydrate dynamics

The term ‘molecular modeling’ is used in many different contexts, from the construction of a single three dimensional molecular structure like the ‘soft modeling’ of nanostructures - such as in Paper III - to molecular dynamics simulations - as used in Paper II and Paper IV - and quantum chemical calculations. In this chapter the basic principles of molecular dynamics in a classical force field will be considered, followed by a selection of the analytical methods, that can be used to validate the theoretical results, in subsequent chapters.

3.1 Molecular dynamics

The method of molecular dynamics simulations was first introduced by Alder et al. [38] to study a system of colliding hard spheres over the time scale of a few picoseconds. It was however soon developed to a point where molecules of biological relevance could also be simulated, as exemplified by an early study where the protein BPTI was simulated in vacuum for 9.2 ps [39]. By the beneficial development in computer technology large system can nowadays be simulated on a much longer timescale and, with supercomputers, getting close to the microsecond timescale that is more relevant biological processes [40]. With the increase in computational power comes the desire to simulate larger and more relevant systems. Specially important is also the awareness of the influence of solvent [41–42], on the conformation and thus the functionality of most molecules in aqueous systems. The inclusion of explicit solvent or condensed phase is a large addition to the complexity of a simulation system, and the major contributor to simulation time.

One of the appeals of molecular simulations is that they make it possible to choose the level of detail, opposed to the time-scale limitations of spectroscopic methods. Thus, if the simulated model can be shown to reproduce experimental results, fast mechanisms at the atomic level can be studied under the physical assumptions of the simulation framework.

The popularity of molecular dynamics simulations, and the many different need of researchers, has spawned multiple software programs with various

parametrizations of molecular systems. The most important programs in this area are CHARMM [43], the AMBER program suite [44–45], GROMOS/GROMACS [46–48], and TINKER [49]. The relevance of them all has been substantially demonstrated in the literature, and one just have to chose, based on the needs for special features. Molecular dynamics should be differentiated from Monte Carlo methods, since the latter does not provide information on the time dependence of events, which is crucial if the purpose is to understand processes. They can though, be used to reproduce averaged properties, as was demonstrated by Peters et al. [50]. In this study they were using a hard sphere exo-anomeric force field to produce ensemble averages for NMR parameters such as NOEs and scalar couplings, and found a good agreement with experimental observations.

3.1.1 Force fields

An empirical force field is the combination of a set of physical parameters and a potential energy function used to evaluate the energy of a system given the coordinates of all atoms and information on the covalent structure. A generalized empirical potential energy function used in many force fields has the form of **Equation 3.1** [51].

$$U(\underline{R}) = \sum_b \overbrace{K_b(b - b_0)^2}^{\text{bonding force}} + \sum_\theta \overbrace{K_\theta(\theta - \theta_0)^2}^{\text{angle force}} + \sum_\omega \overbrace{K_\omega(\omega - \omega_0)^2}^{\text{improper dihedrals}} \quad (3.1a)$$

$$+ \sum_S \overbrace{K_S(S - S_0)^2}^{\text{Urey-Bradley}} \quad (3.1b)$$

$$+ \sum_\phi \overbrace{A_n(1 + \cos(n\phi - \phi_0))}^{\text{torsions}} \quad (3.1c)$$

$$+ \sum_{i < j} \left(\overbrace{\epsilon_{ij} \left[\left(\frac{r_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{\min}}{r_{ij}} \right)^6 \right]}^{\text{Lennard-Jones}} + \overbrace{\frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_i r_j}}^{\text{electrostatic}} \right) \quad (3.1d)$$

The potential energy function is a sum where each term describes a type of molecular mechanics which can typically be measured using x-ray diffraction, vibrational spectroscopy, or be estimated from quantum mechanical calculations.

The first term in Equation 3.1a is a sum of the covalent bond stretching force energies in the system for all bonds b with an equilibrium bond length b_0 approximated by the Hooke's law harmonic potential function using a force constant K_b . Likewise, there are added harmonic descriptions of deformations of angles θ between bonds, with the force constant K_θ , improper dihedrals ω with a force constant K_ω , and in Equation 3.1b the Urey-Bradley interaction for 1,3 pairs [52], also with a force constant K_S . An angular dependence of the twisting deformation around bonds is added by the torsional term in Equation 3.1c, and the final sum in Equation 3.1d are the non-bonded distance dependent effects such as Van der Waals interactions using the Lennard-Jones 6-12 potential [53–54], and Coulombic interactions of charges.

3.1.2 Dynamics

Given a suitable set of parameters for the energy function in Equation 3.1, the classical equations of motion can be solved, for all atoms in the system at hand, by numerical methods to give the time evolution of atomic motion. For a given atom in the system with a known position, the new position after a time period Δt is by Taylor expansion⁴

$$\begin{aligned} x(t + \Delta t) &= \sum_{n=0}^{\infty} \frac{x^{(n)}(t)}{n!} (t + \Delta t - t)^n \\ &= x(t) + \frac{dx}{dt} \Delta t + \frac{1}{2} \frac{d^2x}{dt^2} \Delta t^2 + \dots \end{aligned} \quad (3.2)$$

To find the position of the atom at time $t + \Delta t$, the velocity and acceleration of the atom has to be calculated by numerical approximation. This can be done by integration of Newton's second law

⁴ For the one dimensional case.

$$\underline{F} = m\underline{a} = m \frac{d\underline{v}}{dt} = m \frac{d^2\underline{x}}{dt^2} \quad (3.3)$$

In this differential equation, \underline{F} is the force acting on the particle, m is the constant mass, \underline{v} is the velocity, and \underline{a} the acceleration. It is clear that both the acceleration and the first derivative of the velocity are the second and third term of the Taylor expansion in Equation 3.2, and that these can be obtained from Equation 3.3.

Several algorithms exist for performing integration of Equation 3.3 [55–57]. The most popular method today is the velocity version [58] of the Verlet method [55] in the form of the SHAKE [59] or RATTLE [60] algorithms, which are used to simplify calculations by constraining the stretching motion of covalent bonds to hydrogen atoms.

In principle, the ideal integration time step Δt would be infinitely small. This is - for obvious reasons - not practical. Thus, a compromise has to be made which allows for a reasonable reproducibility of experimental data at the same time as keeping the ‘real-time’ simulation time within practical limits. Among the motional phenomena that we could want to reproduce, stretching vibrations of covalent bond represent some of the fastest motions. As an example, the O-H stretching vibration at approximately 3400 cm^{-1} in the infrared spectrum has a period of around 10^{-14} s. Therefore the time step is often assigned an order of magnitude lower of this to $\Delta t = 1$ fs. Since the subsequent analysis does not always require this level of time resolution, the simulation data can be sampled at a lower frequency.

Numerous specialized force fields for simulation of carbohydrates have been published. The most important ones are the CHARMM based HGFB [61] and CSFF [62] force fields for a variety of sugars, the GLYCAM [63] extension for the AMBER force field [64–65], and the recent less known 45A4 GROMACS force field for hexopyranose based carbohydrates [66]. Some more in depth descriptions of force fields and their properties can be found in the reviews by Perez et al. [67], Mackerell [68], and Hemmingsen et al. [69].

3.1.2.1 Physical conditions

Molecular dynamics simulations can be carried out under many different conditions such as constant pressure, volume, or temperature. These constraints all require special treatments of the previously mentioned algorithms. Such methods were first described by Andersen [70], and further developed versions of these are now integrated parts of simulation softwares. For the simulations in Paper II and Paper IV, the systems were simulated as canonical ensembles only, keeping the temperature and the volume of the systems constant. From the chosen temperature, initial velocities are assigned to all particles so that the rate of stochastic collisions is adequately represented. During the following calculation the system is kept at the wanted temperature by coupling to an external ‘bath’ of a chosen constant temperature T_0 , which adjusts velocities at each time-step if needed [71] to λv by calculating the scaling factor as

$$\lambda = 1 + \frac{\Delta t}{2\tau_T} \frac{T_0}{T - 1} \quad (3.4)$$

Where τ_T is a characteristic relaxation time of the system, that can be chosen to regulate the strength of the coupling between the simulated system and the temperature bath. For all simulations in this thesis the temperature was kept at a constant temperature of 300 K, using temperature correction to readjust deviations larger than 1 K.

Since we are interested in the role of water as a participant in the structuring of carbohydrates, the solvent has to be included in the calculations. This is done by immersion of the solute into a quadratic box of water molecules. As a rule of thumb the box side length should be twice the extension of the solute. This lead to a box of 512 water molecules for the trisaccharides in Paper II, which was the reused for the smaller disaccharides in Paper IV. The solvation procedure cause some overlapping of molecules which is resolved by deletion of overlapping water molecules followed by a readjustment of this box size to approximate a density of 1.

3.1.2.2 Ewald summation

For the simulations of the charged versions of the phosphorylated malto-sides in Section 6.1, Ewald summation [72] was used to take long range

electrostatic interactions into account, by adding a charge-charge potential energy contribution term to Equation 3.1. This term is similar in form to the normal Coulombic interaction, but has on cutoff limit and thus takes interactions with an infinite number of periodic cells into account.

The details of the Ewald summation methods is beyond the scope of this thesis, but is described thoroughly by Allen et al. [73] and also lately reviewed by Fennel et al. [74] along with alternatives that could possibly become faster, and the development of these will therefore be interesting to explore in future work.

3.1.3 Water models

Numerous water models for molecular dynamics simulation purposes exist to describe water as an explicit solvent. The most prominent water models are: The SPC by Berendesen et al. [75] and TIP models by Jorgensen et al. [76], the ST2 model by Stillinger et al. [77], and the AMOEBA water model by Ren et al. [78]. A comparison of such models was published by Jorgensen et al. and van der Spoel et al. However the SPC and TIP3P [76, 79] are by far the most widely used. Recent reviews of water models and their ability to reproduce water structure in simulations has been published by Finney et al. [80] and Errington et al. [81]. Most molecular dynamics softwares are able to handle multiple water models, but the model should be chosen to complement the choice of force field. The selected force field for the simulations included in this thesis is the CSFF [62] parameter set, and the CHARMM based force fields are parametrized using the TIP3P [76, 82] as an integrated water model [51, 83] The use of this setup was also influenced by the preceding recommendations of Corzana et al. [84].

3.1.4 Adiabatic mapping

The first step of a conformational analysis of carbohydrate oligomers is to calculate the adiabatic energy maps of the relevant dihedral angles, to get an impression of where the energetically most stable structures are placed in the applied force-field. Maltose is a well described disaccharide, and the adiabatic map for this molecule has been published in multiple articles [85–87] primarily calculated in the MM3 force-field [88], but also using the

CSFF [62], and more recently DFT based calculations have been published [89–91]. The MM3 adiabatic map for maltose [85] is illustrated in the left part of Figure 3.1 and will be used as the reference for $\alpha(1\rightarrow4)$ glucosidic linkages.

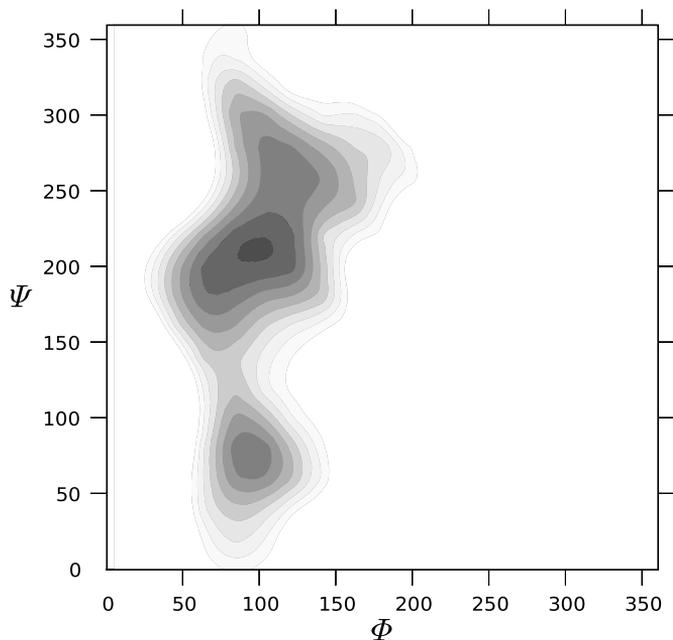


Figure 3.1 Adiabatic MM3 energy map of the $\alpha(1\rightarrow4)$ linkage in maltose

This adiabatic map of the maltosides, and also maltose, was calculated by generating structures with Φ , Ψ torsions increments of 20 degrees giving a 18x18 point grid. For all structures in the grid, every combination of the following degrees of freedom was considered; the clockwise and reverse-clockwise orientation of the secondary hydroxyl groups, the *gg*, *gt* and *tg* conformations of the primary hydroxyl groups. Consequently, the grid point of the adiabatic map was formed from the lowest energy structure out of 36 possible conformations.

The adiabatic energy map for the $\alpha(1\rightarrow6)$ linkage has been published in a similar study for iso-maltose [92], wherein the third dihedral angle ω is also described.

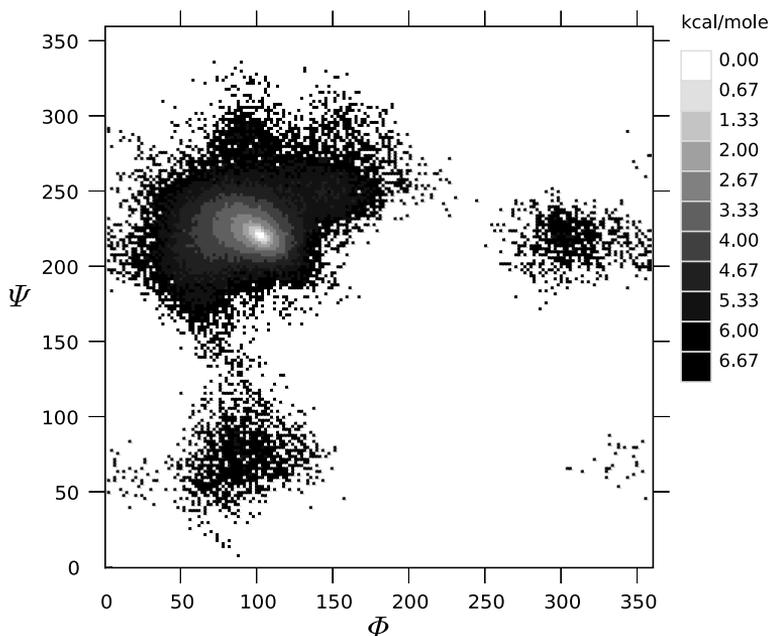


Figure 3.2 High temperature conformational map of the $\alpha(1\rightarrow4)$ linkage in maltose, calculated in the CSFF force field using a continuous solvation model

3.1.5 Fast generation of conformational maps

An alternative approach for rapid generation of conformational energy maps is to run molecular dynamics at an elevated temperature. The elevated temperature increases the rate of conformational change, and makes the system able to cross, higher energy barriers. The molecular dynamics simulation can be run using a continuous solvation model [93–96] to take implicit solvent effects into account as a faster replacement for the explicit solvation model normally used.

This strategy was first proposed by Bruccoleri [97] and recently Frank et al. [98] established that a 10 ns molecular dynamics simulation at 1000 K is sufficient for reaching a conformational equilibrium in glucosidic linkages. The temperature level was decided from a series of simulation at temperatures from 400 K to 1500 K and it was found that the conformational energy landscape was approximately the same as long as a conformational

equilibrium was reached. A 200 ns MD simulation of the tetra-saccharide Sialyl-Lewis^x at 1000 K was used to show that all relevant conformations are well represented after 10 ns. After reaching an equilibrium the potential of mean forces can then be calculated from the conformational probability distribution $p(x)$ by the Boltzmann equation

$$W(x) = -kT \ln(p(x)) \quad (3.5)$$

Setting the highest occupied conformation as a zero minimum, an energy landscape can be generated from the population counts of the simulation time frames, which shows the relative difference in energy with variation of the dihedral angles.

3.2 Trajectory Analysis

The data sampled from the molecular dynamics simulation - called the trajectory - is a series of coordinates for each atom for each time step, which is a multiple of the Δt selected as described in Section 3.1.2. The CHARMM trajectory format in which coordinates for each time-frame is stored as an appended lists of x , y , and z respectively as in the vector,

$$\underline{p} = (x_1, x_2, x_3, \dots, x_n, y_1, y_2, y_3, \dots, y_n, z_1, z_2, z_3, \dots, z_n) \quad (3.6)$$

does not by itself contain information on the sampling time step, or the atom types. This information is stored externally in the setup files and has to be taken into account when running or writing analysis programs. In the case of the TINKER trajectory format, information about the atoms is stored in the trajectory, but not the time frame information. As illustrated in **Figure 3.3**

the coordinates in this format is stored in a sequentially different order. It is important to understand these data structures in order to do the analysis correctly since custom analysis programs was used in this project, and interfaces had to be added in order to switch between formats.

From the molecular dynamics trajectory, a wide range of properties can be calculated. Only a few select examples of analysis will be presented here taken from the applications on the systems studied in Paper I and Paper II.

1491	OT	7.340797	-3.264748	7.770085	720	1492	1493
1491	OT	7.263458	-3.293508	7.713812	720	1492	1493
1491	OT	7.189438	-3.330907	7.749906	720	1492	1493
1491	OT	7.083248	-3.349643	7.766430	720	1492	1493
1491	OT	6.979918	-3.345680	7.710615	720	1492	1493
1491	OT	6.968187	-3.400087	7.554906	720	1492	1493
1491	OT	7.034765	-3.472845	7.395083	720	1492	1493
1491	OT	7.058266	-3.549659	7.399359	720	1492	1493
1491	OT	7.001894	-3.661443	7.518559	720	1492	1493
1491	OT	6.903582	-3.713775	7.598263	720	1492	1493

Figure 3.3 Excerpt from a trajectory file in the TINKER xyz molecular file format, showing sets of a water oxygen coordinates for 9 molecular dynamics time steps. The elements of each line are the; atom number, atom label, x y z coordinates, atom type, and atom numbers of those bound to the entry

3.2.1 Diffusional properties

A simple dynamics property to calculate from the trajectory is the self-diffusion coefficient of a molecule. This quantity is related to the time dependent displacement of the molecular center of mass by a time correlation function which for long time periods can be approximated by an Einstein relation as [73]

$$6tD = \langle |\underline{r}(t) - \underline{r}(0)|^2 \rangle \quad (3.7)$$

The self-diffusion coefficient has been measured by Rayleigh interference for xylose and maltose [99], to be approximately 7.4 and 5.2 cm^2/s for low concentrations. The self-diffusion provide a good parameter for checking the validity of the molecular dynamics setup parameters. In more recent literature the self diffusion coefficient of maltose has been measured to be 4.1 $10^{-6}cm^2/s$ by NMR, and 4.2 $10^{-6}cm^2/s$ by Rayleigh interference [100].

Another diffusion mechanism is caused by the molecular tumbling of the molecule. This has to be described by a time correlation function of a vector that describes the orientational time dependency of the molecule. This is usually done by - following the examples by Rossky et al. [101] - calculating the angular displacement of the molecular dipole moment vector, and fitting the result to an exponential decay function, to estimate the auto-correlation

time τ_c . This was done for the trisaccharide systems in Paper II and the result is illustrated in **Figure 3.4**.

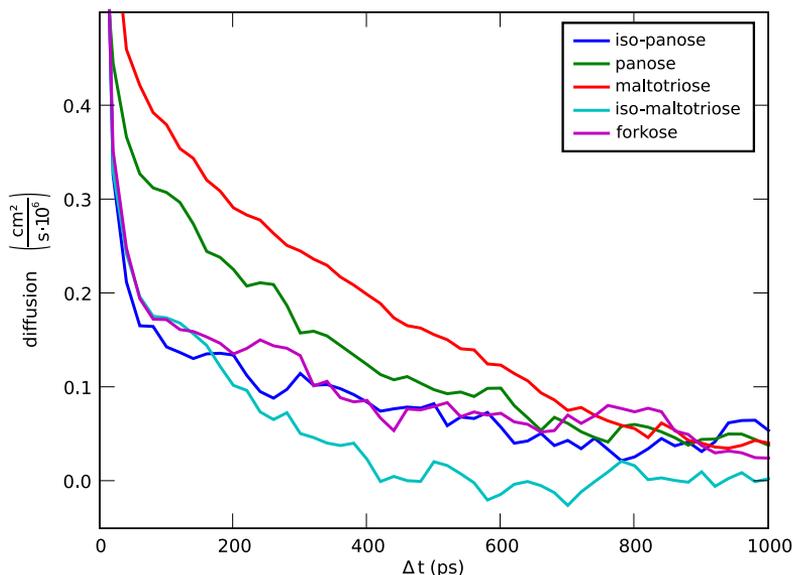


Figure 3.4 Rotational diffusion profiles for the five trisaccharides studied in Paper II, calculated using angular auto-correlation of the dipole moment vector

Using the dipole moment as a representative of the molecular orientation is a robust method only for molecules that are rigid and does not change their overall conformation during dynamics. However, for flexible molecules such as the trisaccharides of Paper II, the dipole moment will not only rotate as a function of molecular tumbling, but will also show rapid fluctuations as a function of conformational changes of the inter-residue linkages.

3.3 Hydration

The main objective of this thesis was to investigate the interaction of starch carbohydrate structure and its surrounding aqueous solvent. A few studies in this area have been published examining, and supporting, the point of

view that there is an important close relationship between the conformation of the glucosidic bonds and the surrounding water molecules [84, 86, 102–105].

For the purpose of using molecular dynamics simulations as an analytical tool, in studies of the detailed behavior of water participating in the structuring of the solute as a more static element than free water, a number of methods have been developed. Many of those have been reviewed by Engelsens [106] in a case study of sucrose, and a few of the approaches, which are used in Paper II and Paper IV, will be introduced in the following sections.

3.3.1 Radial pair distributions

The one dimensional pair-distribution gives the probability of finding a water molecule at the distance r from an atom, compared to the probability of finding two water oxygen atoms separated by the same distance in bulk water [101, 107].

$$g(r) = \frac{N(r)}{\rho_{water} \frac{4}{3}\pi \left((r + \Delta r)^3 - r^3 \right)} \quad (3.8)$$

The calculation can, of course, be performed for any element in the solute, but when looking for structural waters bound to carbohydrates, the oxygen atoms are the most interesting, since they are the primary sites for hydrogen bonding. As an example of the result of such pair distribution calculations is given in **Figure 3.5** for all oxygen atoms of the reducing end of iso-panose. It is seen that the solvent is organized around the oxygen atoms in ‘layers’, which are referred to as hydration shells, and that as many as three of them can be identified for most of the curves. This type of calculation clearly shows which oxygen atoms are more hydrated than others, but not directly if they are sharing the static waters with other sites in the solute.

Two dimensional pair-distribution calculations [108], is a generalization of the one dimensional version from above, and is expressed by Equation 3.9:

$$g(r_1, r_2) = \frac{N(r_1, r_2)}{\rho_{water} V_{intersection}(r_1, r_2, \Delta r)} \quad (3.9)$$

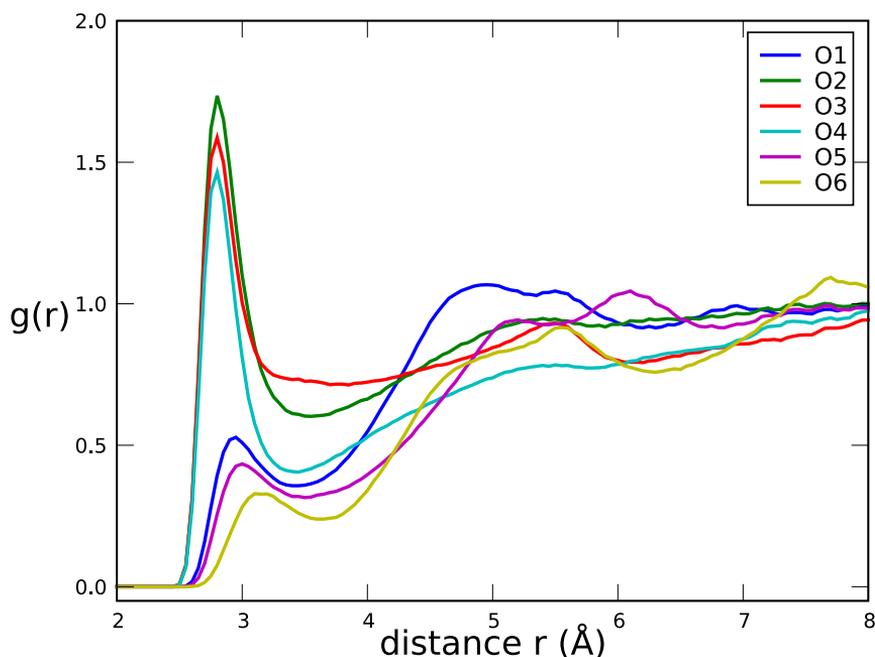


Figure 3.5 The one dimensional radial pair distributions for five oxygen atoms in the reducing end unit of iso-panose. It is seen that the first hydration shell at approximately 2.8 Å has a high density for O2, O3, and O4

This gives a more detailed description of the localized water environment in the area between any two given atoms. This calculation is able to give information which is precise enough to pinpoint the exact location of bridging waters, if they exist.

An example of such a distribution is illustrated in **Figure 3.6**, revealing a water bridge in the iso-panose structure between O1' and O5", where the probability of finding a water molecule is 2.8 times that of the average solvent.

This approach was used extensively in both Paper II and Paper IV - to identify several non-trivial water sharing oxygen pairs which are described in Section 5.2.3 and Section 6.4 - and is preferable to three dimensional density contouring analysis as presented in a study by Liu and Brady [109], which requires step-wise reorientation of the system, and thus becomes more complex and less accurate for flexible solutes. From the results described in

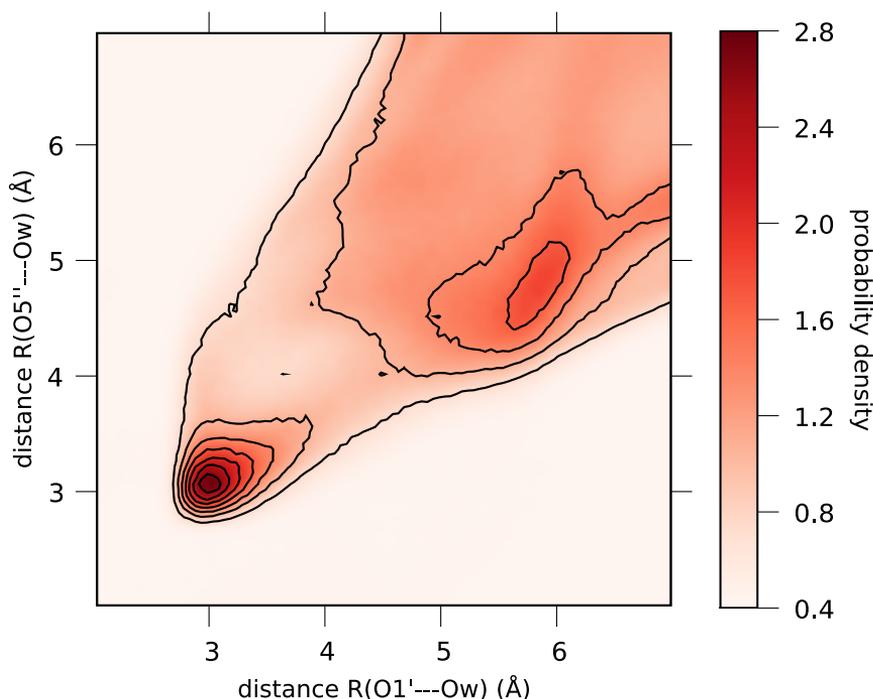


Figure 3.6 The two-dimensional radial pair distribution for bridging water between O1' and O5'' in iso-panose. The maximum is 2.8 times the bulk water density

Section 5.2.3, it is important to consider all possible combinations of solute oxygen atoms since some structural waters are bound in non-obvious sites.

3.3.2 Finding hydrated structures

For further inspection of the structure of the water binding sites, the dynamics of the water binding, and for visualization purposes, it is desirable to find points in time in the molecular dynamics that are the start- and end-point of a period where a specific water molecule is bound to a specific site. First step is to locate a significant water binding site using two dimensional radial pair distribution plots. Next, given the oxygen atoms involved, the nearest water molecule can be found for each time step in the trajectory. Plotting those data gives a picture of how often the site is hydrated, for

how long, and which water molecule is present at any given time. Such an example of a time frame with a present water bridge is illustrated by **Figure 3.7**, stripped for all other water molecules.

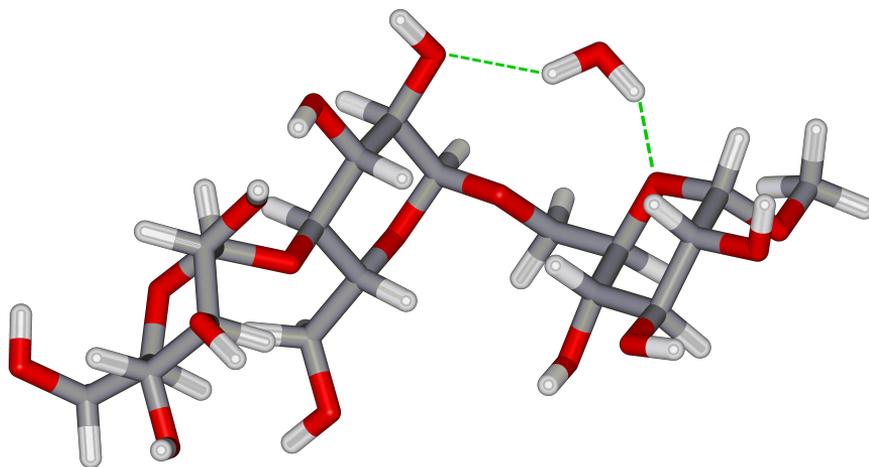


Figure 3.7 The structure of iso-panose, including the highest populated water bridge as also illustrated in **Figure 5.10**

4 Measurement of carbohydrate structure

In this chapter, a short survey of the most important methods for measurement of molecular structure will be summarized, along with important results relating to the field of starch research. Subsequently, an explanation of the chosen experimental protocol for assignment of NMR spectra in the following chapters will be given, followed by the theoretical foundation for relating measurements of scalar couplings and NOEs to carbohydrate structure.

4.1 Single crystal structures

X-ray diffraction has been the general tool of choice used for structure determination of almost any kind of molecular structure for many years since the early demonstration by Bragg [110], and has really picked up momentum after the work of Franklin [111–112], which lead to the famous double helix DNA model by Watson and Crick [113]. Today, state of the art synchrotron X-ray sources produce data of very high resolution that can be better than 1.0 Å. However, one limitation of this technique is that it only gives a static image of the system, which is the energetic minimum of the molecule when arranged in a crystal lattice sometimes including water. While the interpretation of the crystal structure is justifiable as a measure of the energy minimum, it is rarely the same as - but often similar to - the preferred conformation found in solution. This has been shown both by optical rotation experiments [114] and also by several of the NMR studies mentioned in Section 4.2. Another drawback of crystallography is the inherent requirement for the studied species to be able to crystallize. This obstacle is still the primary bottleneck of the method, particularly in the area of carbohydrates, where single crystals are very hard to obtain because of the flexible nature of the molecules.

Some interesting structures - in relation to starch research - has been measured by single crystal diffraction. The structure of α -D-glucose has already been mentioned in **Section 2.2**, and will serve as the basis of all subsequent molecular modeling work.

First and foremost, the structure of α -maltose is interesting in relation to

the α -glucan backbone of starch. The X-ray structure for this disaccharide was solved by Takusagawa et al. [115], who found the conformation of the non-hydrated crystal to be ($\Phi = 116^\circ, \Psi = 242^\circ$), favoring a hydrogen bond between O2 and O3'. β -maltose has also been the subject of extensive structural studies [86, 116–118], yielding similar results, and it seems reasonable to assume that the change in anomeric conformation does not have a dramatic effect on the conformation of the linkage.

The maltose structure is complemented by the structure of maltotriose, the $\alpha(1\rightarrow4)$ linked trisaccharide in **Figure 5.2** [119]. Here the conformation was found to be ($\Phi = 82.2^\circ, \Psi = 211.1^\circ$) between the non-reducing ring and the middle one, and ($\Phi = 82.8^\circ, \Psi = 208.2^\circ$) for the linkage to the reducing ring. It is seen that this molecule provides a much better model for the optimized conformation found in amylose double helical structure as described in Section 2.5.

A crystal structure for the trisaccharide α -panose, which is involved in the amylopectin branch point, has also been solved [120], giving information on both types of glucosidic linkages. The dihedrals found in this structure is ($\Phi = 92.9^\circ, \Psi = 228.7^\circ$) for the $\alpha(1\rightarrow4)$ bond and ($\Phi = 71.4^\circ, \Psi = 165.2^\circ, \omega = 75.7^\circ$) for the $\alpha(1\rightarrow6)$ bond. A further refinement of this structure has later been published by Jeffrey et al. [121] whose results were ($\Phi = 96.8^\circ, \Psi = 225.2^\circ$) for the $\alpha(1\rightarrow4)$ bond and ($\Phi = 70.8^\circ, \Psi = 167.3^\circ, \omega = 74.7^\circ$) for the $\alpha(1\rightarrow6)$ bond. Returning to the amylose double helix structures of Section 2.5, their crystal structure were measured by Imberty et al. [26] by a combination of diffraction on both powder and single crystals prepared from potato starch by a hydrolysis method [122]. This led to the conclusion that the helix was built from a maltotriose strand with dihedrals ($\Phi = 91.8^\circ, \Psi = 206.8^\circ$) and ($\Phi = 85.7^\circ, \Psi = 214.7^\circ$). This result was later optimized by O'Sullivan et al. [123] to fit into a 6-fold helical model with the dihedrals ($\Phi = 92.0^\circ, \Psi = 218.0^\circ$).

It should also be mentioned that one larger carbohydrate have been crystallized. A model of a V-type amylose single helix, cyclomaltohexaicosaoase, has been prepared as a single crystal and the structure was determined [124] at atomic resolution, and the helical conformation was found to be ($\Phi = 103.6^\circ, \Psi = 233.5^\circ$)⁵.

⁵ The Ψ dihedral is translated from Gessler's use of C1-O1-C4-C3 as reference to the definition used in this thesis.

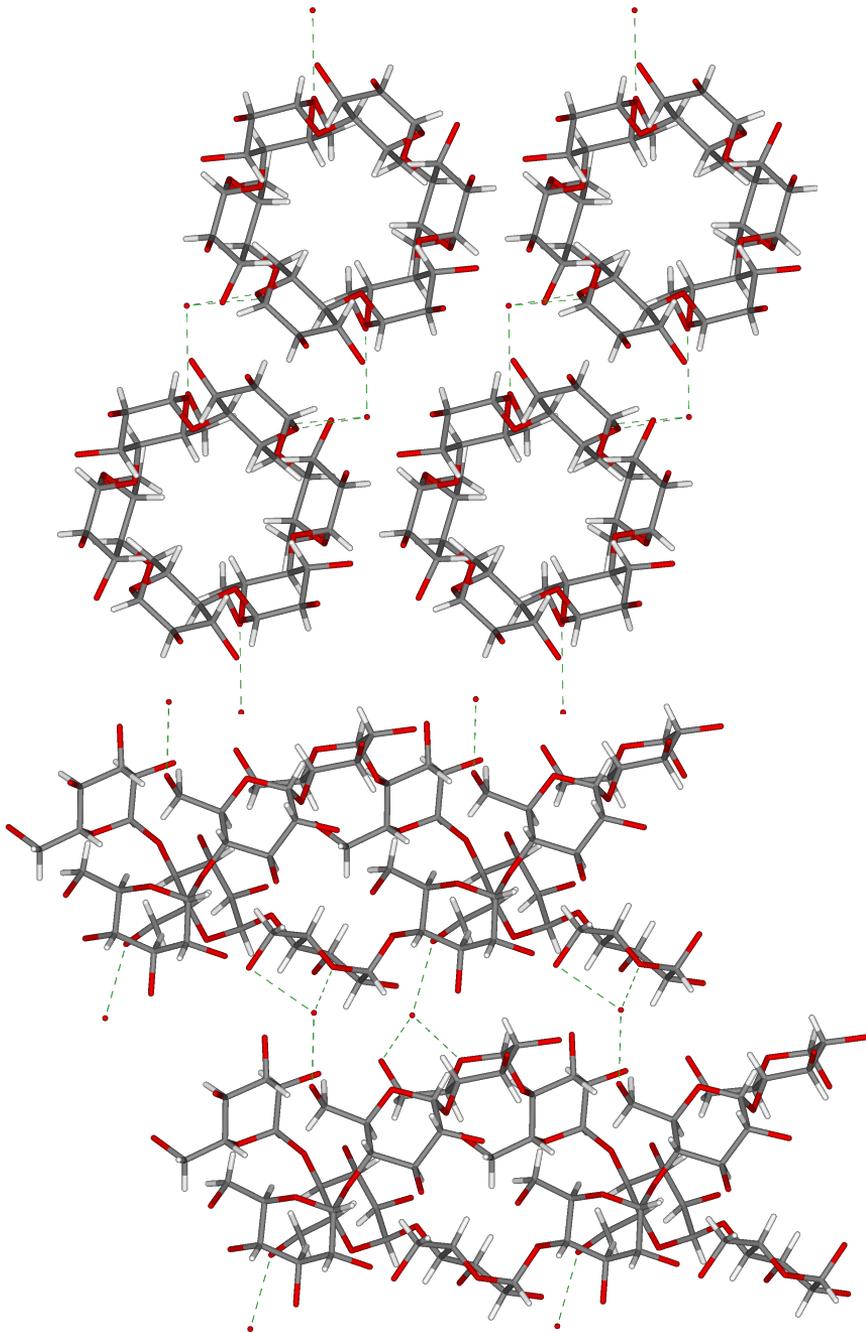


Figure 4.1 Intermolecular hydrogen bonds in the A-type starch crystal, seen as a top view (top) and side view (bottom) Water hydrogen atoms are excluded

Single crystal structures

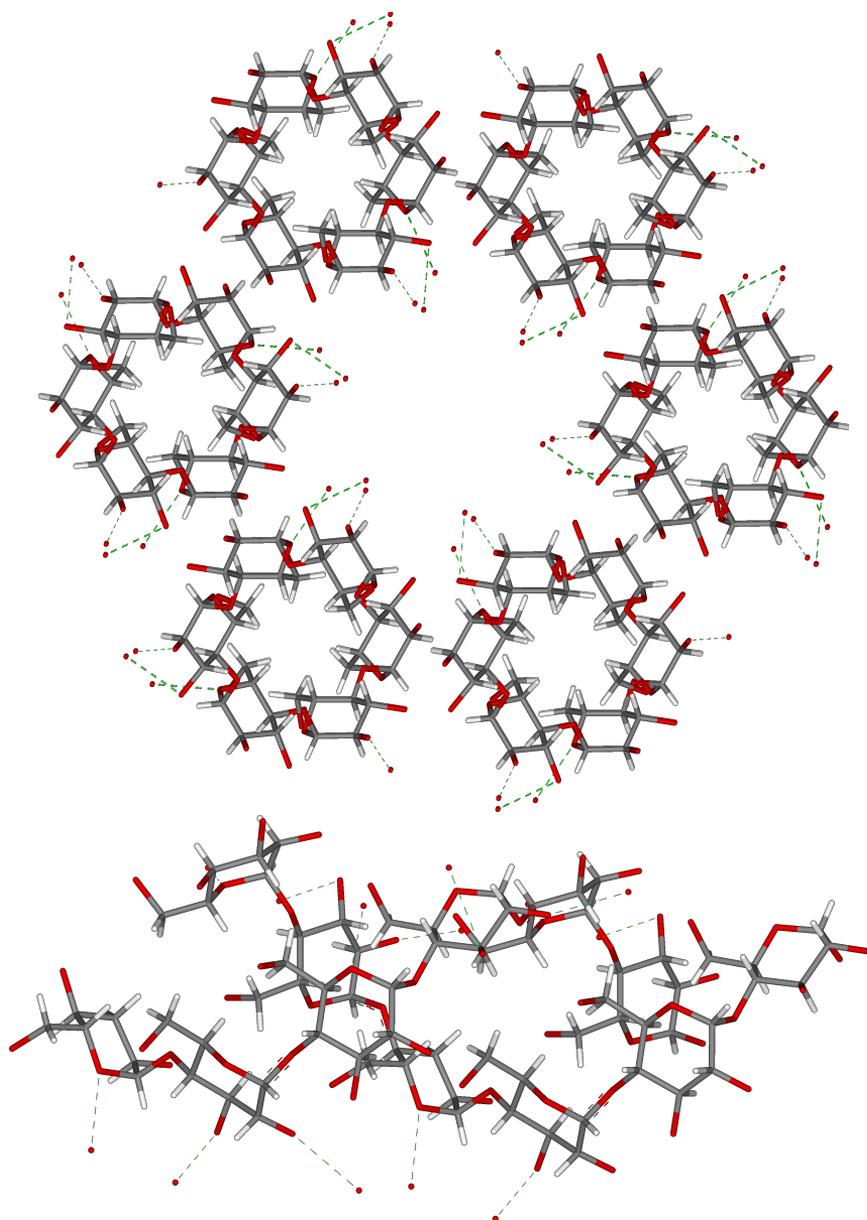


Figure 4.2 Intermolecular hydrogen bonds in the B-type starch crystal, seen as a top view (top) and side view (bottom) Water hydrogen atoms are excluded

While the crystal structures of both maltose and α -panose are not hydrated, maltotriose is reported to be a tetrahydrate [119]. In this crystal water molecules are bound by hydrogen bonds to O2 and O3 respectively in the non-reducing end, O5 in the middle glucose unit, and O4 in the reducing end.

Also the A- and B-type crystal structures each include a number of water molecules, which is interesting in relation to the studies of carbohydrate hydration in Paper II and Paper IV. By closer investigation of the A-type structure illustrated in **Figure 2.8**, it is observed that the water molecules are able to form hydrogen bonds connecting adjacent helices by building bridges between them, using only O5 and O6-H6 atom. No water molecules are found to be bound intra-helical. The possible hydrogen bonding network is given by **Figure 4.1**. In the B-type crystal structure - where much more water is present - as illustrated in **Figure 2.9**, only few water molecules are located close enough to participate in hydrogen bonding. The remaining water molecules have been removed in **Figure 4.2** where the possible are also shown. It is observed here that there are no water molecules bridging between helices, but that the single helices are hydrated systematically in in a two unit sequence with water on O2 and O3 in one glucose unit, and then on O5 in the next.

4.2 NMR spectroscopy

Since the introduction of NMR by Rabi et al. [125–126], its further development by Bloch et al. [127] and Purcell et al. [128], NMR spectroscopy has become one of the most valuable tools for obtaining structural and dynamic information on bio-macromolecules.

The main problem in analysis of NMR spectra of starches and starch models, is the lack of diversity of the residues. As described in **Chapter 2**, the only monomer in the entire polymer is glucose, and even though it is a branched structure with two different possibilities for linkages, the homogeneity is massive. This chemical similarity of residues is expressed in the spectra as a severe degree of signal overlapping, which makes complete chemical shift assignments very difficult, but possible for smaller saccharides.. The development of modern hetero-nuclear 2D NMR experiments has provided tools for making this task easier. Only a few of these experiment will be mentioned here, but for a wider overview of the field see the

reviews by Mulloy [129] and Duus et al. [130].

Since the aim of the present study was to investigate dynamic and hydration properties of starch models through the use of molecular dynamics simulations, it was imperative to obtain experimental data which could support the theoretical results. Therefore a complete resonance assignment is required in order to measure specific scalar couplings, and to assign NOEs, since both of these are sources of structural information. Scalar couplings have been used as experimental support of modeling on several occasions [86, 131], and the approach has recently been validated on a large set of disaccharides by Cheetham et al. [132], who found good agreement in most cases, except for $\beta(1\rightarrow4)$ linkages. How to use the scalar couplings as structural information will be explained in Section 4.2.2.1 and Section 4.2.3.

Scalar couplings was used early on by Parfondry et al. [133] to show that the glucosidic linkages in saccharides tend to favor staggered conformations more in solution than what has been found in crystals.

Examples of the usage of NOEs as a source of structural information on a saccharide is the study by Bothner-By et al. [134] which only gave sequential linkage information and a qualitative idea about the conformation, and also the structural determination of β -maltose, using hetero-nuclear scalar couplings and NOEs has been demonstrated [135].

4.2.1 Assignment of NMR spectra

All the ^1H and ^{13}C resonances from the model saccharides described in Section 5.1 (Paper II) and Section 6.1 (Paper IV) was successfully assigned. To perform the assignment of each structure the used strategy was to begin with the anomeric protons in the COSY [136] spectrum and follow the cross-peaks into the ring-structure until signal overlap is too severe to continue. From that point, most of the remaining resonances could be assigned using the TOCSY [137] spectrum, in many cases using the COSY spectrum as support.

The H6 protons in glucose rings are coupled to a proton (H5) bound to an anomeric carbon (C5) and are therefore not equivalent. Consequently they do not have the same chemical shift and will be split into a multiplet, by

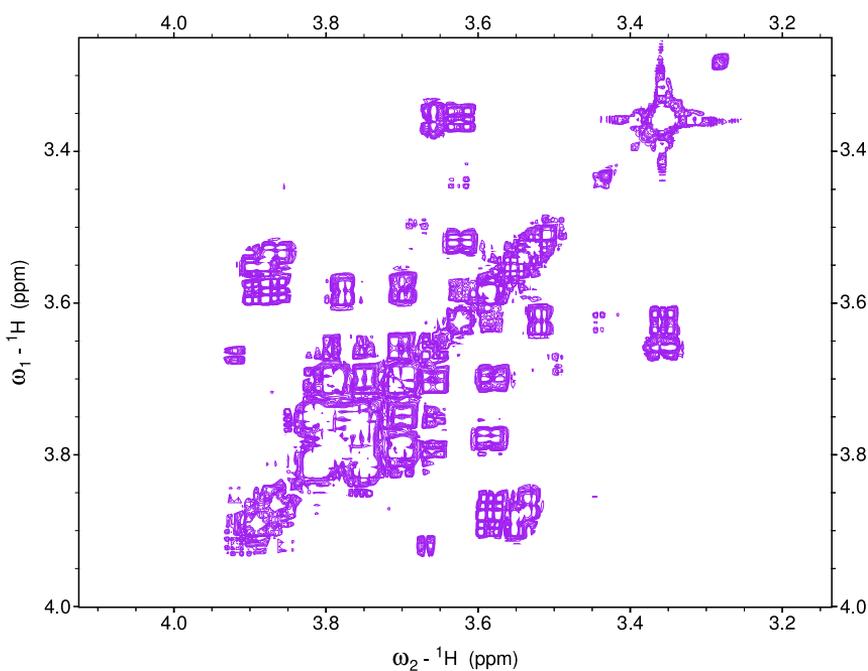


Figure 4.3 Non-anomeric region of the 950 MHz COSY spectrum of maltotriose, showing how the cross peaks close to the diagonal are difficult to resolve

scalar coupling, that is difficult to interpret. The HSQC [139–140] spectrum was often helpful in resolving this, besides providing all the carbon chemical shifts through the hetero-nuclear correlation. An example of the HSQC spectrum can be found in **Figure 4.5**, not including the anomeric region.

Since every glucose ring is an isolated spin system, and that the goal is to measure scalar coupling across the glucosidic linkage, it is necessary to determine the sequence of the rings. For this purpose, HMBC [141] spectra was acquired, in which it was possible to observe correlations between the linkages. In the case of the trisaccharides, the methyl group on the reducing end was easily identifiable and could also be correlated to H1/C1.

In the initial intra-residual assignment, the rings were labeled *a*, *b*, and *c* respectively, in such a way so that the resonances in the ring with the anomeric proton at lowest field were labeled *a* and the ring with the anomeric

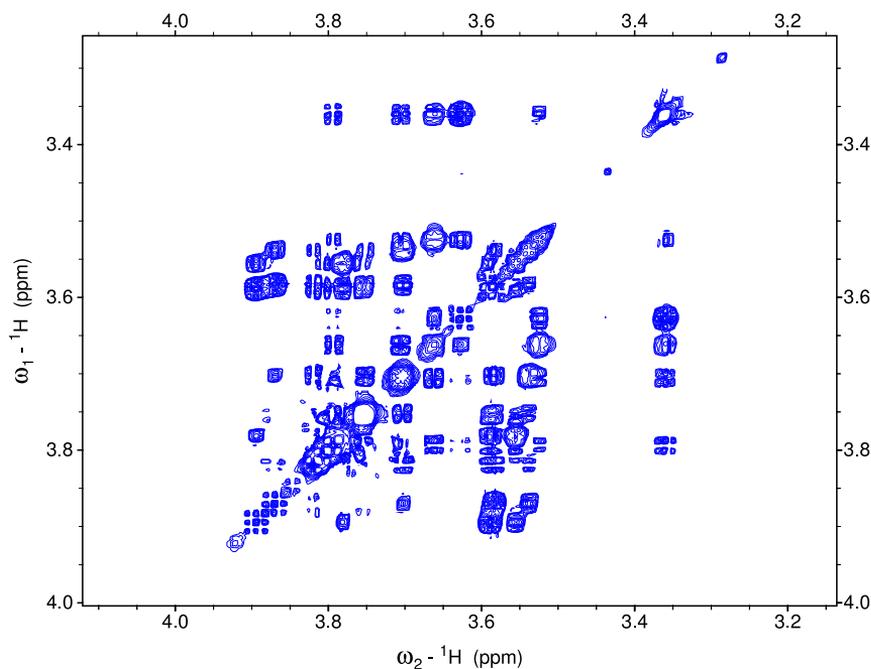


Figure 4.4 Non-anomeric region of TOCSY spectrum of maltotriose, showing how the extra correlations and more suppressed diagonal of the TOCSY spectrum can help assigning spin systems. The spectrum was recorded at 950 MHz, with a mixing time of 80 ms using the MLEV-17 [138] pulse train

proton at the highest field *c*. Subsequently the set of rings could be translated to an ordered sequence by using HMBC inter-glucosidic connectivities. As an example, the sequential assignment of the α -forkose linkage structure in the right part of **Figure 5.4** was:



The HMBC spectra also proved to be useful in cases with partially overlapping ^1H - ^{13}C resonances in HSQC. These could then be solved by correlation with protons bound to another carbon.

For the handling of NMR data the processing tools NMRPipe/NMRDraw

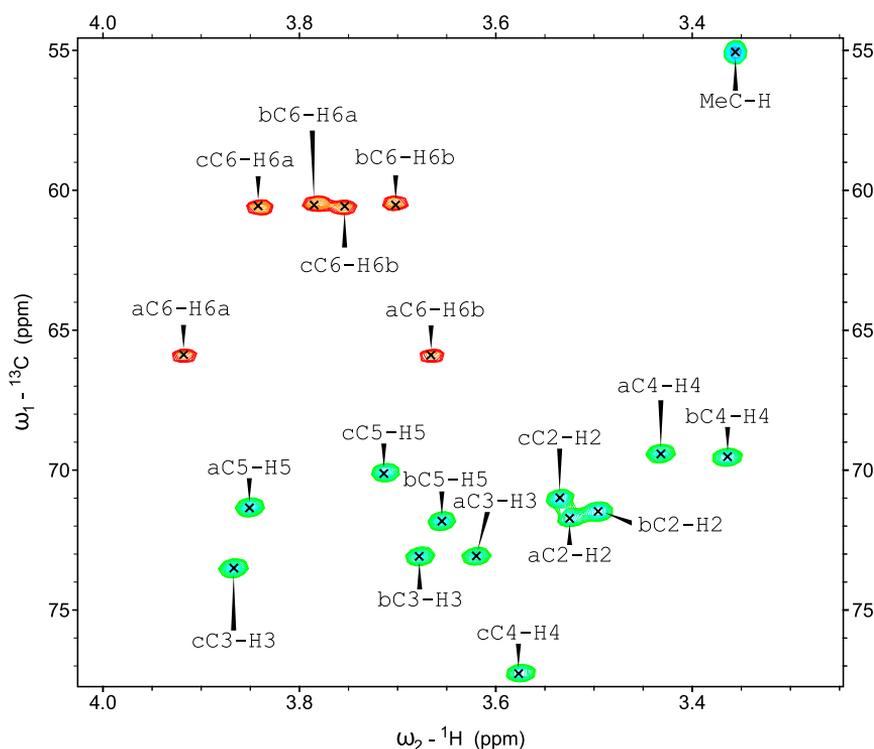


Figure 4.5 950 MHz HSQC spectrum of the non-anomeric region for α -panose, including assignment labels

from Delaglio et al. [142] was used for transformation and phase correction. The assignment of resonances was performed with the aid of the Sparky program from Goddard et al. [143]

4.2.2 Measurement of hetero-nuclear scalar couplings

One of the most valuable informations we can get from NMR experiments, is geometric information about the dihedral angles of the glycosidic bond connecting the glucose units. In the simplest case of maltose derivatives, two angles are needed to define the $\alpha(1\rightarrow4)$ linkage. One possible way of measuring these structural parameters, is through the hetero-nuclear $^3J_{CH}$ scalar couplings of the C-O-C-H connectivity. The relevant couplings across the $\alpha(1\rightarrow4)$ linkage are $^3J_{H1-C4'}$ and $^3J_{C1-H4'}$, whereas for the $\alpha(1\rightarrow6)$ bond the two couplings $^3J_{H1-C6'}$ and $^3J_{C1-H6'_{a,b}}$ are of interest.

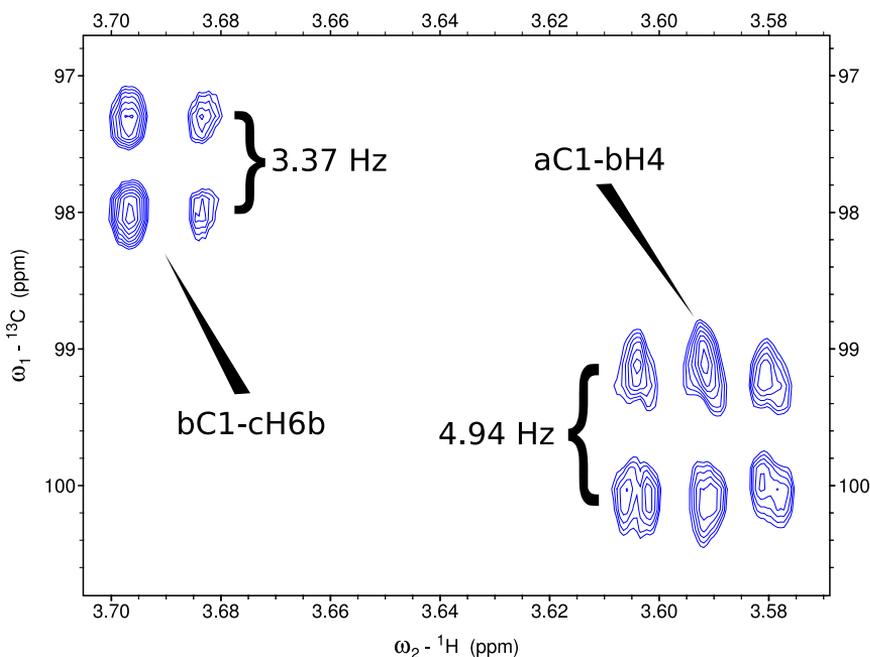


Figure 4.6 A region in the 800 MHz J-HMBC spectrum of *iso*-panose, showing hetero-nuclear 3-bond correlation across both the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkage, using a scaling factor of $\kappa = 40$

These couplings indirectly represent the Φ and Ψ dihedrals defined in **Figure 2.3**. Since the measurement of this type of coupling is not trivial, different experiments were tested before selection of the J-HMBC experiment from Meissner and Sørensen [144]. This experiment seems to offer a good signal to noise ratio and a straightforward measurement of the desired coupling constants.

4.2.2.1 The Tvaroska parametrization

A Karplus-like [145–146] description of the C-O-C-H dihedral angle found in the glucosidic bond has been published by Tvaroska et al. [147], which included more experimental data than a previous study [148]. The hetero-nuclear coupling constant is in this way described as a function of an angle by Equation 4.1a. This parametrization was performed by measuring the relevant coupling constant for a series of synthetic mono-saccharide model

compounds known to be conformationally restricted, and for which the X-ray structure has been solved.

$${}^3J_{CH}(\phi) = A + B \cos \phi + C \cos^2 \phi \quad (4.1a)$$

$$A = 0.5, B = -0.6, C = 5.7 \quad (4.1b)$$

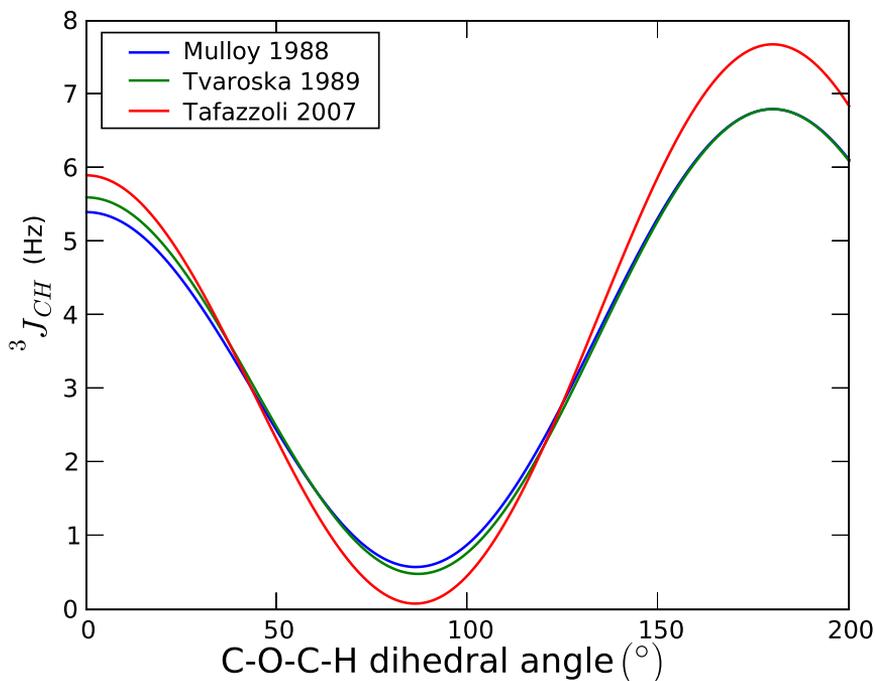


Figure 4.7 Hetero-nuclear ${}^3J_{CH}$ -coupling across the glucosidic linkage parametrized as a function of the C-O-C-H dihedral angle by equation Equation 4.1a, showing reasonable agreement between experiment and theory

This type of relation is a useful tool as it directly relates the torsional conformation to a measurable quantity. The model has recently been shown to be in good agreement with predicted values from density functional theory calculations [149] on substituted glucose and galactose, which suggests some corrections to the parameters in Equation 4.1b. The new parameters calculated from theory are:

$$A = 0.11, B = -0.89, C = 6.68 \quad (4.2)$$

Both parametrizations are plotted in **Figure 4.7**, using the model function of Equation 4.1a, showing minor differences of the results. In the interpretation of coupling constants of this type in Paper II and Paper IV, a decision was made to rely only on the model derived from experimental results, instead of validating simulation data with a theoretical result. Still, the good agreement with theory further strengthens the model, since a weak point of the experimental study was the lack of structures in the 10° – 60° area. However, this functional relationship should be used with caution. While Equation 4.1a does permit the calculation of a corresponding coupling constant from a given dihedral angle, it is not as trivial to derive an angle directly from the measurement of a coupling constant. Thus, when used in this thesis, time averaged coupling constants are calculated from molecular dynamics trajectories and compared to experimental results. This provides a measure of the validity of the simulation data, but is difficult to interpret as a direct measure of the molecular conformation.

4.2.3 Homo-nuclear scalar couplings

The geometry of the ω dihedral can be obtained using the standard ${}^3J_{HH}$ coupling between H5 and H6, again with the requirement of H6 resonance assignment, which is often difficult. The homo-nuclear scalar couplings are present in both the COSY and TOCSY experiments, and should be interpreted via another Karplus type parametrization.

4.2.3.1 The Haasnoot-Altona parametrization

Using the generalized parametrization of Haasnoot et al. [150], the relationship between the scalar coupling constant and the proton-proton dihedral angle can be described by equation 4.3a.

$$J(\phi) = A \cos(2\phi) + B \cos(\phi) + C \sin(2\phi) + D \quad (4.3a)$$

$$A = 4.205, B = -0.990, C = 0.104, D = 5.591 \quad (4.3b)$$

$$A = 4.205, B = -0.990, C = -2.211, D = 5.591 \quad (4.3c)$$

Two sets of parameters are given since the two protons H₆₁ and H₆₂ are different as illustrated in **Figure 4.8** by the three staggered conformations around the C₅ – C₆ bonds found in saccharides.

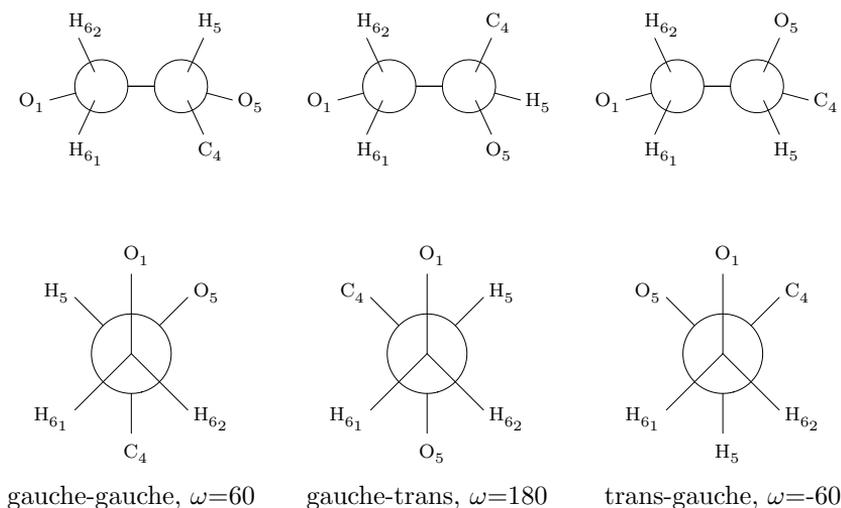


Figure 4.8 The staggered conformations, around the C₅-C₆ bond of a glucose molecule, seen from a side view (top) and the corresponding Newman projections (bottom), looking along the bond from the C₆ end

The functional relationships between the conformation and scalar coupling are plotted for both H₆ protons in **Figure 4.9**.

For this type of parametrization, the same warnings about back calculation of dihedral angles as mentioned in Section 4.2.2.1 are valid.

4.2.4 Nuclear Overhauser Effects

The nuclear Overhauser effect [151–152] (NOE), which is a through space magnetization transfer effect, is widely used to provide structural restraints on molecular structures, using the two dimensional NOESY [153], or the rotating frame cross-relaxation [134] equivalent experiment, ROESY [154]. NOEs can in some cases help with assignment problems by providing an inter-residue relationship, and often help with intra-residual assignment. The use of inter-residue NOEs for studying structural conformation are

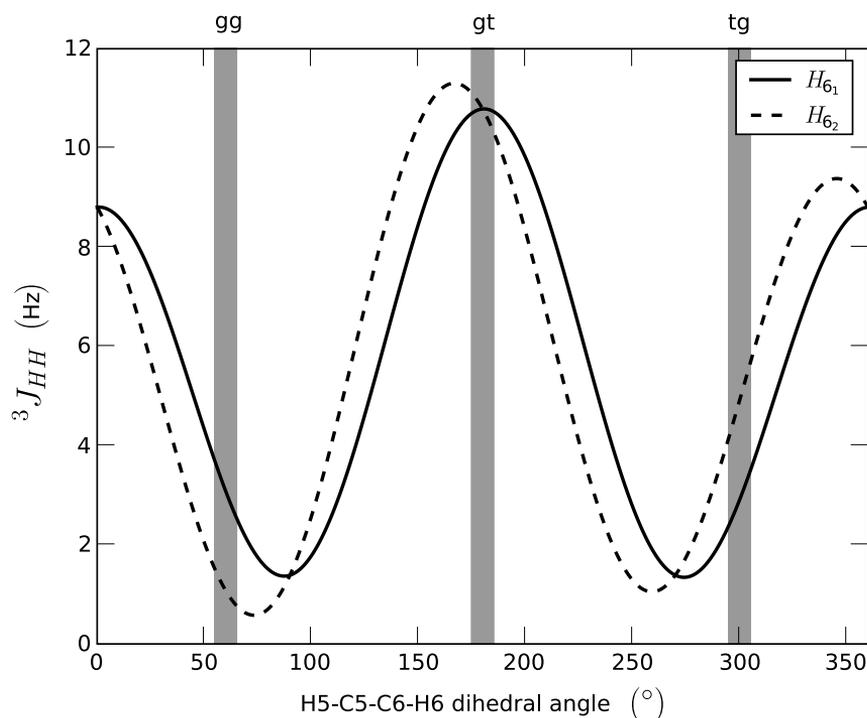


Figure 4.9 $^3J_{HH}$ -coupling between H5 and each of the H6 protons described as a function of the dihedral angle by equation 4.3. The vertical bars indicate the three most stable conformations *gg*, *gt*, and *tg*

limited in carbohydrates, since most often only one constraint can be observed, for example between H1 and H4'. In addition, the very flexible nature of saccharides will cause the proton-proton distance to be widely distributed because of the many conformations visited, which will make the effect weaker than for the rigid intra-ring proton pairs.

An example of the low number of NMR measurable constraints in saccharides are illustrated in **Figure 4.10** and compared to a segment of the peptide backbone, where several NOEs in addition to a scalar coupling are often available. As mentioned in Section 4.2.2.1, the interpretation of scalar couplings can be difficult, as up to three solutions exist to the inverse Karplus equation. Having multiple NOE distances is a stronger constraint on the conformational space, which is why the four measurables for the peptide bond outweigh the three for the glucosidic counterpart. This

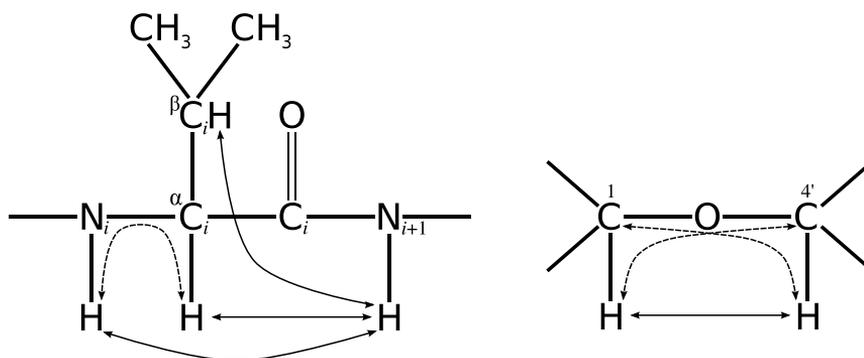


Figure 4.10 Comparison of the measurable constraints on the conformation around a peptide bond and a glucosidic linkage. Dashed arrows indicate scalar couplings and filled arrows are NOEs. Inspired by Fig. 8.1 in Wüthrich [155] page 131

lack of distance information makes it very difficult to carry out a structure determination of linear saccharides, which is why accurate hetero-nuclear coupling and good Karplus parametrizations are most important for the work in this field.

4.2.5 Diffusion edited spectroscopy

On a side-note, the differences in diffusional properties of even very similar molecules, as described in Section 3.2.1 was used in Paper V in combination with parallel factor analysis [156–159] to extract individual NMR spectra from a complex mixture of solutes. Following a preliminary study [160] we acquired 2D diffusion edited NMR (DOSY) spectra on a designed series of mixtures of the three saccharides glucose, maltose, and maltotriose. These saccharides have similar diffusion properties [161], and exhibit severe signal overlap in the 1D proton NMR spectrum. The signal intensity in the DOSY spectrum of mixture k can be described a function of the chemical shift Ω , the gradient field strength δ , and the concentrations f of the constituent solutes in the mixture

$$I_{\Omega\delta k} = \sum_{f=1}^F S_{\Omega f} A_{\delta f} C_{kf} \quad (4.4)$$

In this notation $S_{\Omega f}$ is the intensity measured at Ω for compound f only, $A_{\delta f}$ is the reduction in signal amplitude of compound f due to the gradient field, and C_{kf} is simply the concentration of compound f in mixture k . The description of the DOSY signal intensity in Equation 4.4 satisfies the general algebraic expression for the PARAFAC model structure [158]

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (4.5)$$

for which a unique solution can be found. For the designed set of saccharide mixtures it was shown that the DOSY data was well described by the PARAFAC model, and that physically meaningful pure spectra of each component could be estimated by this mathematical approach, even for solutes with very similar diffusion constants. The result of the preliminary experiment [160] is illustrated in **Figure 4.11** showing highly pure spectra of each component.

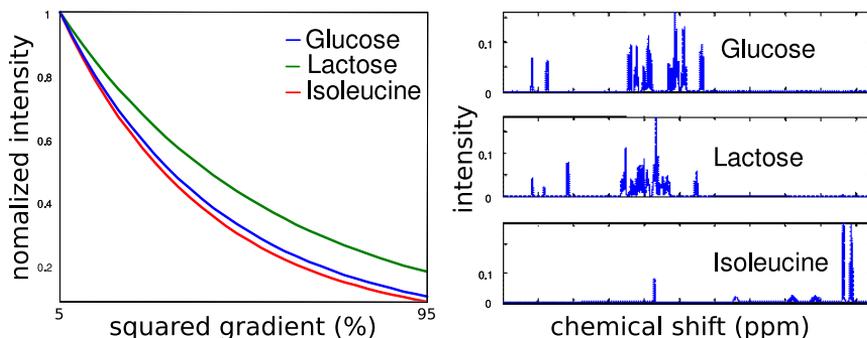


Figure 4.11 PARAFAC resolved diffusion profiles (left) as a function of gradient field strength (percent of maximum) and estimated spectra (right) of the components of a designed mixture set of glucose, lactose and isoleucine. Adapted from Bro et al. [160]

This methodology was further tested on a complex set of metabonomic data from measuring on 20 blood plasma samples. The methylene region of the proton NMR spectra contains information on the concentration of lipo-proteins. The quantification of the amount and distribution of lipo-protein classes is important for diagnostics of diabetes and coronary heart disease, thus a reliable spectroscopic method would be preferable over time

consuming fractionation techniques. Since the lipo-proteins vary in size the Stokes-Einstein relation

$$D = \frac{k_B T}{6\pi\eta r} \quad (4.6)$$

predicts a dependence of the diffusion coefficient on the radius of the molecule.

The DOSY data in this case was resolved by a four-component PARAFAC model, as shown in **Figure 4.12** which corresponds to the four lipo-protein groups, Very-Low-Density- (VLDL), Intermediate-Density- (IDL), Low-Density- (LDL), and High-Density-Lipoproteins (HDL).

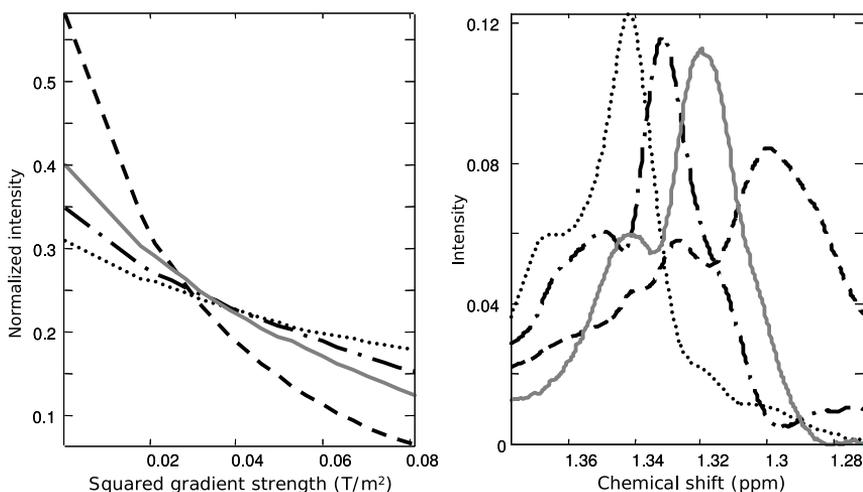


Figure 4.12 PARAFAC resolved diffusion profiles (left) and estimated pure spectra (right) of the four lipo-protein classes in a set of blood plasma samples. Adapted from Dyrby et al. [162]

Assuming spherical geometry of the molecules, the sizes estimated from the diffusion profiles are 42, 26, 18, and 8 nm, in good agreement with the experimental size range estimations previously published by Dyrby et al. [162].

Unfortunately this novel technique is not yet ready to be applied to the starch oligomers under investigation here. Still, the methods hold great promise for the purpose of future comparative studies of diffusional properties of starch oligomers.

5 Starch building blocks

The very basic building block of starch is the α -glucose unit, for which the structure has already been introduced in **Section 2.2** by **Figure 2.1**. Moving up in complexity, the next model systems are the combinations of two glucose units using the two possible linkages. This yields the two molecules α -maltose and iso-maltose. The dynamics and hydration of these two fundamental disaccharides have already been the topic of numerous molecular dynamics studies [85–86, 163–167], all finding that explicit solvent has a significant influence on the molecular structure.

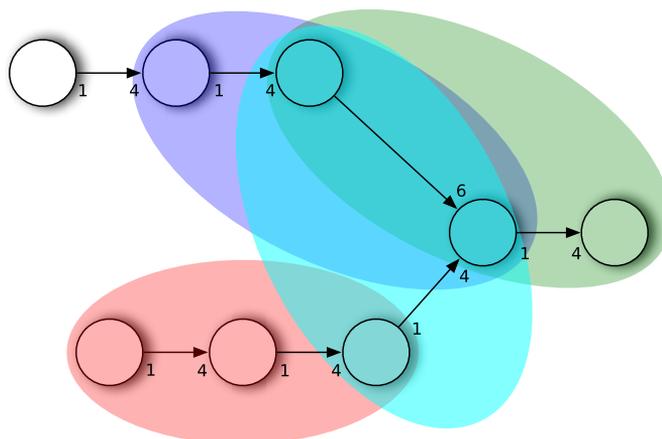


Figure 5.1 Schematic overview of the placement of essential trisaccharides in the amylopectin branchpoint. iso-panose (magenta), α -panose (green), maltotriose (red), and α -forkose (cyan)

Several larger starch fragments have also been studied in the literature, A tetrasaccharide was used by Best et al. [168], and a pentasaccharide by Corzana et al. [84] both with the purpose to be a model of the amylopectin branchpoint. A study by Naidoo et al. of maltohexaose [169] has been published, as a model system for amylose, followed by the work of Momany et al. with maltodecaose [170]. All studies find that the conformation of the saccharides are intimately related to their ability to disrupt or restructure the surrounding water, and stabilized by bridging water hydrogen bonding.

Most recently a series of structural amylopectin motifs have been studied [171] by dynamic light scattering, where it was concluded that saccharides containing $\alpha(1\rightarrow6)$ glucosidic bonds tend to fold into compact structures, and that more than 5 glucose units in a structure causes it to prefer helical-like conformations.

To provide a continued systematic base of knowledge on starch structure - relating both the branchpoint and amylose features - we have constructed the natural follow-up on the maltose, iso-maltose studies by Corzana et al. [86] and Best et al. [166], by synthesis, NMR analysis and modeling of all the different trisaccharides that can be found in starch. While the glucosidic bonds are well characterized, the addition of an extra glucose ring opens up for the possibility of bend structures as a function of inter-ring interactions, like in the case of the pentasaccharide branchpoint model [84], where bridging water molecule was found to play an important role in the structuring of the solute.

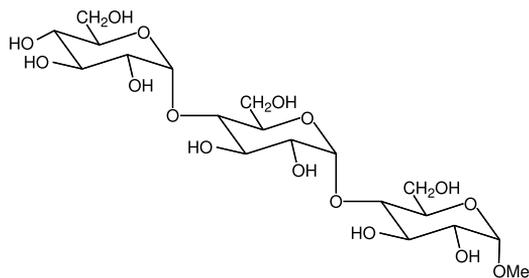


Figure 5.2 Chemical structure of the exclusively $\alpha(1\rightarrow4)$ linked tri-saccharide maltotriose

The placement of these structural motifs in the amylopectin branchpoint is illustrated schematically in **Figure 5.1**. All of the five trisaccharides are methylated in the C-1 position to prevent mutarotation.

5.1 The five starch trisaccharides

Clearly the most abundant tri-saccharide present in starch is maltotriose. As illustrated in **Figure 5.2** this trimer is built using $\alpha(1\rightarrow4)$ linkages exclusively and is, next to maltose, the primary model compound for amylose.

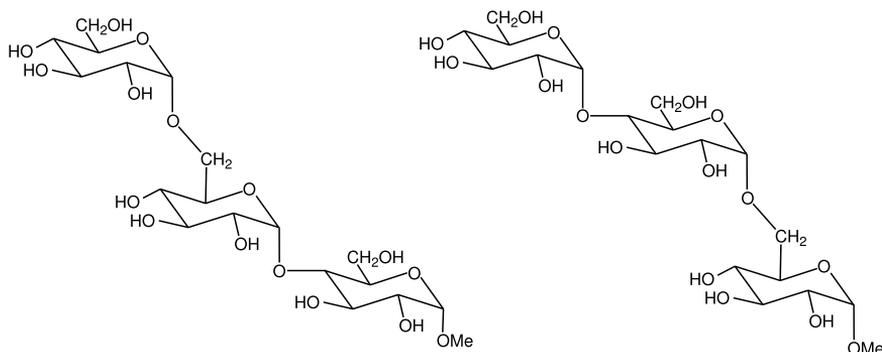


Figure 5.3 Chemical structure of the two panoses, α -panose (left) and iso-panose (right). Both saccharides contain one $\alpha(1\rightarrow4)$ and one $\alpha(1\rightarrow6)$ glucosidic linkage, but in reverse order

The chemical composition of maltotriose was first elucidated by Thompson et al. [172], and a crystal structure - as mentioned in Section 4.1 has been solved by X-ray crystallography by Pangborn et al. [119]. Most recently a DFT study by Schnupf et al. [173] have indicated that solvation plays a decisive role in the structural arrangement of maltotriose.

The structure of α -panose is the most studied of the trisaccharides here, both by X-ray diffraction [120–121], and also by simulation methods [168]. The saccharide iso-panose, contain the basic structure linking the amylose side chain to the 6-position in the amylopectin branchpoint.

Whether or not iso-maltotriose exists in the amorphous linker regions of amylopectin is uncertain as demonstrated on maize starch by Kainuma and French [174], and in rice starch by Umeki and Yamamota [175]. Still, it was included in the study for the sake of completeness. If not present in starch, it is a central structural component of Dextran, an important blood anti-coagulation agent. Compared to starch, the structural buildup of Dextran is reverse in the way that the backbone is build using $\alpha(1\rightarrow6)$ glucosidic bonds, which is then branched by $\alpha(1\rightarrow4)$ glucosidic bonds.

The saccharide illustrated in the right side of **Figure 5.4** is the central branching motif in amylopectin. It is the only structure in this study which is not built sequentially, but has a central glucose unit which is the reducing end since it is the endpoint of two linkages. Being the structure that splits or ‘forks’ amylopectin into amylose strands that are able to curl into double

helices, and lacking a suitable trivial name for the molecule, it was given the name α -forkose.

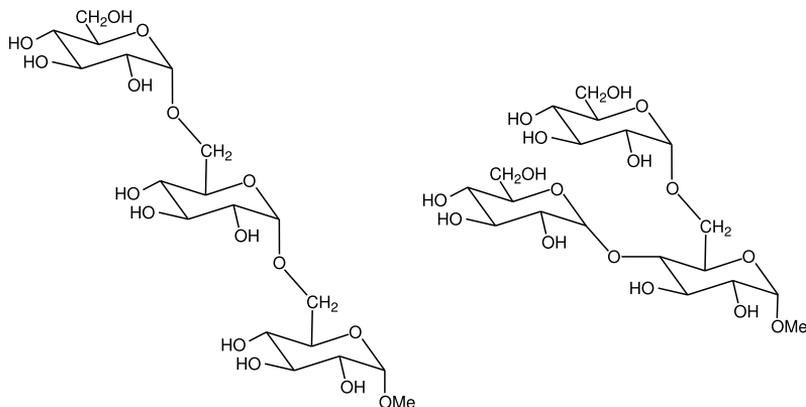


Figure 5.4 Chemical structure of the trisaccharide iso-maltotriose (left) arguably present in the amorphous parts of starch and common in glycogen, and the central branch point structure α -forkose (right)

Starting structures for all five trisaccharides were built with the polysaccharide builder program POLYS [176] using monomer units from the MONOBANK [177] database with the addition of a methylated glucose unit.

5.2 Trajectory analysis

For each of the five trisaccharides a 15 nanosecond molecular dynamics trajectory was calculated. From the analysis of the data a number of structural properties was calculated, among these are conformational distributions, diffusional properties, molecular flexibility, and finally several structural water molecules were found.

5.2.1 Conformational distribution

To quantify the preferred conformations of the trisaccharides, time series of Φ , Ψ , and ω dihedral angles, as defined in Section 2.3 was calculated from

the trajectories, assuming that length of the simulations provides a good representation of equilibrium between structural populations.

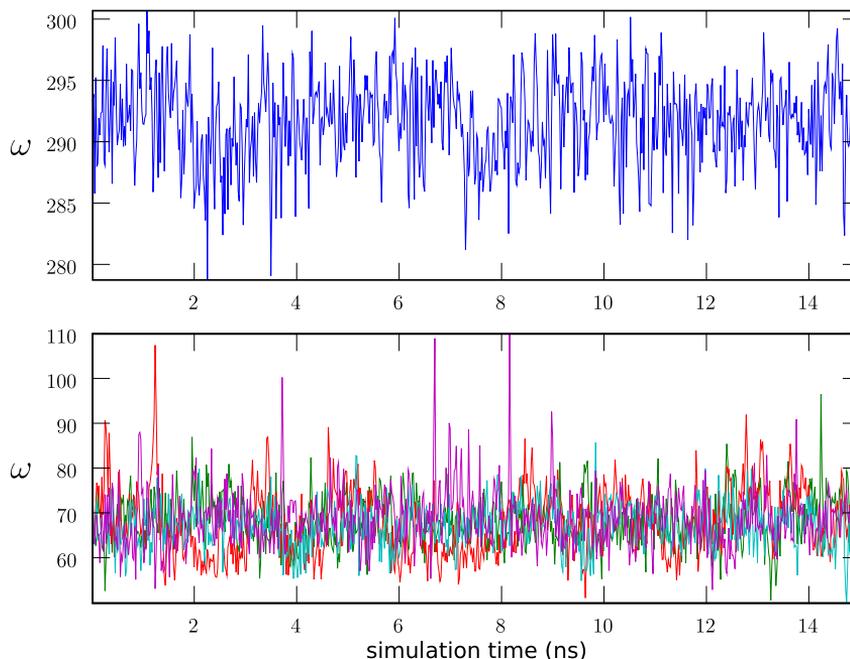


Figure 5.5 The ω dihedral angle as a function of time for the trisaccharides described in Section 5.1. (—) iso-panose (top), (—) α -panose, (—) iso-maltotriose link to non-reducing end, (—) iso-maltotriose link to reducing end, (—) α -forkose

From the dihedral angle analysis it is observed that the ω dihedral angle is stable for the five $\alpha(1\rightarrow6)$ linkages throughout the entire time series. As illustrated in **Figure 5.5** the ω dihedral is most flexible in iso-panose, where it exhibits a 20° variation around the *tg* conformation. For the remaining four linkages, the ω dihedral is always in the *gg* conformation, with only 10° variation.

The inflexible nature of the ω dihedral lead to a change of focus to the Φ and Ψ dihedrals, of which time series were also calculated. These data were then transformed into population density maps corresponding to each glucosidic linkage. The result is a 3D landscape as shown in **Figure 5.6**.

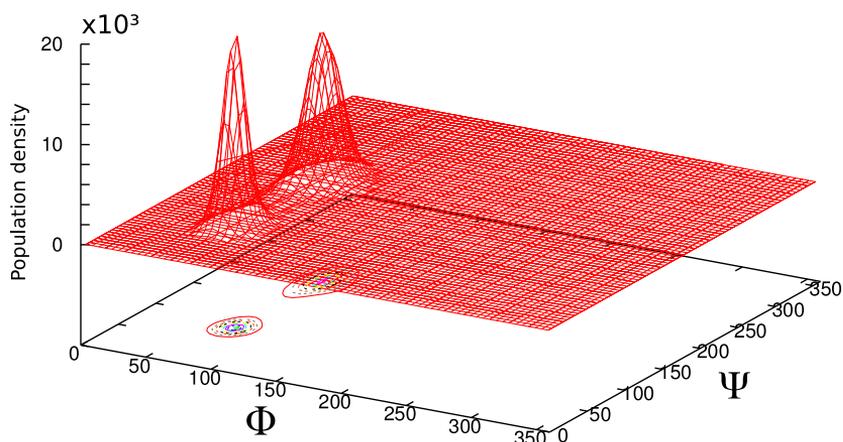


Figure 5.6 Population density map for the $\alpha(1\rightarrow6)$ linkage to the non-reducing end in *iso*-maltotriose, as a function of the Φ and Ψ dihedral angles. As the third dihedral was found to be stable around $\omega \approx 300^\circ$ this plot describes the conformational space of this linkage

In Paper II contoured versions of the population density maps is compared to the energetically preferred minima in the adiabatic map of maltose, and there is overall excellent agreement with the simulation data. The population density results from molecular dynamics is plotted in **Figure 5.7** on top of the adiabatic map from **Figure 3.1** together with the literature results on single crystals.

It is observed that the $\alpha(1\rightarrow4)$ linkage is relatively flexible. In most cases variation of $50\text{--}60^\circ$ is allowed for both dihedrals, except for α -panose which is more rigid. In several cases the conformational distribution is split into two close local minima. However, none of the linkages jumps to another minimum on the adiabatic map.

Looking at the results for the $\alpha(1\rightarrow6)$ glucosidic bonds, *iso*-panose is behaving differently by having $\omega \approx 300^\circ$, but still in Φ, Ψ space the conformation is relatively stable around $(70^\circ, 187^\circ)$. Maltotriose has no $\alpha(1\rightarrow6)$ bond, and the result for the remaining saccharides are plotted in **Figure 5.8**. Little variation, 30° or less, of the Φ dihedral is observed for the $\alpha(1\rightarrow6)$ glucosidic bonds. The Ψ dihedral, on the other hand, is more flexible. For

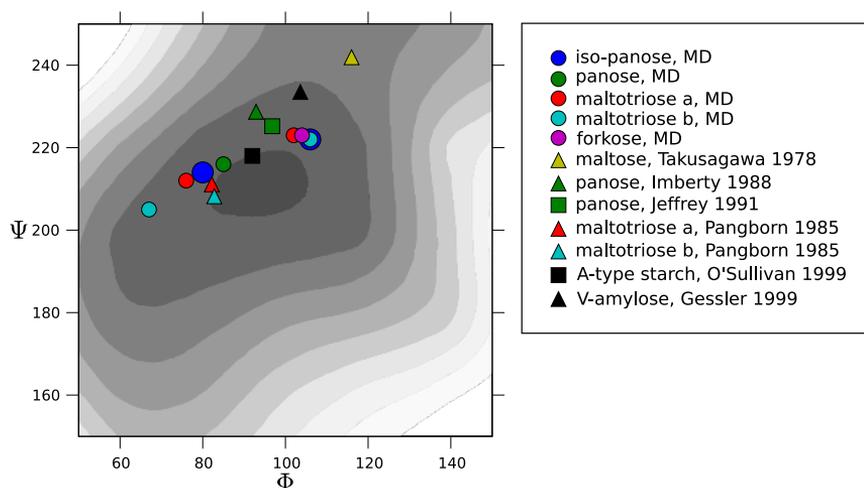


Figure 5.7 Primary Φ, Ψ conformations of all $\alpha(1\rightarrow4)$ linkages from the molecular dynamics runs (circles). The markers for iso-panose are enlarged to reveal overlap. Additional data is included from single crystal studies [26, 115, 119–121, 123–124]. The background adiabatic map is identical to **Figure 3.1**

this dihedral variation of up to 60° is observed where $\Psi \approx 190^\circ$ and in both iso-maltotriose and α -forkose this conformation also jumps between this *gt*-like conformation, and a *gg*-like conformation where $\Psi \approx 70^\circ$. As a special case, the linkage between the non-reducing end and the central glucose unit in iso-panose, the *gg* conformation is preferred.

The conclusion drawn from these conformational data must be that the geometry of iso-panose fits perfectly well with the adiabatic map, and therefore is either unaffected by the explicit solvent, or that this particular conformation is stabilized by the water. Also α -panose is in a stable overall conformation which is in good agreement with both X-ray data [120–121], and the adiabatic map. However, in contrast to iso-panose, it exhibits increased flexibility in the Ψ angle. For maltotriose, four closely related conformers are possible by the fine splitting of the maximum populations in both linkages, which is different from the crystal structure [119] where the geometry of the two linkages are almost the same. In iso-maltotriose, four very different conformers are made possible by the previously mentioned

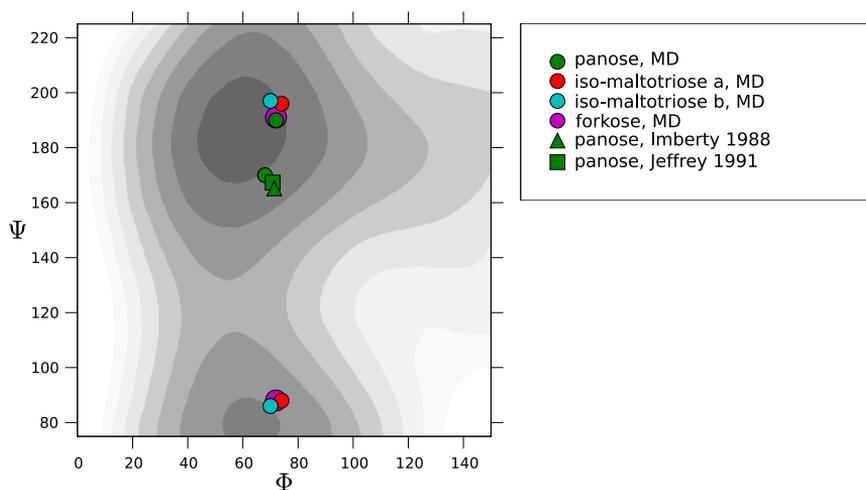


Figure 5.8 Primary Φ, Ψ conformations of all $\alpha(1\rightarrow6)$ linkages from the molecular dynamics runs (circles) on the adiabatic map corresponding to the case where $\omega \approx 60^\circ$. The markers for α -forkose are enlarged to reveal overlap. Additional data is included from single crystal studies [120–121]

jump between different Ψ dihedrals. Similar to this, two conformations are populated for the $\alpha(1\rightarrow6)$ linkage in α -forkose, with a strong preference for $\Psi = 191^\circ$.

5.2.2 Molecular flexibility

The distance between ring oxygen atoms in each end of the molecules, that is from O5 to O5'', was calculated for all time frames in the simulations, in order to study the effect of the conformational dynamics of the molecular extension⁶. The result of this analysis is illustrated in **Figure 5.9**.

The O5-O5'' distance information depends both on the flexibility of the entire molecule, but it is also sensitive to rotation of the end ring independently of the remaining part of the molecule. The dynamic behavior of the

⁶ By the defined notation O5 is placed in the non-reducing end and O5'' in the reducing end of the saccharide. Since α -forkose has two non-reducing ends, the non-reducing end of the $\alpha(1\rightarrow4)$ link will be defined as having no prime and the other non-reducing ring as double prime.

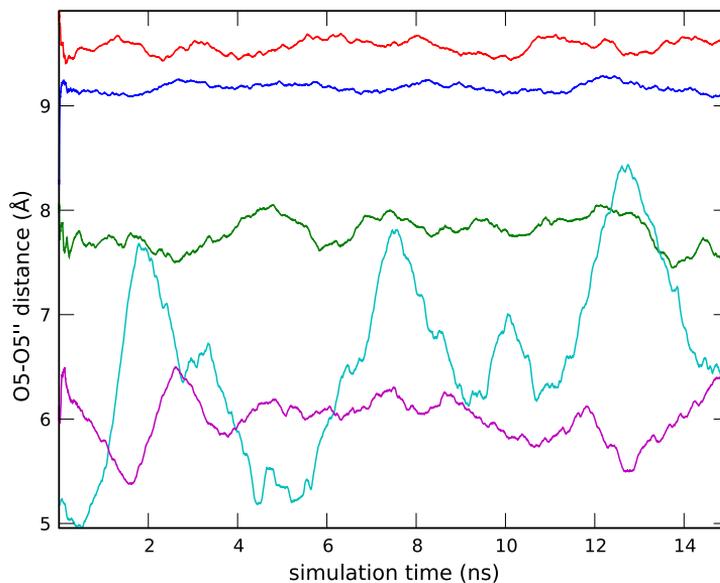


Figure 5.9 Time evolution of the O5-O5'' distance, for each of the trisaccharides described in Section 5.1, showing the differences in overall flexibility. (—) iso-panose, (—) α -panose, (—) maltotriose, (—) iso-maltotriose, (—) α -forkose.

five trisaccharides in this respect is very different. As seen from the data in Section 5.2.1 iso-panose, α -panose, and maltotriose is conformationally rather rigid, which also clearly shows in the extension. The jump between local minima in the maltotriose conformation explains the small fluctuations in this distance, as is also the case for α -forkose. The most dramatic behavior is observed for iso-maltotriose, showing large jumps in molecular extension. This corresponds to the large separation in conformational populations visited, and four different conformational levels can be observed in the O5-O5'' distance plot.

5.2.3 Structural water molecules

To conduct an in depth examination of the role that water plays in the structuring of the starch tri-saccharides, two dimensional water distribu-

tions were calculated for the complete list of solute oxygen-oxygen combinations using the 2D pair distribution principles introduced in Section 3.3.1. The pair distributions was calculated for distances shorter than 3.5 Å from each oxygen, to include the first hydration shell, and resulting maximum densities larger than 1.0 considered as indicative of water binding sites. In saccharides, several highly populated water binding sites are always found between neighboring intra-ring oxygen atoms. What is more interesting is when inter-ring oxygen pairs can be identified as a water bridging site, as this puts a strong restraint on the structure of the molecule.

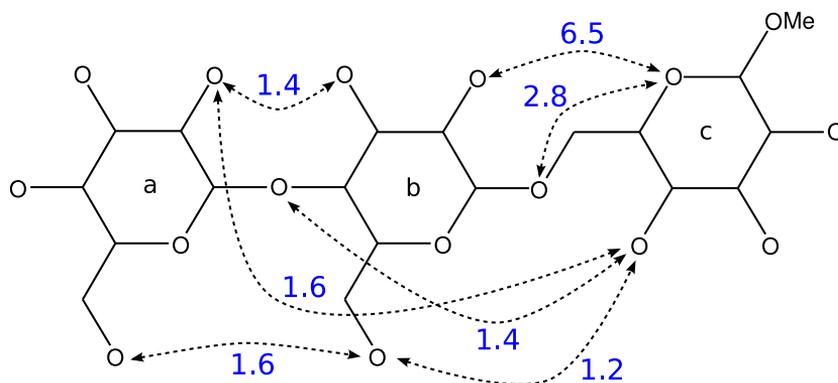


Figure 5.10 Seven significant localized water densities found between inter-ring oxygen atoms pairs in iso-panose. Densities are given relative to bulk water

Several significant water bridges were identified in the trisaccharide set. Most notably, a complex water binding network was found in iso-panose, which is illustrated in **Figure 5.10**. The density of 6.5 times that of bulk water, between O2' and O5'' is the highest one found in the five systems. The strong hydration of this site, and the additional bridging across this glucosidic linkage reflects the confined conformational space visited, as was seen in Section 5.2.1. This hydration site is in good agreement with the one found in a the larger penta-saccharide branch-point model by Corzana et al. [84].

A strong water bridge of density 3.0 was also found in α -panose, along with two weaker ones. The two maltotrioses was also found to be hydrated by at network of weak water bridges. Maltotriose, which was found to

be relatively rigid, is has a total of five inter-residue waterbridges. Three of them, all between the middle unit and the reducing end, has a density of 1.0 which must be considered as very weak. The remaining two has densities of 1.4 and 1.6 respectively and is bound to the pairs O6-O6' and O2-O3'. None of these hydration patterns is found in either the maltotriose crystal structure or the A- and B-type amylopectin crystals as described in Section 4.1.

While maltotriose was found to be conformationally stable, the similar weak hydration of the $\alpha(1\rightarrow6)$ bonds in iso-maltotriose are not strong enough to prevent the dihedral conformation, from changing between energetic minima in the adiabatic map. The existence of multiple stable overall conformations of iso-maltotriose, as indicated by the extension plot in **Figure 5.9**, is able to explain the two somewhat ambiguous bridging water molecules between non-reducing and reducing end of the tri-saccharide as illustrated in **Figure 5.11**, which are clearly not possible in the extended form of the molecule.

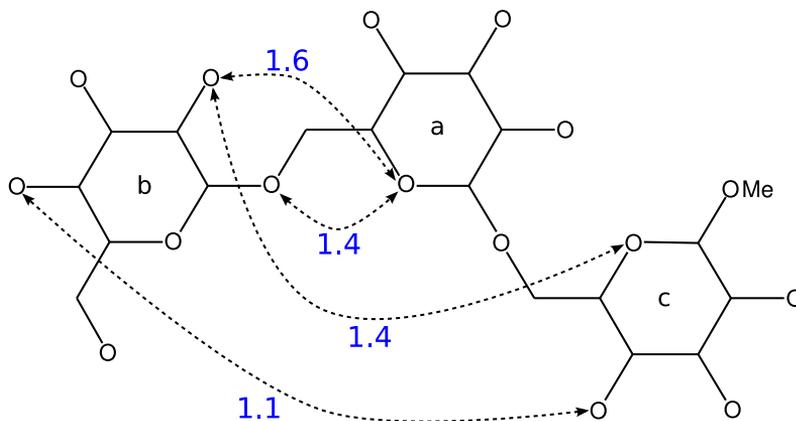


Figure 5.11 The four significant localized water densities found between inter-ring oxygen atoms pairs in iso-maltotriose. Densities are given relative to bulk water

Interestingly, the branch point saccharide, α -forkose is only significantly hydrated in a single inter-residue site from O2 to O3' with a density of 1.4. This type of bridging water is characteristic of the $\alpha(1\rightarrow4)$ bond as seen from the collected data in Paper II and also the results in hydrational

investigations of Corzana et al. [84, 86]. The apparent lack of hydrational influence on the structure of this saccharide indicate that primarily intramolecular are governing the dihedrals. This is in agreement with the tetra-saccharide branch point model by Best et al. [168].

5.3 Experimental results

To evaluate the above results from molecular dynamics, it is informative to compare the distribution of dihedral arrangements with structural data obtained by experiment, such as X-ray crystallography, as was done for α -panose and maltotriose in Section 5.2.1. This is, however, a limited comparison since there are no crystal structures available for the remaining three tri-saccharides. Furthermore, the differences in environments between the crystalline phase and explicitly solvated state as discussed in Section 4.1, add complexity to the interpretation of the results. This was seen for the structure of maltotriose where the simulated data predicts the conformations of the two glucosidic linkages to be clearly different (**Figure 5.7**). These conformations were found to be almost identical in the crystal structure [119]. In contrast, the $\alpha(1\rightarrow4)$ linkage in the crystal structure of α -panose is in reasonable agreement with the dynamics data. The Ψ angle of the $\alpha(1\rightarrow6)$, which has a wide distribution, also has a local minimum that match the crystal structure. The conclusion of such comparisons emphasize that the crystal structures are too static, and thus insufficient as the only reference points when it comes to dynamic systems which might have more than one favored stable conformational state.

To provide additional information on the conformational arrangements of the tri-saccharides, structural data in the form of hetero-nuclear scalar coupling constants across the glucosidic linkages was measured using NMR spectroscopy, according to the experimental strategy that was outlined in Section 4.2.

From the NMR spectra recorded on the five tri-saccharide samples, as described in Section 4.2.1, all proton and carbon resonances were successfully assigned. Combining this knowledge with the J-HMBC [144] spectra, it was possible to measure a complete set of ${}^3J_{CH}$ couplings for the linkages in all tri-saccharides.

Using the labeling scheme mentioned in section Section 4.2.1, the complete list of heteronuclear coupling constants in presented in Table 5.1.

iso-panose	${}^3J_{H1-C4}$	CSFF	${}^3J_{C1-H4'}$	CSFF	${}^3J_{H1-C6'}$	CSFF	${}^3J_{C1-H6'}$	CSFF
$a \rightarrow b$	4.0	3.8	4.9	4.1				
$b \rightarrow c$					3.7	2.7	3.4/3.4	1.0/4.6
α -panose								
$a \rightarrow c$	3.7	4.0	4.4	3.1				
$a \rightarrow b$					3.5	2.6	3.4/3.3	2.1/2.4
maltotriose								
$a \rightarrow c$	3.4	3.5	4.2	4.0				
$b \rightarrow a$	3.8	3.7	4.5	4.2				
iso-maltotriose								
$a \rightarrow c$					3.6	2.6	3.5/3.3	1.1/1.1
$a \rightarrow b$					3.5	2.7	3.3/3.4	3.9/2.1
α -forkose								
$a \rightarrow c$	3.9	3.9	4.7	4.1				
$a \rightarrow b$					3.6	2.7	3.7/3.7	2.9/1.8

Table 5.1 3-bond heteronuclear coupling constant in Hz from Paper II, comparing experimental values with all results from molecular dynamics calculations. For the $\alpha(1\rightarrow6)$ linkages the two H6 protons are listed in the same entry separated by a ‘slash’ character

Theoretical values of these coupling constants were calculated from each trajectory using the geometry-coupling relationship of Equation 4.1a. The calculation was performed by calculating the appropriate dihedral angle ϕ from the atomic coordinates at each time-step and converting it to the corresponding coupling constant. This yields the time dependent evolution of the coupling from which a time average can be calculated and subsequently be compared to the experimental value.

From the set of coupling constant data published in Paper II, there was found a good agreement between theory and experiment in most cases where the glucose units are connected by an $\alpha(1\rightarrow4)$ linkage. Interestingly, the data contradicts the previously mentioned crystal structure of maltotriose, and clearly shows - in support of the simulated data - that the two linkages assume different conformations. In contrast, the agreement

with experiment is consistently poor for all $\alpha(1\rightarrow6)$ linkages. This was also the experience by Corzana et al. [86], who similarly found that molecular dynamics was not able to reproduce the measure values, and that this was the case for three different carbohydrate force fields. Apparently there is a problem with describing this structural element, by either the force field parametrization of this type of linkage, or alternatively by the applied Karplus relation which may need to be re-parametrized.

Some supporting information was obtained from the NOESY spectra, in form of inter residue cross peaks. This information must be regarded as qualitative, since the resolvable intra residue NOEs were not consistent in intensity, and no NOE buildup rate measurements were performed.

For iso-panose, three NOEs was found for both linkages. Two of them are strong in intensity, from H1 to H4' and H1' to H6". These NOEs confirm results from the conformational and hydrational analysis in that this structure is stabilized. The same is observed for the $\alpha(1\rightarrow4)$ linkage in α -panose with a strong NOE between H1' and H4". In contrast the two weak NOEs across the $\alpha(1\rightarrow6)$ bond support the increased flexibility of this linkage. For maltotriose, only a single medium intensity NOE was found between H1' and H3" which again supports that this linkage is different than the other one in this molecule when observed in solution.

Also in α -forkose NOEs was found to be correlating the H1 and H4' protons across the $\alpha(1\rightarrow4)$, which was stabilized by a water bridge. Interestingly, a strong NOE was found between H1" and H6' which was not hydrated, implying that intra-molecular forces alone stabilize this structure. A weak NOE was further observed between H3 and H1" which from distance calculation on the conformational population maxima can be attributed to one of the rarely populated structures. This NOE was the only one in the entire study which is a correlation between the end rings of a tri-saccharide. Such a distance constraint is as important as the strong ones observed since it is not originating from a proton pair that is close in space by default - as with the ones directly across linkages - but require a folding of the molecule which is not necessarily to be expected. One of the stable structures found in the conformational analysis in Section 5.2.1, which is illustrated in **Figure 5.12**, is able to explain the observed weak NOE, even though it is not the most populated conformation.

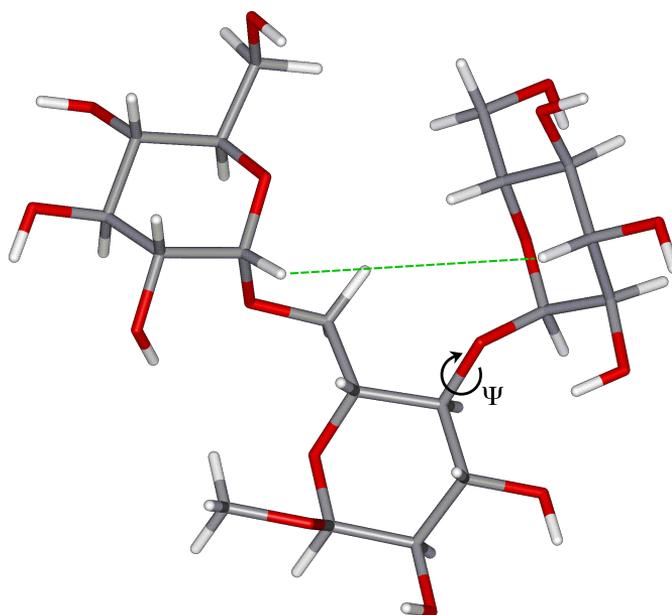


Figure 5.12 The structure of α -forkose in the least populated of the two main conformations where $\Psi = 88^\circ$. In this structure H3 and H1'' comes closer than 5 Å which explains the observed weak NOE between the two hydrogen atoms.

6 Starch phosphorylation

Besides the amylose and amylopectin structural motifs generated by combinations of the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ glucosidic bonds, another, less common, chemical modification of the glucose unit is known to occur during the synthesis and degradation of starch in the plants. That modification is the condensation reaction of phosphoric acid with the saccharide into a mono-ester.

The phosphate groups, which are covalently bound to starch, has been shown to be primarily located in the C-3 and C-6 positions [178–179], with approximately 30% in the C-3 position and 70% in the C-6 position. From recent work, using ^{31}P -NMR spectroscopy and ^{33}P radioactive labeling, it has been established that the two enzymes Glucan-water-dikinase and Phosphoglucan-water-dikinase are the specific catalysts for C-6 and C-3 phosphorylation respectively [180]. It has been suggested that 1% of the phosphate groups in potato starch is bound in the C-2 position [181], but this is still debated in the literature. However, if this substitution exist, the concentration is so low that we cannot study it by NMR spectroscopy. Recent work by Blennow et al. [182] has indicated a relationship between starch phosphorylation and a restructuring of the starch granules that changes the degradability of the granules, and evidence that links starch phosphorylation with the activity of degradation enzymes has been published recently by Edner et al. [183]. Phosphorylation of starch has multiple applications as phosphorylation can induce large changes of the physical properties of starch. Chemically phosphorylated starches are currently used in a variety of industrial contexts, for example as a component in the paper-making process [184], in the area of water purification [185], and as a biodegradable complex binder of heavy

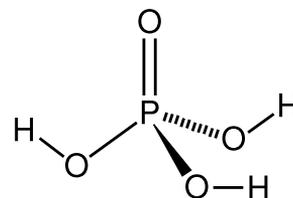


Figure 6.1 Chemical structure of phosphoric acid in its fully protonated state

metal cations. Traditionally, phosphorylated starches have been manufactured by chemical processing⁷ where the starch is treated with phosphoric acid and urea. After the discovery of the enzymes that regulate the natural phosphorylation process in the plants, it has become possible to replace some of the traditional chemical processes with alternative enzyme technologies that are more friendly to the environment. An advantage of this, is a more specific control of the substitution reaction, and thereby the functional properties of the starch.

In the food industry, where - as described in Section 2.1 - the occurrence of starch as a functional ingredient is vast, new applications are emerging of for example genetically engineered hydrocolloids with improved properties such as higher gelling temperatures and stronger textures of the gels [186]. Potato starch is known to have a higher content of covalent phosphate than found in other plants, and recent investigations of the physical chemical properties of potato starch granules of varying size [14], showed that the smaller granules have a larger concentration of phosphate groups and bind more cations, have a higher viscosity and are more easily digested by glucoamylases. These properties have made potatoes an attractive subject for the exploration of starch phosphorylation, specially in relation to the crystalline parts of the granules [19, 187–188], which will also be the primary reference structure in the rest of this chapter.

Phosphorylation of glucose-6-phosphate has previously been investigated by a combined x-ray and molecular dynamics study by Engelsen et al. [189], where the phosphate group was found to be able to fit into natural cavities when imposed on the amylopectin helical structure. In the light of the importance of hydration on the glucosidic linkage structure, two model compounds were designed in order to study the effect of the two types of phosphorylation on the $\alpha(1\rightarrow4)$ linkage, which is the subject of the work in Paper IV.

⁷ A special feature of the chemical phosphorylation of starch, is the ability to create di- and tri-phosphate esters, in addition to one the mono-esters found in plant starch. Another is cross-linked phosphates, which is also not observed in natural starches.

6.1 Phosphorylation models

The two models synthesized for the combined molecular dynamics and NMR spectroscopic study, is maltose-3-phosphate and maltose-6-phosphate.

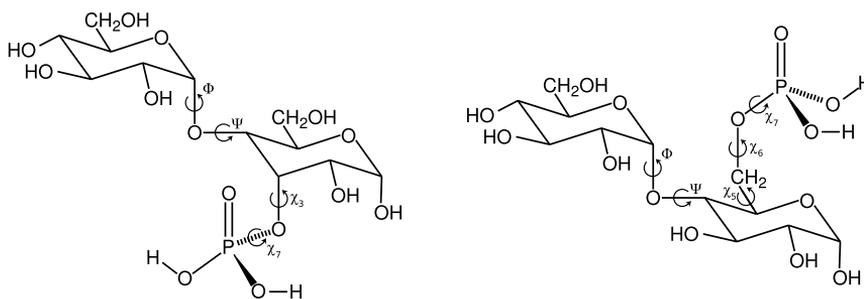


Figure 6.2 Chemical structure of the maltose-3-phosphate (left) and maltose-6-phosphate (right) molecule including definitions of dihedral angles, both for the $\alpha(1\rightarrow4)$ linkage and the exocyclic angles giving the conformation of the phosphate group

The chemical structure of the two model compounds are illustrated in **Figure 6.2**, along with the definitions of both the dihedral angles in the glucosidic bond but also the dihedrals needed to describe the conformation of the phosphate groups which will be the topic of some of the following discussions. The drawings are of the phosphorylated maltoses in their fully protonated forms. Depending on the pH of the aqueous environment, the phosphate group will deprotonate according to the acidity of the protons.

After the phosphate mono-ester condensation reaction the disassociation constants for the remaining labile protons are changed from the ones for phosphoric acid. pKa values that should be similar to the ones in the phosphorylated maltoses are known for the corresponding phosphorylations of glucose. These are given in Table 6.1, and can be used to assess the degree of protonation of the molecules by a given pH value. The full picture of the possible protonation states as a function of the pH is given by the Bjerrum diagram **Figure 6.3** defined by the pKa values, and

	pKa ₁	pKa ₂
glucose-3-phosphate	0.84	5.67
glucose-6-phosphate	0.94	6.11

Table 6.1 Proton disassociation constants for 3- and 6-phosphorylated glucose.

from this it is clear that around the neutral $\text{pH}=7$, the phosphate groups is almost completely deprotonated in both molecules.

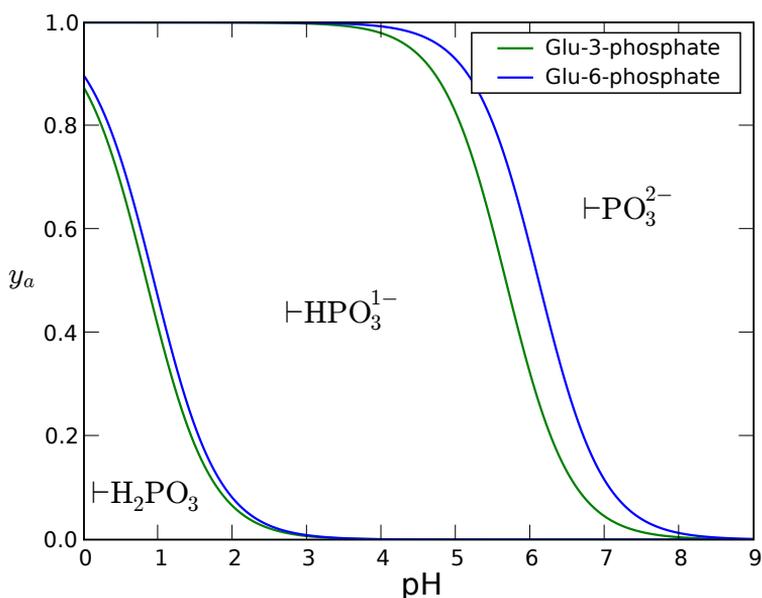


Figure 6.3 Bjerrum diagram of the protonation states of phosphorylated glucose, showing the fraction of acid y_a as a function of pH. The symbol \vdash represents the remaining part of the phosphorylated molecule

6.2 Molecular dynamics

Following the considerations about the protonation states of the phosphorylated maltoses, the conclusion must be that they are most likely completely deprotonated, which means they have two negative charges. However, since biological systems or food matrices with a pH lower than neutral do exist, all protonation states are taken into account in the subsequent molecular dynamics simulations. The two phosphorylated disaccharides were built using the POLYS program [176] using a custom built phosphorylated glucose unit.

6.2.1 Force-field parameters

Using molecular dynamics calculations to study phosphorylated carbohydrates, has only been done on a few occasions. Thus, the force-field parameters for the phosphate group attached to a sugar-ring have never been thoroughly optimized for this system, like they have for eg. the nucleic acids. A couple of different approaches to adapt parameters from other force fields has been demonstrated in the literature, in both cases from the area of nucleic acids. An example is the study of 6-phosphorylated α -glucose by Engelsen et al. [189], where a set of force-field parameters was adopted from the work Dupré et al. [190], who parametrized the phosphodiester linkage.

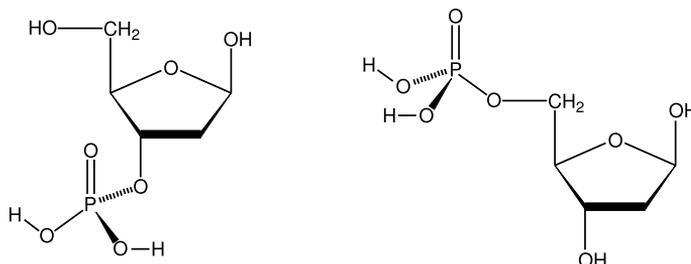


Figure 6.4 Chemical structure of deoxy-ribose-3'-phosphate (left). and deoxy-ribose-5'-phosphate (right)

Another adaptation of phosphate parameters was done by Crouzy et al. [191] who used the part of the CHARMM nucleic acid parameters for phosphate bound to the nucleotide terminals. For the work presented in Paper IV, the latter approach was chosen and the relevant force field parameters was merged with the CSFF [62] force field. Phosphorus is a much more abundant element in nucleic acids, whose backbone is linked by phosphodiester, and the phosphate mono-ester is present at the 3' and 5' ends, where it is attached to the deoxyribose 5-ring structure. From the molecular structures illustrated in **Figure 6.4** it is easily seen, when compared to **Figure 6.2**, that there is a close structural resemblance to the glucose ring structure, and that the substitution sites closely resembles the ones found in the natural starches. To verify the similarity of the chemical environments of the deoxyribose terminals to the corresponding ones in glucose, *ab initio*

calculations were used to provide grounds of comparison between the electronic distribution in the two ring systems. The calculations were performed on simplified models built from a glucose ring which was phosphorylated at either the 3- or 6- position by substitution of the hydroxyl proton by the phosphate group. To perform the quantum mechanical calculations, the software suite MPQC [192] was used at the B3LYP/6-31G* level of theory which is used extensively in the parametrization of the CHARMM22 force field [68] that is the basis for the CSFF force field. The charges of the atoms was found to be very similar in the 5- and 6-ring systems, and only small adjustments were required to obtain residues that are neutral, or has an integral charge of -1 or -2.

6.3 Conformational analysis

To get an initial impression of the effect of the phosphorylations on the linkage structure, a set of conformational maps were calculated using the high temperature dynamics procedure outlined in Section 3.1.5. These are illustrated only for the neutral systems in **Figure 6.5**.

From the figure, it is observed that the maltose-3-phosphate prefers a single most stable conformation at ($\Phi = 70^\circ, \Psi = 245^\circ$), which is far from the Φ dihedral in the crystal structure of α -maltose where ($\Phi = 116^\circ, \Psi = 242^\circ$). For maltose-6-phosphate the conformational map bears a closer resemblance to the adiabatic map in **Figure 3.1** visiting two conformation but with a preference for ($\Phi = 88^\circ, \Psi = 197^\circ$).

Conventional molecular dynamics simulations was also run for these system including explicit solvation in form of TIP3P water molecules and at 300 K. The result of these simulations were also very similar across the possible degrees of protonation, with only slight variations. The population maximum for maltose-3-phosphate in this case is ($\Phi = 70^\circ, \Psi = 200^\circ$) and for maltose-6-phosphate two closely related conformations ($\Phi = 65^\circ, \Psi = 200^\circ$) and ($\Phi = 100^\circ, \Psi = 210^\circ$) are obtained, with a preference towards the latter as protons are lost from the phosphate group. The population distributions for the fully deprotonated forms are plotted in **Figure 6.6** and **Figure 6.6** together with an indication of the α -maltose crystal structure and helical parameters for different constructions of an amylopectin double helix. This

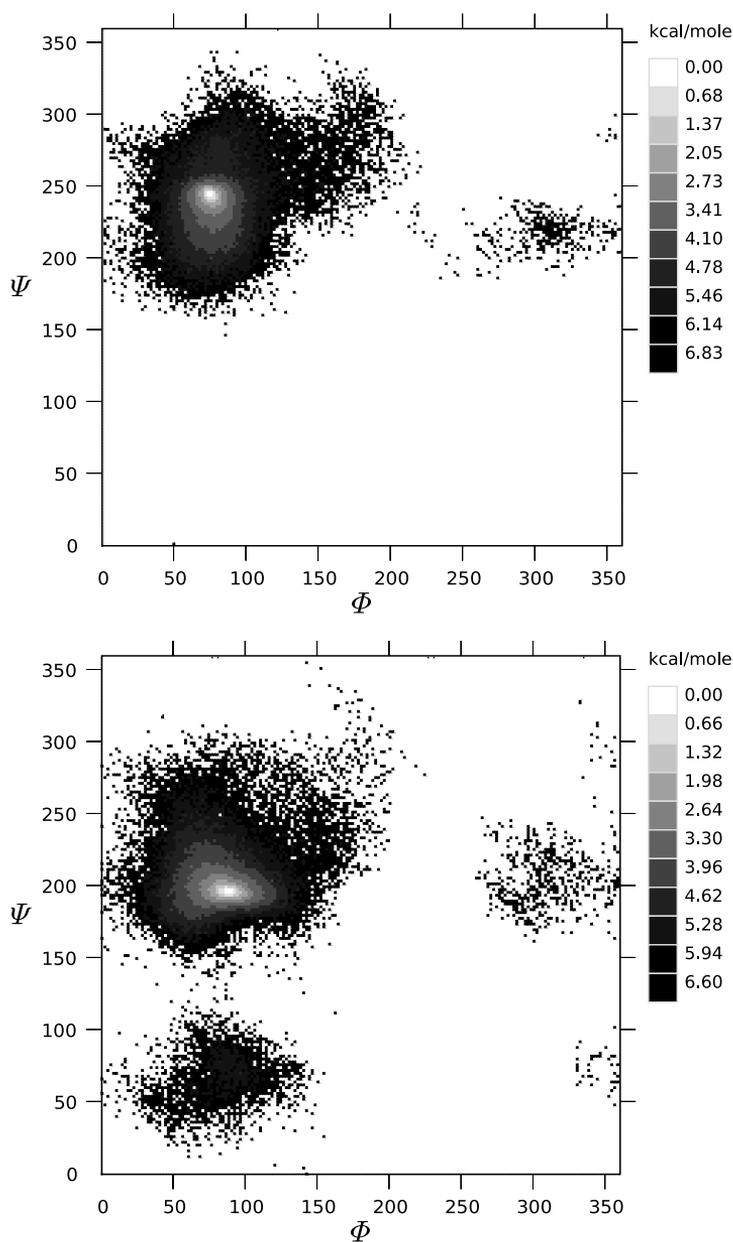


Figure 6.5 Conformational maps of maltose-3-phosphate (top) and maltose-6-phosphate (bottom), resulting from molecular dynamics simulation at 1000 K, using a continuous solvation model

type of plot has previously been used by Pérez et al. [26, 193] to conclude that the optimal structure of the unmodified double helix is a 6-fold left-handed helix with a helical pitch of $h = 3.5 \text{ \AA}$.

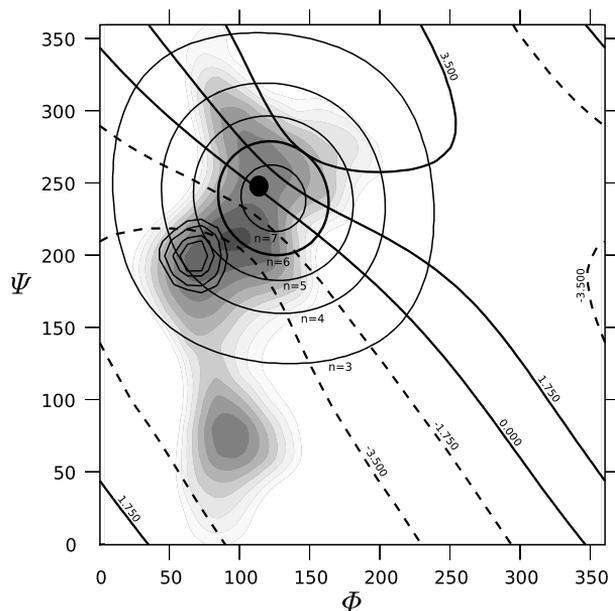


Figure 6.6 Adiabatic map for the $\alpha(1\rightarrow4)$ linkage in maltose as background for double helical amylopectin parameters. The contour lines with floating point values are the helical rise h , per residue, in Angstrom parallel to the helix axis. The values for h are positive for right hand helices and negative for left handed. The integral valued contours are the number of residues per helical turn. Superimposed on top are the conformational distribution found for maltose-3-phosphate. The maltose crystal structure [115] is indicated with a black dot. The figure is inspired by Blackwell et al. [28]

By combining the conformational analysis with the adiabatic map and what is known about the optimal structure of the amylopectin double helix, it

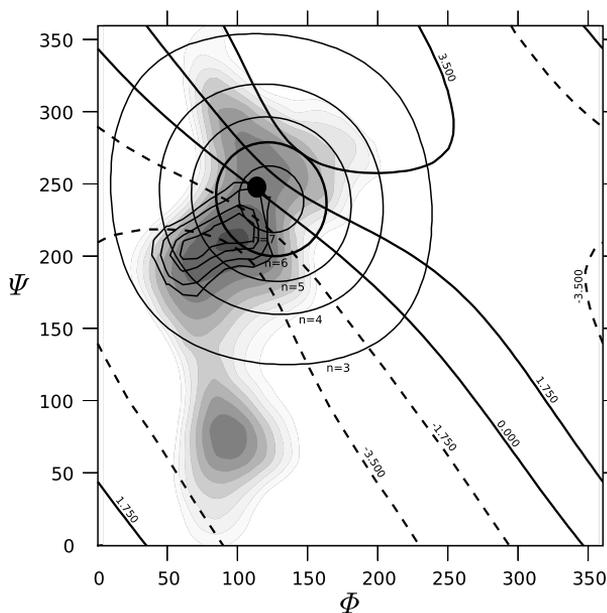


Figure 6.7 Adiabatic map of the maltose dihedral angles, and amylopectin helical parameter plotted by the principles given for **Figure 6.6**. Superimposed on top are the conformational distribution found for maltose-6-phosphate.

must be concluded that the conformational change that is induced by phosphorylation in the 3-position does not conform with the helical structure. In the case of 6-phosphorylation, the preferred conformer of the deprotonated form fits very well with the helical optimum. For both substitutions, the orientation of the phosphate group was found align well in the openings of a double helix if built onto this.

6.4 Hydrational analysis

A complete search for bridging water molecules was conducted in a similar manner as the work in Paper II; by calculation of all possible 2-dimensional radial pair distributions. The phosphorylated maltoses did not exhibit the same amount water sharing across the glucosidic linkages, as was found for the trisaccharides but instead showed a hydrational pattern similar to what

has been observed for α -maltose [86, 164–165]. In maltose-3-phosphate, the most significant inter-residue water-bridge was found between O2 and O3' with a probability density of 1.1 which is weaker than the 1.4 expected from α -maltose, and for the 6-phosphorylated maltose, no such bridge was found.

To further investigate the hydration of the solutes, all maxima from radial pair distributions were compiled into hydrational map, showing hydrational interactions between all oxygen atoms in the saccharides. An example of such a map is given in **Figure 6.8** for maltose-6-phosphate at neutral pH.

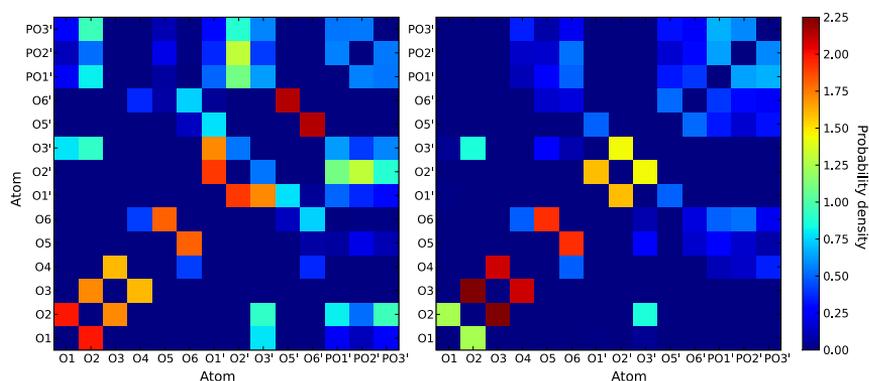


Figure 6.8 Hydrational map of the fully deprotonated forms of maltose-3-phosphate (left) and maltose-6-phosphate (right). Every square represents the maximum value of the two dimensional radial pair distribution function corresponding, within 3.5 Å, for the two coordinate oxygen atoms.

With all the pair distribution data visible in this way, it is clear that there exist some collective interaction between the three oxygen atoms of the phosphate group and different sites elsewhere in the maltose. Specially for maltose-3-phosphate the inter residue interactions between any phosphate oxygen and O2 are 0.8, 0.5, and 1.0 respectively, which together adds up to a significant stabilization.

6.5 Experimental data

Following the same procedures as in the trisaccharide study, a complete assignment of proton and carbon resonances was obtained. The assign-

ments were compared to literature data measured on maltose by Hoffman et al. [194], with only a few differences that could easily be explained by the structural changes by phosphorylation, which causes a downfield shift of the resonance of the substituted carbon and the attached proton. Likewise, a difference was found for the methylation of maltose by Corzana et al. [86], which causes an upfield shift for H1' and C1'.

A set of ${}^3J_{CH}$ scalar coupling constants was measured and found to be comparable to previous reports for maltose by Parfondry et al. [133]. In that study they found scalar coupling of 3.5 Hz or less across the glucosidic linkage. In the obtained data from Paper IV, the ${}^3J_{H1-C4'}$ coupling is measured to be 4.6 Hz, which is higher than the 3.5 Hz from literature. This high value does not agree with the molecular dynamics simulation. However, a similar difference between experiment and theory was also reported by Corzana et al. [86], which indicates that the phosphorylation is not the cause of this. For both the measurements in Paper IV and the ones from Corzana et al. the coupling constants was calculated using Equation 4.1. By comparison with theoretical values calculated from all the molecular dynamics trajectories, the best agreement was found when using the explicit solvation model, and when the maltoses are deprotonated. This supports the validity of the force field parameters for the phosphate group, and the hypothesis that water is important for the conformation of the saccharide. A summary of the measured and calculated coupling constants is presented in **Table 6.2**, along with relevant values from published studies of maltose.

From the obtained ROESY spectra, very few NOEs could be assigned, which correlated protons across the $\alpha(1\rightarrow4)$ bond. Only a single NOE was observed between H1 and H4' for both solutes. The cross peak was found to be of medium intensity for maltose-3-phosphate and is strong for maltose-6-phosphate. This indicates that the two protons are closer in space for maltose-6-phosphate than for maltose-3-phosphate, which is also observed in the molecular dynamics trajectory.

Maltose-3-phosphate	${}^3J_{H1-C4'}$	${}^3J_{C1-H4'}$
Experimental	2.3	3.4
CSFF-GBSA	3.0	5.5
CSFF-TIP3P charge 0	2.5	3.1
CSFF-TIP3P charge -1	2.5	3.1
CSFF-TIP3P charge -2	2.3	3.0
Maltose-6-phosphate		
Experimental	4.6	3.2
CSFF-GBSA	4.1	2.9
CSFF-TIP3P charge 0	3.2	3.9
CSFF-TIP3P charge -1	3.3	3.9
CSFF-TIP3P charge -2	3.5	4.0
Maltose		
Experimental, Corzana et al. [86]	4.1	4.9
CSFF-TIP3P, Corzana et al. [86]	3.4	3.9
X-ray Takusagawa et al. [115]	5.6	5.6
MM3 Dowd et al. [85]	5.0	5.2

Table 6.2 3-bond heteronuclear coupling constant in Hz, comparing experimental values with all results from calculation in Paper IV. Previous results for maltose from literature are included as references

7 Thesis out-look and perspectives

The next natural step to follow this study is to fully characterize all possible glucose tetramers, to investigate if the fourth glucose unit adds some special functionality which is not present in any of the trisaccharides. Even more interesting would be to pursue our original goal and study a model large enough to form a double helix structure *in vitro*.

The initial aim of this project was to explore the tetradecamer of the amylopectin double helical structure, as illustrated in **Figure 7.1**, by means of molecular dynamics simulations and NMR spectroscopic studies.

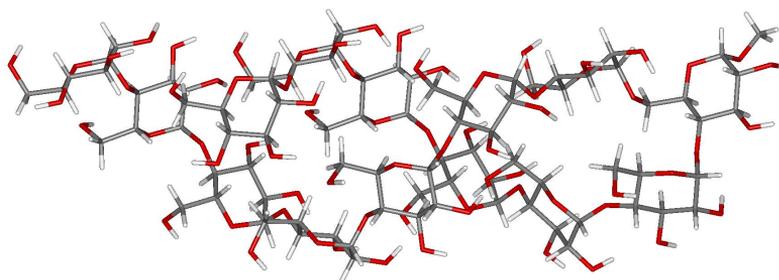


Figure 7.1 Model of the amylopectin double helix, including the branchpoint, built from 14 glucose units. The two side chains of the model is long enough to describe a full turn of the double helix

Large parts of the modeling work was performed, but since the synthesis of the molecule yielded too little material to record useful NMR data, we turned the focus toward simpler systems.

From the molecular dynamics run of the double helix model in explicit water, several interesting features can be observed. For example, the helical structure is still intact after 10 ns, with some motions of the non-reducing units at the end of the helix. Calculations of the excluded volume of this molecule compared to an unfolded equivalent, indicate that despite the absence of water molecules in the center of the helix, the unfolded molecule actually occupies a larger volume. This in turn indicate that the folded structure provide a better packing of the solute into the aqueous solution.

Also in the case of the phosphorylation studies, these would benefit from investigating larger model compounds. For example phosphorylation of amylosidic chains large enough to form helical structures would be extremely interesting.

7.1 Experimental

7.1.1 Crystal structures

New and more accurate experimental data are always informative, and in the case of carbohydrate studies better structural data are required. As mentioned previously, there is a shortage of X-ray reference data in the area of carbohydrate structure in particular carbohydrate oligomers. This is, of course, caused by the inherent difficulty in crystallizing the very flexible saccharides. However, the results from Paper II suggest that besides α -panose and maltotriose, both iso-panose and α -forkose are also fairly rigid molecules. Attempts to crystallize these compounds using modern crystallization techniques, will, if successful, be very informative about the structural preferences of the trisaccharides and perhaps their hydration. In particular information on the structure of the the $\alpha(1\rightarrow6)$ glucosidic linkage, which is not well described in literature, will be valuable.

7.1.2 NMR structures

Measurement of residual dipolar couplings could be a source of stronger distance restraint on the carbohydrate structures [195–197]. This approach should however not be applied without thoroughly investigating the consequences of changing the solute environment to a partially ordered phase. The whole purpose of this type of studies is to understand the water-solute interactions, and introducing an agent that most likely disturbs the structure and dynamics of the water network could possibly lead to misleading results.

Similar considerations should be applied to the proposed usage of super-cooled NMR measurements on the hydroxyl protons [198]. The requirement for this type of experiment, is a very low temperature obtained by using

acetone as an additive to the solvent. In this approach there is likewise a fair possibility that the water network will change, and this will have to be studied further by comparison of well described model systems such as maltose or sucrose.

A limited number of inter residue NOEs were identified from the NOESY and ROESY spectra recorded for both Paper II and Paper V. Given that the full assignment is now available, it will be informative to conduct more detailed studies using series of NOESY spectra with varying mixing times. Using this approach, quantitative distance information should be obtainable.

Furthermore, measurements of the rotational diffusion constants for the five trisaccharide samples could provide a good basis for development of a better model to describe this type of motion than the currently used dipole moment.

Regarding the observation of nano-crystal morphology as in Paper III, these models can be significantly improved by additional and improved experimental data, and by application of more refined methods for image analysis. Until now the TEM images have been analyzed by hand using a standard protractor. This method is tedious and prone to errors, so efforts should be given to the development of an automation scheme such as computerized image analysis. Another problem is that for some starch samples - for unknown reasons - it is very difficult to obtain a separation of crystallites good enough to yield TEM images suitable for further analysis.

7.2 Computational

Computer aided methods to assist with assignment of resonances in NMR spectra and structural analysis should be explored further. Some initiatives are already trying to do this. For example the CASPER program [199], and the system proposed by Lipkind et al. [200] are promising possibilities.

7.2.1 Karplus parametrization

It has become clear from the results in Paper II, and other results in the literature, that the $\alpha(1\rightarrow6)$ glucosidic bond is not well characterized by the explored molecular mechanics force fields or is not well described by the available Karplus like parametrizations of the ${}^3J_{CH}$ coupling in this particular structural motif. The relation used in these studies [147] and even the recent similar result from Tafazzoli et al. [149], are derived using either static crystallographic data, or DFT calculations on mono-saccharides with smaller substituents⁸. This is seemingly not adequate and perhaps a re-parametrization should be performed from analysis of disaccharide structures, including both $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ glucosidic linkages.

7.2.2 Force field parametrization

The problems as mentioned in Section 7.2.1 could easily be caused by inadequate force field parametrization of the primary hydroxyl groups. Even though the CSFF force field [62] is designed with this exact goal in mind, it is clear that further improvements are required. The next generation of carbohydrate force fields needs to be parametrized not only with the free primary hydroxyl group in mind, but also the $\alpha(1\rightarrow6)$ glucosidic linkage. Thus, focus should be turned toward extensive study of iso-maltose using DFT calculations and experimental methods to gain the same amount of information on this fundamental structure as exist for maltose.

For molecular dynamics calculations, the wide selection of force fields brings up a dilemma in choosing the appropriate parameter set. Ideally several force fields should be tested on each new system, following the methodology of Perez et al. [67] and Hemmingsen et al. [69]. This is not always practical because of the time consuming nature of these calculations.

Another aspect to keep an eye on is the development of the molecular dynamics software. During this project we started using TINKER, both to gain the competence, but also because of its flexibility by for example being able to easily switch between force-fields. In the same period of time, the folding@Home project [201–202] switched to GROMACS for explicit

⁸ Substituents such as methyl- or ethyl-groups, or halogens.

solvation calculations due to speed considerations. For future projects, which will inevitably involve larger solvation systems, the choice of tool to be used for the calculations must be considered carefully.

8 Conclusions

In the work included in this thesis molecular modeling was applied to starch model systems at both microscopic scale and at atomic resolution, and it was demonstrated how this semi-empirical theoretical tool in combination with experimental techniques, of such variety as TEM imaging and NMR spectroscopy, can bring new insight into detailed structural features and fast time scale dynamic interactions.

A series of models were built to investigate subtle differences in geometric features of 2D nanometer scale starch crystallites when subjected to variations in the systematic packing of the helices. This method worked very well to explain experimental TEM observations of hydrolyzed A-type starch samples that was not in agreement with the currently accepted models of this particular packing type. By building an alternative A-type crystallite using a simple principle of inter-helical displacement, agreement with the experimental data was achieved. In a similar manner, the structure of B-type starch as reported in the literature was confirmed.

The result from Paper II reviewed in Chapter 5 shows a full characterization of all trisaccharide fragments of the starch polymer and how they fit within the crystalline double helical amylopectin regions and also suggest how they could be hydrated in the amorphous parts of the system. By performing a full conformational and hydration analysis of the trisaccharides, the role of water as an important structural element was further established. It was shown how the amylopectin branch-point conformation is induced by the stable conformation of the central point trisaccharide, α -forkose.

It was furthermore suggested, that a shared water molecule between two inter ring oxygen atoms must have a density of 3.0 or more, when compared to bulk water, in order to force a dihedral arrangement into a stable conformation. This hypothesis will hopefully be supported in future studies.

The same techniques were applied to study the structural consequences of maltose phosphorylation, by simulation and measurements on model compounds representing the two known forms of starch phosphorylation. It was found that phosphorylation in the 3-position of the glucose ring is

not compatible with the ordered phase of amylopectin, in which case it will function as a structure breaker. Phosphorylation in the 6-position match into the double helical structure, and can therefore readily be accepted.

This study also revealed that a functional group with considerable rotational freedom, such as a phosphate group, participates in the water sharing network in a more complex manner than the hydroxyl groups. Interactions of all three oxygen atoms in the phosphate with other water binding sites of the solute has to be taken into account as there is a distributed contribution from these that should not be ignored.

The successful use of the PARAFAC model as a tool to extract pure sub-spectra and diffusional profiles from 2D diffusion edited NMR spectra was demonstrated in Paper V and Paper VI, on diverse systems such as the medically important lipoprotein systems, and also on mixtures of small carbohydrates. This technique has a promising future since it is not limited to the presently studied samples, but can possibly be applied in a wide range of settings, for example in future investigations of degradation products in the enzymatic breakdown of starch.

- Peter I. Hansen, May 2008

9 References

- [1] F. Franks, *Water Relationships in Food* (Plenum Press: New York 1991), 1–19
- [2] S. B. Engelsens, C. Herve du Penhoat, and S. Perez, *Molecular relaxation of sucrose in aqueous solutions: how a nanosecond molecular dynamics simulation helps to reconcile NMR data*, *The Journal of Physical Chemistry* **99** (1995), 13334–13351
- [3] R. P. Ellis, M. P. Cochrane, M. F. B. Dale, C. M. Duffus, A. Lynn, I. M. Morrison, R. D. M. Prentice, J. S. Swanston, and S. A. Tiller, *Starch production and industrial use*, *J. Sci. Food Agric.* **77** (1998), 289–311
- [4] P. H. Richardson, R. Jeffcoat, and Y. C. Shi, *High-amylose starches: from biosynthesis to their use as food ingredients*, *MRS Bulletin* **25** (2000), 20–24
- [5] D. T. Gordonet *al.*, *Resistant starch: physical and physiological properties*. (ATL Press, Inc. Science Publishers, Shrewsbury, MA, 1997)
- [6] D. L. Topping, M. Fukushima, and A. R. Bird, *Resistant starch as a prebiotic and synbiotic: state of the art*, *Proceedings of the Nutrition Society* **62** (2003), 171–176
- [7] S. J. Angyal, *The composition of reducing sugars in solution*, *Advances in carbohydrate chemistry and biochemistry* **42** (1984), 15–68
- [8] The IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), *Symbols for specifying the conformation of polysaccharide chains*, *Pure & Applied Chemistry* **55** (1983), 1269–1272
- [9] G. M. Brown and H. A. Levy, *α -D-Glucose: Precise Determination of Crystal and Molecular Structure by Neutron-Diffraction Analysis*, *Science* **147** (1965), 1038–1039
- [10] J. W. Brady, *Molecular dynamics simulations of α -D-glucose in aqueous solution*, *J. Am. Chem. Soc.* **111** (1989), 5155–5165
- [11] A. Buléon, P. Colonna, V. Planchot, and S. Ball, *Starch granules: structure and biosynthesis*, *Int. J. Biol. Macromol.* **23** (1998), 85–112
- [12] D. B. Thompson, *On the non-random nature of amylopectin branching*, *Carbohydr. Polym.* **43** (2000), 223–239

- [13] D. J. Gallant, B. Bouchet, and P. M. Baldwin, *Microscopy of starch: evidence of a new level of granule organization*, Carbohydr. Polym. **32** (1997), 177–191
- [14] T. Noda, S. Takigawa, C. Matsuura-Endo, S. J. Kim, N. Hashimoto, H. Yamauchi, I. Hanashiro, and Y. Takeda, *Physicochemical properties and amylopectin structures of large, small, and extremely small potato starch granules*, Carbohydr. Polym. **60** (2005), 245–251
- [15] P. J. Jenkins and A. M. Donald, *The influence of amylose on starch granule structure*, Int. J. Biol. Macromol. **17** (1995), 315–321
- [16] D. French, *Organization of starch granules*, Starch: Chemistry and Technology **2** (1984), 183–247
- [17] J. Blanshard, D. Bates, A. Muhr, D. Worcester, and J. Higgins, *Small-angle neutron scattering studies of starch granule structure*, Carbohydr. Polym. **4** (1984), 427–442
- [18] G. T. Oostergetel and E. F. J. Van Bruggen, *On the origin of a low angle spacing in starch*, Stärke **41** (1989), 331–335
- [19] P. Muhrbeck, *On crystallinity in cereal and tuber starches*, Stärke **43** (1991), 347–348
- [20] G. T. Oostergetel and E. F. J. Van Bruggen, *The crystalline domains in potato starch granules are arranged in a helical fashion*, Carbohydr. Polym. **21** (1993), 7–12
- [21] T. A. Waigh and A. M. Donald, *Analysis of the Native Structure of Starch Granules with Small Angle X-Ray Microfocus Scattering*, Faraday Discuss. **103** (1996), 325–337
- [22] T. A. Waigh, A. M. Donald, F. Heidelberg, C. Riekkel, and M. J. Gidley, *Analysis of the native structure of starch granules with SAXS microfocus diffraction*, Biopolymers **49** (1999), 91–105
- [23] E. Bertoft, *On the nature of categories of chains in amylopectin and their connection to the super helix model*, Carbohydr. Polym. **57** (2004), 211–224
- [24] A. Sarko and H. C. H. Wu, *The crystal structures of A-, B- and C-polymorphs of amylose and starch*, Starch/Stärke **30** (1978), 73–78
- [25] M. J. Gidley and S. M. Bociak, *Molecular organization in starches: a carbon 13 CP/MAS NMR study*, J. Am. Chem. Soc. **107** (1985), 7040–7044
- [26] A. Imberty, H. Chanzy, S. Perez, A. Buleon, and V. Tran, *The double-helical nature of the crystalline part of A-starch*, J. Mol. Biol. **201** (1988), 365–378

- [27] D. R. Kreger, *The configuration and packing of the chain molecules of native starch as derived from x-ray diffraction of part of a single starch grain*, *Biochim. Biophys. Acta* **6** (1951), 406–425
- [28] J. Blackwell, A. Sarko, and R. H. Marchessault, *Chain conformation in B-amylose*, *J. Mol. Biol.* **42** (1969), 379–383
- [29] H. C. H. Wu and A. Sarko, *The double-helical molecular structure of crystalline B-amylose*, *Carbohydr. Res* **61** (1978), 7–25
- [30] A. Imberty and S. Perez, *A revisit to the three-dimensional structure of B-type starch*, *Biopolymers* **27** (1988), 1205–1221
- [31] T. Y. Bogracheva, V. J. Morris, S. G. Ring, and C. L. Hedley, *The granular structure of C-type pea starch and its role in gelatinization*, *Biopolymers* **45** (1998), 323–332
- [32] D. L. Wild and J. M. V. Blanshard, *The relationship of the crystal structure of amylose polymorphs to the structure of the starch granule*, *Carbohydr. Polym.* **6** (1986), 121–143
- [33] A. C. Eliasson, K. Larsson, S. Andersson, S. T. Hyde, R. Nesper, and H. G. Von Schnering, *On the structure of native starch- An analogue to the quartz structure*, *Stärke* **39** (1987), 147–152
- [34] J. P. Robin, C. Mercier, F. Duprat, R. Charbonnieare, and A. Guilbot, *Lintnerized starches—Chromatographic and enzymatic studies of insoluble residues from acid hydrolysis of various cereal starches, particularly waxy maize starch*, *Starch/Starke* **27** (1975), 36–45
- [35] H. Angellier, J. Putaux, S. Molina-Boisseau, D. Dupeyre, and A. Dufresne, *Starch nanocrystals fillers in an acrylic polymer matrix*, *Macromol. Symp.* **221** (2005), 95–104
- [36] A. Pohu, J. L. Putaux, V. Planchot, P. Colonna, and A. Buléon, *Origin of the Limited-Amylolysis of Debranched Maltodextrins Crystallized in the A Form: A TEM Study on Model Substrates*, *Biomacromolecules* **5** (2004), 119–125
- [37] J. L. Putaux, S. Molina-Boisseau, T. Momaur, and A. Dufresne, *Platelet nanocrystals resulting from the disruption of waxy maize starch granules by acid hydrolysis*, *Biomacromolecules* **4** (2003), 1198–1202
- [38] B. J. Alder and T. E. Wainwright, *An experimental study of radial distribution of particles in a circulating fluidized bed*, *J. Chem. Phys.* **33** (1960), 1439–1451
- [39] J. A. McCammon, B. R. Gelin, and M. Karplus, *Dynamics of folded proteins*, *Nature* **267** (1977), 585–590

- [40] M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules*, Nature Structural Biology **9** (2002), 646–652
- [41] F. Franks, D. S. Reid, and A. Suggett, *Conformation and hydration of sugars and related compounds in dilute aqueous solution*, J. Solution Chem. **2** (1973), 99–118
- [42] K. N. Kirschner and R. J. Woods, *Solvent interactions determine carbohydrate conformation*, Proceedings of the National Academy of Sciences **98** (2001), 10541–10545
- [43] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations*, J. Comput. Chem. **4** (1983), 187–217
- [44] P. K. Weiner and P. A. Kollman, *AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions*, J. Comput. Chem. **2** (1981), 287–303
- [45] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *The Amber biomolecular simulation programs*, J. Comput. Chem **26** (2005), 1668–1688
- [46] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. **91** (1995), 43–56
- [47] E. Lindahl, B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, Journal of Molecular Modeling **7** (2001), 306–317
- [48] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *GROMACS: fast, flexible, and free*, J. Comput. Chem **26** (2005), 1701–1718
- [49] J. W. Ponder and others, *TINKER: Software Tools for Molecular Design* (2000).
- [50] T. Peters, B. Meyer, R. Stuike-Prill, R. Somorjai, and J. R. Brisson, *A Monte Carlo method for conformational analysis of saccharides*, Carbohydr. Res. **238** (1993), 49–73
- [51] A. D. MacKerell Jr, J. Wiorkiewicz-Kuczera, and M. Karplus, *An all-atom empirical energy function for the simulation of nucleic acids*, J. Am. Chem. Soc. **117** (1995), 11946–11975

- [52] B. M. Pettitt and M. Karplus, *Role of electrostatics in the structure, energy and dynamics of biomolecules: a model study of N-methylalanylacetamide*, J. Am. Chem. Soc. **107** (1985), 1166–1173
- [53] J. E. Lennard-Jones, *Cohesion*, Proceedings of the Physical Society **43** (1931), 461–482
- [54] A. T. Hagler, E. Huler, and S. Lifson, *Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals*, Journal of the American Chemical Society **96** (1974), 5319–5327
- [55] L. Verlet, *Computer ‘experiments’ on condensed fluids I. Thermodynamical properties of Lennard–Jones molecules*, Phys. Rev **159** (1967), 98–103
- [56] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*. (Prentice Hall PTR Upper Saddle River, NJ, USA, 1971)
- [57] D. Beeman, *Some Multistep Methods for Use in Molecular Dynamics Calculations*, Journal of Computational Physics **20** (1976), 130–139
- [58] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*, The Journal of Chemical Physics **76** (1982), 637–649
- [59] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *Numerical integration of the Cartesian equations of proteins and nucleic acids*, J. Comput. Phys **23** (1977), 327–341
- [60] H. C. Andersen, *Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations*, Journal of Computational Physics **52** (1983), 24–34
- [61] S. N. Ha, A. Giammona, M. Field, and J. W. Brady, *A revised potential-energy surface for molecular mechanics studies of carbohydrates*, Carbohydr. Res. **180** (1988), 207–221
- [62] M. Kuttel, J. W. Brady, and K. J. Naidoo, *Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations*, J. Comput. Chem. **23** (2002), 1236–1243

- [63] R. J. Woods, R. A. Dwek, C. J. Edge, and B. Fraser-Reid, *Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development*, The Journal of Physical Chemistry **99** (1995), 3832–3846
- [64] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *A new force field for molecular mechanical simulation of nucleic acids and proteins*, J. Am. Chem. Soc. **106** (1984), 765–784
- [65] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *An all atom force field for simulations of proteins and nucleic acids*, J. Comput. Chem. **7** (1986), 230–252
- [66] R. D. Lins and P. H. Hünenberger, *A new GROMOS force field for hexopyranose-based carbohydrates*, J. Comput. Chem **26** (2005), 1400–1412
- [67] S. Pérez, A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jimenez-Barbero, A. Poveda, J. F. Espinosa, B. P. van Eyck, and G. Johnson, *A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates*, Carbohydr. Res. **314** (1998), 141–155
- [68] A. D. Mackerell, *Empirical force fields for biological macromolecules: Overview and issues*, J. Comput. Chem. **25** (2004), 1584–1604
- [69] L. Hemmingsen, D. E. Madsen, A. L. Esbensen, L. Olsen, and S. B. Engelsen, *Evaluation of carbohydrate molecular mechanical force fields by quantum mechanical calculations*, Carbohydr. Res. **339** (2004), 937–948
- [70] H. C. Andersen, *Molecular dynamics simulations at constant pressure and/or temperature*, The Journal of Chemical Physics **72** (1980), 2384–2393
- [71] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *Molecular dynamics with coupling to an external bath*, The Journal of Chemical Physics **81** (1984), 3684–3690
- [72] P. P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*, Annalen der Physik **369** (1921), 253–287
- [73] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. (Oxford University Press, USA, 1989)

- [74] C. J. Fennell and J. D. Gezelter, *Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics*, *The Journal of Chemical Physics* **124** (2006), 234104–234116
- [75] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Interaction models for water in relation to protein hydration*, *Intermolecular Forces* **1** (1981), 331–342
- [76] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating water*, *J. Chem. Phys.* **79** (1983), 926–935
- [77] F. H. Stillinger and A. Rahman, *Improved simulation of liquid water by molecular dynamics*, *J. Chem. Phys.* **60** (1974), 1545–1557
- [78] P. Ren and J. W. Ponder, *Polarizable atomic multipole water model for molecular mechanics simulation*, *J. Phys. Chem. B* **107** (2003), 5933–5947
- [79] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen, *A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field*, *J. Chem. Phys.* **108** (1998), 10220–10230
- [80] J. L. Finney, *The water molecule and its interactions: the interaction between theory, modelling, and experiment*, *J. Mol. Liq.* **90** (2001), 303–312
- [81] J. R. Errington and P. G. Debenedetti, *Relationship between structural order and the anomalies of liquid water*, *Nature* **409** (2001), 318–321
- [82] W. L. Jorgensen, *Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water*, *J. Am. Chem. Soc.* **103** (1981), 335–340
- [83] A. D. MacKerell Jr and N. Banavali, *Development and Current Status of the CHARMM Force Field for Nucleic Acids*, *Biopolymers (Nucleic Acid Sciences)* **56** (2001), 257–265
- [84] F. Corzana, M. S. Motawia, C. Herve du Penhoat, F. van den Berg, A. Blennow, S. Perez, and S. B. Engelsen, *Hydration of the amylopectin branch point. Evidence of restricted conformational diversity of the α -(1 \rightarrow 6) linkage*, *J. Am. Chem. Soc.* **126** (2004), 13144–13155
- [85] M. K. Dowd, J. Zeng, A. D. French, and P. J. Reilly, *Conformational analysis of the anomeric forms of kojibiose, nigerose, and maltose using MM3*, *Carbohydr. Res.* **230** (1992), 223–244

- [86] F. Corzana, M. S. Motawia, C. H. Du Penhoat, S. Perez, S. M. Tschampel, R. J. Woods, and S. B. Engelsen, *A Hydration Study of (1 → 4) and (1 → 6) Linked α -Glucans by Comparative 10 ns Molecular Dynamics Simulations and 500-MHz NMR*, *J. Comput. Chem.* **25** (2004), 573–586
- [87] M. M. Kuttel and K. J. Naidoo, *Free energy surfaces for the $\alpha(1 \rightarrow 4)$ -glycosidic linkage: Implications for polysaccharide solution structure and dynamics*, *Journal of physical chemistry. B, Condensed matter, materials, surfaces, interfaces, & biophysical chemistry* **109** (2005), 7468–7474
- [88] N. L. Allinger, M. Rahman, and J. H. Lii, *A molecular mechanics force field (MM3) for alcohols and ethers*, *J. Am. Chem. Soc.* **112** (1990), 8293–8307
- [89] A. D. French, A. M. Kelterer, G. P. Johnson, M. K. Dowd, and C. J. Cramer, *HF/6-31 G* energy surfaces for disaccharide analogs*, *Journal of Computational Chemistry* **22** (2001), 65–78
- [90] F. A. Momany, S. U., J. L. Willett, and W. B. Bosma, *DFT study of α -maltose: influence of hydroxyl orientations on the glycosidic bond*, *Struct. Chem.* **18** (2007), 611–632
- [91] U. Schnupf, J. L. Willett, W. B. Bosma, and F. A. Momany, *DFT studies of the disaccharide, α -maltose: relaxed isopotential maps*, *Carbohydr. Res.* **342** (2007), 2270–2285
- [92] M. K. Dowd, P. J. Reilly, and A. D. French, *Relaxed-residue conformational mapping of the three linkage bonds of isomaltose and gentiobiose with MM 3 (92)*, *Biopolymers* **34** (1994), 625–638
- [93] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *Semi-analytical treatment of solvation for molecular mechanics and dynamics*, *J. Am. Chem. Soc.* **112** (1990), 6127–6129
- [94] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *Pairwise solute descreening of solute charges from a dielectric medium*, *Chem. Phys. Lett.* **246** (1995), 122–129
- [95] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, *The GB/SA Continuum Model for Solvation-A Fast Analytical Method for the Calculation of Approximate Born Radii*, *J. Phys. Chem. A* **101** (1997), 3005–3014
- [96] A. Onufriev, D. Bashford, and D. A. Case, *Modification of the generalized Born model suitable for macromolecules*, *J. Phys. Chem. B* **104** (2000), 3712–3720

- [97] R. E. Bruccoleri and M. Karplus, *Conformational sampling using high-temperature molecular dynamics*, *Biopolymers* **29** (1990), 1847–1862
- [98] M. Frank, *Rapid Generation of a Representative Ensemble of N-Glycan Conformations*, In *Silico Biology* **2** (2002), 427–439
- [99] H. Uedaira and H. Uedaira, *Diffusion Coefficients of Xylose and Maltose in Aqueous Solution*, *Bull. Chem. Soc. Jpn.* **42** (1969), 2140–2142
- [100] C. Monteiro and C. H. du Penhoat, *Translational Diffusion of Dilute Aqueous Solutions of Sugars as Probed by NMR and Hydrodynamic Theory*, *J. Phys. Chem. A* **105** (2001), 9827–9833
- [101] P. J. Rossky and M. Karplus, *Solvation. a molecular dynamics study of a dipeptide in water*, *J. Am. Chem. Soc.* **101** (1979), 1913–1937
- [102] S. B. Engelsen and S. Perez, *The hydration of sucrose*, *Carbohydr. Res.* **292** (1996), 21–38
- [103] S. B. Engelsen, C. Monteiro, C. Hervé de Penhoat, and S. Pérez, *The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy*, *Biophys. Chem.* **93** (2001), 103–127
- [104] K. J. Naidoo and J. Y. J. Chen, *The role of water in the design of glycosidic linkage flexibility*, *Mol. Phys.* **101** (2003), 2687–2694
- [105] A. Almond, *Towards understanding the interaction between oligosaccharides and water molecules*, *Carbohydr. Res.* **340** (2005), 907–920
- [106] S. B. Engelsen, *Molecular Dynamics Simulations of Carbohydrate-Water Interactions*, *Water in Biomaterials Surface Science* (2001), 53–89
- [107] M. J. Field, *A Practical Introduction to the Simulation of Molecular Systems*. (Cambridge University Press, 1999)
- [108] C. Andersson and S. Balling Engelsen, *The mean hydration of carbohydrates as studied by normalized two-dimensional radial pair distributions*, *Journal of Molecular Graphics and Modelling* **17** (1999), 101–105
- [109] Q. Liu and J. W. Brady, *Anisotropic solvent structuring in aqueous sugar solutions*, *J. Am. Chem. Soc.* **118** (1996), 12276–12286
- [110] W. L. Bragg, *Die Struktur einiger Kristalle, gedeutet nach ihrer Beugung von Roentgen-Strahlen*, *Zeitschrift fuer anorganische Chemie* **90** (1914), 185–218

- [111] R. E. Franklin and R. G. Gosling, *Molecular Configuration in Sodium Thymonucleate*, *Nature* **171** (1953), 740–741
- [112] R. E. Franklin and R. G. Gosling, *Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate*, *Nature* **172** (1953), 156–157
- [113] J. D. Watson and F. H. C. Crick, *Molecular structure of nucleic acids*, *Nature* **171** (1953), 737–738
- [114] E. S. Stevens and B. K. Sathyanarayana, *Potential-energy surfaces of cellobiose and maltose in aqueous solution: a new treatment of disaccharide optical rotation*, *J. Am. Chem. Soc.* **111** (1989), 4149–4154
- [115] F. Takusagawa and R. A. Jacobson, *The crystal and molecular structure of α -maltose*, *Structural Crystallography and Crystal Chemistry* **34** (1978), 213–218
- [116] G. J. Quigley, A. Sarko, and R. H. Marchessault, *Crystal and molecular structure of maltose monohydrate*, *J. Am. Chem. Soc.* **92** (1970), 5834–5839
- [117] S. M. E. Gres and G. A. Jeffrey, *A neutron diffraction refinement of the crystal structure of [beta]-maltose monohydrate*, *Structural Crystallography and Crystal Chemistry* **33** (1977), 2490–2495
- [118] S. Perez, F. Taravel, and C. Vergelati, *Experimental evidences of solvent induced conformational changes in maltose*, *Nouveau Journal de Chimie*(1977) **9** (1985), 561–564
- [119] W. Pangborn, D. Langs, and S. Perez, *Regular left-handed fragment of maylose: crystal and molecular structure of methyl- α -maltotrioxide, $4H_2O$* , *Int. J. Biol. Macromol* **7** (1985), 363–369
- [120] A. Imberty and S. Perez, *Crystal structure and conformational features of α -panose*, *Carbohydr. Res.* **181** (1988), 41–55
- [121] G. A. Jeffrey and D. B. Huang, *Hydrogen bonding in the crystal structure of alpha, beta-panose*, *Carbohydr. Res.* **222** (1991), 47–55
- [122] F. Duprat, D. Gallant, A. Guilbot, C. Mercier, and J. Robin, *L'amidon (Starch)*, *Les Polymeres Végétaux (Vegetable Polymers)* (1980), 176–231
- [123] A. C. O'Sullivan and S. Perez, *The relationship between internal chain length of amylopectin and crystallinity in starch*, *Biopolymers* **50** (1999), 381–390

- [124] K. Gessler, I. Usón, T. Takaha, N. Krauss, S. M. Smith, S. Okada, G. M. Sheldrick, and W. Saenger, *V-Amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltohexaicosaoose)*, Proc Natl Acad Sci US A **96** (1999), 4246–4251
- [125] G. Breit and I. I. Rabi, *Measurement of Nuclear Spin*, University of California, Berkeley **38** (1931), 2082–2083
- [126] I. Rabi, *Space Quantization in a Gyating Magnetic Field*, Physical Review **51** (1937), 652–654
- [127] F. Bloch, W. W. Hansen, and M. Packard, *The Nuclear Induction Experiment*, Physical Review **70** (1946), 474–485
- [128] E. M. Purcell, H. C. Torrey, and R. V. Pound, *Resonance Absorption by Nuclear Magnetic Moments in a Solid*, Physical Review **69** (1946), 37–38
- [129] B. Mulloy, *High-field NMR as a technique for the determination of polysaccharide structures*, Molecular Biotechnology **6** (1996), 241–265
- [130] J. Duus, C. H. Gotfredsen, and K. Bock, *Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations*, Chem Rev **100** (2000), 4589–4614
- [131] T. Rundlof, A. Kjellberg, C. Damberg, T. Nishida, and G. Widmalm, *Long-range proton-carbon coupling constants in conformational analysis of oligosaccharides*, Magn. Reson. Chem. **36** (1998), 839–847
- [132] N. W. H. Cheetham, P. Dasgupta, and G. E. Ball, *NMR and modelling studies of disaccharide conformation*, Carbohydr. Res. **338** (2003), 955–962
- [133] A. Parfondry, C. N., and A. S. Perlin, *^{13}C - ^1H inter-residue coupling in disaccharides, and the orientations of glycosidic bonds*, Carbohydr. Res **59** (1977), 299–309
- [134] A. A. Bothner-By, R. L. Stephens, J. Lee, C. D. Warren, and R. W. Jeanloz, *Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame*, J. Am. Chem. Soc. **106** (1984), 811–813
- [135] A. S. Shashkov, G. M. Lipkind, and N. K. Kochetkov, *Nuclear Overhauser effects for methyl β -maltoside and the conformational states of maltose in aqueous solution*, Carbohydr. Res. **147** (1986), 175–182

- [136] W. P. Aue, E. Bartholdi, and R. R. Ernst, *Two-dimensional spectroscopy. Application to nuclear magnetic resonance*, The Journal of Chemical Physics **64** (1976), 2229–2246
- [137] L. Braunschweiler and R. R. Ernst, *Coherence transfer by isotropic mixing: application to proton correlation spectroscopy*, J. Magn. Reson **53** (1983), 521–528
- [138] A. Bax and D. G. Davis, *MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy*, J. Magn. Reson. **65** (1985), 355–360
- [139] G. Bodenhausen and D. J. Ruben, *Heteronuclear 2D correlation spectra with double in-phase transfer steps*, Chem. Phys. Letters **69** (1980), 185–189
- [140] W. Willker, D. Leibfritz, R. Kerssebaum, and W. Bermel, *Gradient selection in inverse heteronuclear correlation spectroscopy*, Magn. Reson. Chem **31** (1993), 287–292
- [141] A. Bax and M. F. Summers, *^1H and ^{13}C assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR*, J. Am. Chem. Soc. **108** (1986), 2093–2094
- [142] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *NMRPipe: A multidimensional spectral processing system based on UNIX pipes*, J. Biomol. NMR **6** (1995), 277–293
- [143] T. D. Goddard and D. G. Kneller *SPARKY 3* (unpublished). University of California, San Francisco.
- [144] A. Meissner and O. W. Sørensen, *Measurement of $J(H, H)$ and long range $J(X, H)$ coupling constants in small molecules. Broadband XLOC and J-HMBC*, Magn. Reson. Chem **39** (2001), 49–52
- [145] M. Karplus, *Contact electron-spin coupling of nuclear magnetic resonance*, J. Chem. Phys **30** (1959), 11–15
- [146] M. Karplus, *Vicinal Proton Coupling in Nuclear Magnetic Resonance*, J. Am. Chem. Soc. **85** (1963), 2870–2871
- [147] I. Tvaroska, M. Hricovini, and E. Petrakova, *An attempt to derive a new Karplus-type equation of vicinal proton-carbon coupling constants for C-O-C-H segments of bonded atoms*, Carbohydr. Res. **189** (1989), 359–362
- [148] B. Mulloy, T. A. Frenkiel, and D. B. Davies, *Long-range carbon-proton coupling constants: application to conformational studies of oligosaccharides*, Carbohydr. Res. **184** (1988), 39–46

- [149] M. Tafazzoli and M. Ghiasi, *New Karplus equations for $^2J_{HH}$, $^3J_{HH}$, $^2J_{CH}$, $^3J_{CH}$, $^3J_{COCH}$, $^3J_{CSCH}$, and $^3J_{CCCH}$ in some aldohexopyranoside derivatives as determined using NMR spectroscopy and density functional theory calculations*, Carbohydr. Res. **342** (2007), 2086–2096
- [150] C. A. G. Haasnoot, F. A. A. M. de Leeuw, and C. B. Altona, *The relationship between proton-proton nmr coupling constants and substituent electronegativities—1. an empirical generalization of the karplus equation*, Tetrahedron **36** (1980), 2783–2792
- [151] A. W. Overhauser, *Polarization of Nuclei in Metals*, Physical Review **92** (1953), 411–415
- [152] T. R. Carver and C. P. Slichter, *Polarization of Nuclear Spins in Metals*, Physical Review **92** (1953), 212–213
- [153] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst, *Investigation of exchange processes by two-dimensional NMR spectroscopy*, The Journal of Chemical Physics **71** (1979), 4546–4553
- [154] A. Bax and D. G. Davis, *Practical aspects of two-dimensional transverse NOE spectroscopy*, J. Magn. Reson. **63** (1985), 207–213
- [155] K. Wüthrich, *NMR of Proteins and Nucleic Acids*. (John Wiley and Sons, inc. New York, 1986)
- [156] R. B. Cattell, *“Parallel proportional profiles” and other principles for determining the choice of factors by rotation*, Psychometrika **9** (1944), 267–283
- [157] J. Carroll and J. J. Chang, *Analysis of individual differences in multidimensional scaling via an n -way generalization of “Eckart-Young” decomposition*, Psychometrika **35** (1970), 283–319
- [158] R. A. Harshman, *Foundations of the PARAFAC procedure: model and conditions for an ‘exploratory’ multi-mode factor analysis*, UCLA Working Papers in Phonetics **16** (1970), 1–84
- [159] R. Bro, *PARAFAC. Tutorial and applications*, Chemom. Intell. Lab. Syst. **38** (1997), 149–171
- [160] R. Bro, P. I. Hansen, N. Viereck, M. Dyrby, H. T. Pedersen, and S. B. Engelsen, *A new principle for unique spectral decomposition of 2D NMR data*, Magnetic Resonance in Food Science. The Multivariate Challenge (2004), 195–203
- [161] J. W. Lee, T. O. Kwon, and I. S. Moon, *Adsorption of monosaccharides, disaccharides, and maltooligosaccharides on activated carbon for separation of maltopentaose*, Carbon **42** (2004), 371–380

- [162] M. Dyrby, M. Petersen, A. K. Whittaker, L. Lambert, L. Nørsgaard, R. Bro, and S. B. Engelsen, *Analysis of lipoproteins using 2D diffusion-edited NMR spectroscopy and multi-way chemometrics*, *Anal. Chim. Acta* **531** (2005), 209–216
- [163] S. N. Ha, L. J. Madsen, and J. W. Brady, *Conformational analysis and molecular dynamics simulations of maltose*, *Biopolymers* **27** (1988), 1927–1952
- [164] J. W. Brady and R. K. Schmidt, *The role of hydrogen bonding in carbohydrates: molecular dynamics simulations of maltose in aqueous solution*, *J. Phys. Chem.* **97** (1993), 958–966
- [165] K. H. Ott and B. Meyer, *Molecular dynamics simulations of maltose in water*, *Carbohydr. Res.* **281** (1996), 11–34
- [166] R. B. Best, G. E. Jackson, and K. J. Naidoo, *Molecular Dynamics and NMR Study of the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ Glycosidic Linkages: Maltose and Isomaltose*, *J. Phys. Chem. B* **105** (2001), 4742–4751
- [167] M. Umemura, S. Hayashi, T. Nakagawa, H. Urakawa, and K. Kajiwara, *Structure of water molecules in aqueous maltose and cellobiose solutions using molecular dynamics simulation. II. Dynamics*, *Journal of Molecular Structure: THEOCHEM* **636** (2003), 215–228
- [168] R. B. Best, G. E. Jackson, and K. J. Naidoo, *Modeling the $\alpha(1\rightarrow6)$ Branch Point of Amylopectin in Solution*, *J. Phys. Chem. B* **106** (2002), 5091–5098
- [169] K. J. Naidoo and M. Kuttel, *Water structure about the dimer and hexamer repeat units of amylose from molecular dynamics computer simulations*, *J. Comput. Chem.* **22** (2001), 445–456
- [170] F. A. Momany and J. L. Willett, *Molecular dynamics calculations on amylose fragments. I. Glass transition temperatures of maltodecaose at 1, 5, 10, and 15.8% hydration*, *Biopolymers* **63** (2002), 99–110
- [171] M. S. Motawia, I. Damager, C. E. Olsen, B. L. Møller, S. B. Engelsen, S. Hansen, L. H. Øgendal, and R. Bauer, *Comparative Study of Small Linear and Branched r -Glucans Using Size Exclusion Chromatography and Static and Dynamic Light Scattering*, *Biomacromolecules* **6** (2005), 143–151
- [172] A. Thompson and M. L. Wolfrom, *The Structure of Maltotriose*, *J. Am. Chem. Soc.* **74** (1952), 3612–3614
- [173] U. Schnupf, J. L. Willett, W. B. Bosma, and F. A. Momany, *DFT conformational studies of alpha-maltotriose*, *J. Comput. Chem.* (2007)

- [174] K. Kainuma and D. French, *Action of pancreatic amylase on starch oligosaccharides containing single glucose side chains.*, FEBS Letters **5** (1969), 257–261
- [175] K. Umeki and T. Yamamoto, *Structures of Multi-Branched Dextrins Produced by Saccharifying {alpha}-Amylase from Starch*, Journal of Biochemistry **78** (1975), 897
- [176] S. B. Engelsen, S. Cros, W. Mackie, and S. Pérez, *A molecular builder for carbohydrates: application to polysaccharides and complex carbohydrates*, Biopolymers **39** (1996), 417–433
- [177] S. Perez and M. M. Delage, *A database of three-dimensional structure of monosaccharides from molecular-mechanics calculations*, Carbohydr. Res. **212** (1991), 253–259
- [178] S. Lim and P. Seib, *Location of phosphate esters in a wheat starch phosphate by ³¹P-nuclear magnetic resonance spectroscopy*, Cereal Chem. **70** (1993), 145–152
- [179] A. Blennow, S. B. Engelsen, L. Munck, and B. L. Møller, *Starch molecular structure and phosphorylation investigated by a combined chromatographic and chemometric approach*, Carbohydr. Polym. **41** (2000), 163–174
- [180] G. Ritte, M. Heydenreich, S. Mahlow, S. Haebel, O. Kötting, and M. Steup, *Phosphorylation of C6-and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases*, FEBS Lett. **580** (2006), 4872–4876
- [181] S. Tabata and S. Hizukuri, *Studies on starch phosphate. Part 2. Isolation of glucose-3-phosphate and maltose phosphate by acid hydrolysis of potato starch*, Starch/Stärke **23** (1971), 267–272
- [182] A. Blennow, T. H. Nielsen, L. Baumsgaard, R. Mikkelsen, and S. B. Engelsen, *Starch phosphorylation: a new front line in starch research*, Trends Plant Sci. **7** (2002), 445–450
- [183] C. Edner, J. Li, T. Albrecht, S. Mahlow, M. Hejazi, H. Hussain, F. Kaplan, C. Guy, S. M. Smith, and M. Steup, *Glucan, Water Dikinase Activity Stimulates Breakdown of Starch Granules by Plastidial beta-Amylases*, Plant Physiol. **145** (2007), 17–28
- [184] H. W. Maurer and M. D. Laurel, *Opportunities and Challenges for Starch in the Paper Industry*, Biol. Chem **54** (1990), 2793–2799
- [185] U. Heinze, D. Klemm, E. Unger, and F. Pieschel, *New starch phosphate carbamides of high swelling ability: Synthesis and characterization*, Stärke **55** (2003), 55–60

- [186] A. Blennow, B. Wischmann, K. Houborg, T. Ahmt, K. Jørgensen, S. B. Engelsen, O. Bandsholm, and P. Poulsen, *Structure function relationships of transgenic starches with engineered phosphate substitution and starch branching*, *Int. J. Biol. Macromol.* **36** (2005), 159–168
- [187] A. Blennow, A. M. Bay-Smidt, C. E. Olsen, and B. L. Møller, *The distribution of covalently bound phosphate in the starch granule in relation to starch crystallinity*, *Int. J. Biol. Macromol.* **27** (2000), 211–218
- [188] A. Blennow, K. Houborg, R. Andersson, E. Bidzinska, K. Dyrek, and M. Labanowska, *Phosphate positioning and availability in the starch granule matrix as studied by EPR*, *Biomacromolecules* **2006** (2006), 965–974
- [189] S. B. Engelsen, A. Ø. Madsen, A. Blennow, M. S. Motawia, B. L. Møller, and S. Larsen, *The phosphorylation site in double helical amylopectin as investigated by a combined approach using chemical synthesis, crystallography, and molecular modeling*, *FEBS Lett.* **541** (2003), 137–144
- [190] D. B. DuPré and M. C. Yappert, *Conformational simulation of phospholipids by molecular mechanics*, *Journal of Molecular Structure: THEOCHEM* **467** (1999), 115–133
- [191] S. Crouzy, F. Fauvelle, J. C. Debouzy, M. Göschl, and Y. Chapron, *Investigation of the α -cyclodextrin-myoinositol phosphate inclusion complex by NMR spectroscopy and molecular modeling*, *Carbohydr. Res.* **287** (1996), 21–35
- [192] C. L. Janssen, I. B. Nielsen, M. L. Leininger, E. F. Valeev, and E. T. Seidl, *The Massively Parallel Quantum Chemistry Program (MPQC), Version 2.3.1*, Sandia National Laboratories, Livermore, CA, USA (2004)
- [193] S. Pérez, M. Kouwijzer, K. Mazeau, and S. B. Engelsen, *Modeling polysaccharides: Present status and challenges*, *Journal of Molecular Graphics* **14** (1996), 307–321
- [194] R. E. Hoffman, J. C. Christofides, D. B. Davies, and C. J. Lawson, *Scope and limitations of the simple ^{13}C -N.M.R. method of structural analysis of carbohydrates: glucodisaccharides*, *Carbohydr. Res.* **153** (1986), 1–16

- [195] M. Martin-Pastor and C. A. Bush, *The use of NMR residual dipolar couplings in aqueous dilute liquid crystalline medium for conformational studies of complex oligosaccharides*, Carbohydr. Res. **323** (1999), 147–155
- [196] M. Martin-Pastor and C. A. Bush, *Conformational studies of human milk oligosaccharides using ^1H - ^{13}C one-bond NMR residual dipolar couplings*, Biochemistry **39** (2000), 4674–4683
- [197] A. Almond, J. Bunkenborg, T. Franch, C. H. Gotfredsen, and J. Ø. Duus, *Comparison of aqueous molecular dynamics with NMR relaxation and residual dipolar couplings favors internal motion in a manose oligosaccharide*, J. Am. Chem. Soc **123** (2001), 4792–4802
- [198] S. Bekiroglu, L. Kenne, and C. Sandstrom, *^1H NMR studies of maltose, maltoheptaose, alpha-, beta-, and gamma-cyclodextrins, and complexes in aqueous solutions with hydroxy protons as structural probes*, J. Org. Chem. **68** (2003), 1671–1678
- [199] P. E. Jansson, L. Kenne, and G. Widmalm, *CASPER: a computer program used for structural analysis of carbohydrates*, Journal of Chemical Information and Computer Sciences **31** (1991), 508–516
- [200] G. M. Lipkind, A. S. Shashkov, Y. A. Knirel, E. V. Vinogradov, and N. K. Kochetkov, *A computer-assisted structural analysis of regular polysaccharides on the basis of ^{13}C -N.M.R. data*, Carbohydr. Res. **175** (1988), 59–75
- [201] M. Shirts and V. S. Pande, *COMPUTING: Screen Savers of the World Unite!*, Science **290** (2000), 1903–1904
- [202] S. M. Larson, C. D. Snow, M. Shirts, and V. S. Pande, *Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology*, Computational Genomics (2002), 1–31

Paper I

P. I. Hansen and S. B. Engelsen, *Modeling, Simulation and Visualization of the Hydration of Starch Glucans*. Starch - Biology, Structure and Functionality, Renewable Resources, Wiley and Sons Ltd. 2008. Editors: Anton Huber, Werner Praznik, ISBN-10: 0470014628, ISBN-13: 978-047001462. **In press**

Starch - Biology, Structure and Functionality

Modeling, Simulation and Visualization of the Hydration of Starch Glucans

Peter Ibsen Hansen and Søren Balling Engelsen

Quality & Technology, Department of Food Science, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark

Introduction

This chapter presents a short survey of current methods used to analyze and visualize the structures, hydration and properties of α -glucans in aqueous solution. The emphasis will be on molecular dynamics simulations of starch α -glucan models. While physical-chemical methods are only able to provide time-averaged data on the hydration of starch glucans, molecular modeling is the only feasible method for obtaining insight into the structural mechanisms and dynamics of the hydration at the atomic level.

Besides being the most important energy source in food and feed, starch is also used extensively as an ingredient to manipulate the quality of our food and as an inexpensive, versatile, renewable and biodegradable polymer in a wide range of material applications, such as a thickener and stabilizer for controlling consistency and as a texture enhancer¹. Irrespective of food, food ingredient or material application, practically all aspects of starch functionality are related to the hydration of the starch structure. In order to fully exploit the potential of starch material and ingredient applications an improved understanding of starch-water interactions at the molecular level is a prerequisite. Only when we understand more about the complex

relationships between the hydration of starch structure and its functionality can new starch materials with well-defined functionalities be designed, making it possible to manufacture foods and degradable and renewable biomaterials with improved functionalities².

Natural starch is found packed into dehydrated granules (containing about one water molecule per glucose unit) with origin-specific shape and size. When starch granules are soaked or cooked in water they swell, which is an important functional property of the starch complex and largely determined by the granule size distribution. The gelling abilities of starch rely on absorption of water and leakage of amylose during cooking. When the mixture is cooled, the extracted amylose forms a gel network. Starch gels thus consist of swollen (amylopectin-rich) granules held in an amylose-water network and the gel functionality depends on the state of the swollen granules as well as on the amylose network. Upon cooling and storage, starch gels normally become firmer (retrogradation), often associated with syneresis. This is due to precipitation of dissolved amylose onto the amylose network, which then becomes thicker.

Starches isolated from different botanical sources are known to have different molecular structures, resulting in a wide range of functionalities. The differences in functionality can be attributed not only to the morphology and size of the starch granules, but also to the assembly and structure of the starch molecules within the starch granules³. However, there is a rising demand for new, improved or tailored functional properties of starch. Much attention has been given to the ratio between amylose and amylopectin when considering starch functionality, since this variable has profound effects on starch paste rheology, as shown for amylose-free potato starch⁴ and high-amylose starch⁵. Other structural parameters, such as the starch molecular

weight distribution ⁶ and the degree of amylopectin branching, are also known to have specific and important impacts on the functional properties of starches⁷. The only non-glucose modification of starch found in nature is the phosphorylation found in starches from tuberous plants. They are unique by having covalently bound phosphate esterified to a relatively large proportion (approximately 0.5 %) of the glucose residues in the amylopectin. A number of studies have (now) shown that many thermal and rheological properties of potato starch are related to the degree of phosphorylation⁸.

Computer modeling of starch glucans

- Introduction

In recent years it has become commonly accepted that molecular modeling by computer simulation is the only available method to gain insight into the dynamic behavior of molecules below the nanosecond time scale. There exist two very different approaches to carry out simulations of molecular systems. One approach is the so-called *ab initio* calculations which are based on the theory of quantum mechanics. However, due to the immense calculation efforts required, we are not likely to be able to perform high (theory) level *ab initio* calculations of larger α -glucan models such as amylose and amylopectin in the foreseeable future.

A more pragmatic method is molecular mechanics, which is an empirically based method that uses a number of approximations to provide a simpler model to simulate the conformational and dynamical properties of molecular structures. The simplifications applied in molecular mechanics make it possible to simulate dynamics of molecular systems on a time scale long enough to give a fairly good reproduction of experimental physical observables, rendering molecular mechanics a practical tool for studying the structure and dynamics of

polysaccharides including α -glucans. We will briefly outline the method and give examples of the application to starch molecules.

- Theory of molecular mechanics

While the main focus in quantum mechanics is on the distribution of electrons, protons and neutrons of a molecular system, molecular mechanics simplifies molecular calculations one order of magnitude further by assuming that the atoms are the basic constituents of a molecular system as a necessary simplification in order to facilitate simulation of sufficiently large systems.

Central for the computations is the definition of a potential energy function relating the molecular structure (\vec{R}) to the energy of the system investigated (U). The combination of a potential energy function and a set of physical parameters used with that function constitutes what is called a force field. An example of a potential energy function is given in a recent review by Mackerell⁹.

$$U(\vec{R}) = V_{bonds} + V_{angles} + V_{dihedrals} + V_{impropers} + V_{nonbond}$$

In the terms concerning bonds, angles and impropers the energy is normally described by harmonic potentials. The harmonic bond potential corresponds to a set of particles combined by a set of mutually independent springs or harmonic oscillators. In this potential each particle exhibits harmonic oscillations around the equilibrium state. The bond energy term has the form:

$$V_{bonds} = \frac{1}{2} \sum_n K_b (r_n - r_o)^2$$

This quadratic potential gives an approximation of the strain energy caused by separating two bonded atoms at a distance r_n , where r_o

represents the separation at the potential equilibrium energy for the specified bond type (eg. C-H) and K_b is the force constant.

The non-bonded interactions in molecular mechanics are divided into two categories: The van der Waals potential and the electrostatic Coulomb potential. In regions where the molecule is hydrophobic the dominating potential is the van der Waals potential which attracts other neutral objects. In regions where the molecule is hydrophilic the Coulomb potential is most significant and it attracts other polar objects. The van der Waals potential consists of a repulsive term, which accounts for the excluded volume, and an attractive term which accounts for dispersion attraction. The van der Waals potential is most widely adapted to molecular mechanics such as the Lennard Jones 12-6 potential which has the form:

$$V_{nb} = \sum_i^N \sum_{j>i} \rho^4 \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_n} \right)^{12} - \left(\frac{\sigma_{ij}}{r_n} \right)^6 \right]$$

Where r_n is the non-bonded distance, ϵ is the well depth and σ is the van der Waals (or Lennard Jones) diameter. ρ is a non-bonded intramolecular scaling factor ¹⁰.

According to the classical theory of electrostatics, interactions between diffuse electron clouds are calculated by a series expansion differentiating between charge-charge, charge-dipole and dipole-dipole interactions. For interactions between ions, the Coulomb potential is the leading term:

$$V_{Cou} = \sum_i^N \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon r_{ij}}$$

where r is the distance between the charges, q is the charge and ϵ is the effective dielectric constant. In molecular mechanics, the polarization of a molecular system is either modeled by bond dipoles or by fractional charges located in the center of the atoms. It is in the non-bonded

electrostatic potential we find the most fundamental approximations in molecular mechanics. Firstly, they are modeled by effective pair-potential, neglecting three-body and higher interactions. Secondly, they are assumed to work on both intra- and intermolecular interactions, although they have been derived from strictly intermolecular (ionic) experiments. Thirdly, as molecular charges are not explicitly taken into account, all molecular charges are assumed distributed as fractional charges located in the nucleus of each atom. This approximation is known to be wrong, which is shown experimentally by the differences between the measured positions of hydroxyl hydrogen in x-ray diffraction (electron density) and neutron diffraction (nuclei positions) where the hydrogen electron density is located between the two nuclei. Finally, there is the controversial problem with the dielectric constant, which is basically a measurement of the modification of the potential made by the intervening medium. While the dielectric constant is well-defined on a molecular basis in surrounding media like vacuum and solutes, problems may arise when considering the intramolecular (or interatomic) dielectric constant.

To correct for some of the inadequacies in the non-bonded potential energy functions it is necessary to add a number of intramolecular correction potentials which might be implemented in a very similar manner to the bonded potential. The valence angle potential is included to correct for geminal repulsions for which the van der Waals potential gives an inadequate description. The torsional angle potential is included to enable hindered rotation of groups around a bond. A correction term has become necessary, since van der Waals interaction large enough to create hindered rotation about sp^3 - sp^3 bonds will create unphysical behavior in condensed phase simulations. In order to create hindered rotation about sp^2 - sp^2 bonds the torsional potential is needed to maintain planarity. The out-of-plane correction potential is included to enforce planarity about sp^2 -hybridized groups.

When combined and properly parameterized the potential energy functions constitute a molecular mechanics *force field* in which structures (Cartesian coordinates and molecular topology) of a given class of molecular systems, e.g. the α -glucans, can be evaluated energetically. The choice and parameterization of the potential energy functions are crucial for the end performance of a given force field. Due to the limited amount of experimental data on carbohydrates with atomic resolution the parameterization normally includes data obtained on adequate model substances such as the hydrocarbons, the ethers and the alcohols and in this context such molecular mechanics can be considered as an extrapolation technique.

- Available empirical force fields for glucans

Development of empirical force fields for modeling carbohydrates has proven to be a challenging task and the evolution of this area has been lagging behind others classes of compounds such as peptides and nucleic acids. The main challenges are the flexible nature of the carbohydrate pyranose ring structure and the mutarotation due to ring opening which makes redistribution of the electron system possible, causing multiple conformational changes and preferences due to anomeric, exo-anomeric and gauche effects. Adding to the complexity is the rich possibility to establish hydrogen-bonding networks of the hydroxyl groups with structural water molecules and the rotational freedom of the glycosidic linkages.

All atoms force fields designed for simulation of carbohydrates are currently available for most of the commonly used software simulation packages. For the CHARMM package¹⁰, models for the hexopyranoses have been developed from crystallographic and vibrational spectroscopy data¹¹. This force field was later improved and re-optimized by Brady and coworkers¹² which resulted in an improved model. Multiple force fields for carbohydrates have also been developed for AMBER¹³ of which

the GLYCAM force field¹⁴ is widely used. Also the GROMOS¹⁵ force field has been extensively used for modeling carbohydrates including a force field refinement developed using maltose¹⁶. An extension of the OPLS¹⁷ force field has been developed for carbohydrates¹⁸, and is based on optimization of torsional parameters using quantum mechanical calculation of the conformations of a series of monosaccharides. The OPLS force field can be used with the freely available Gromacs¹⁹ software package.

It is generally difficult to decide on which force field to apply to a given molecular system, as all of them have different properties and will provide at least slightly different answers to the questions under investigation. As an example, it was found in the force field comparison study by Corzana et al.²⁰ that in a group of tested force fields one was better at reproducing the molecular diffusion, whereas the others yielded better reproduction of molecular structure when compared to experimental data from NMR spectroscopy. An extensive comparison of the *in vacuo* performance of different carbohydrate force fields were made by Perez et al.²¹ which was followed by a quantum mechanical comparison 'validation' study by Hemmingsen et al.²².

- Molecular dynamics

In addition to rotational and vibrational motions, molecules also exhibit thermal motions, and when a molecular mechanics force field is at hand, it is possible to assign velocities to the molecular system under investigation and to integrate Newton's equations of motion for the system over a (short) period of time. It is the most important task of molecular dynamics to calculate and analyze thermal motions. The molecular dynamics method concept is to calculate the displacement coordinates in time (a trajectory) of a system at a given temperature. Mathematically, molecular dynamics correspond to solving a set of coupled second-order differential equations:

$$\underline{F}(r) = \sum_i^N m_i \frac{\partial^2 r_i}{\partial t^2} = -\nabla_i V(r_i)$$

where r is the positional vectors of the atoms, F is the forces acting on the system and m_i the atomic masses. In molecular dynamics it is important to 'digitize' the trajectories with step sizes as large as possible. This is normally done by the so-called Verlet algorithm²³ which is based on two assumptions: that the average velocity $\langle v \rangle$ and the average of the acceleration $\langle a \rangle$ are equal to the instant values in the mid-point ($v(t+\Delta t/2)$ and $a(t+\Delta t/2)$) of the integrating interval (Δt). In the original formulation:

$$x_i(t + \Delta t) = x_i(t) + v_i(t)\Delta t + a_i(t)\frac{1}{2}\Delta t^2$$

the theoretical instability limit is expressed by: $\Delta t = \frac{\omega_{\min}}{\pi}$ where ω_{\min} is the period length of the fastest motion in the system, which is usually the N-H bond stretch, having a frequency of approximately 3600 cm^{-1} which corresponds to a largest possible step of about 2 *femto* seconds. The use of a step size in the femtosecond regime has some very important implications when studying atomic motions, as the calculation of a single trajectory in one microsecond involves 1 billion evaluations of the potential energy functions using the Verlet algorithm, for which reason one cannot overestimate the importance of optimizing the speed of the function evaluation.

The main advantage of molecular dynamics is the possibility to simulate liquids and molecules dissolved in water or other solvents. If the objective is to simulate a continuous condensed phase, the central asymmetric unit of the molecular system under investigation is repeated in the axis directions in order to minimize special surface effects. The most efficient method to accomplish this is called *minimum image convention periodic boundary conditions*²⁴. It enables us to explicitly

dissolve α -glucan molecules in a box with water molecules provided that we also have a force field for water.

The result from molecular dynamics simulations is a time series of coordinates for all atoms included in the molecular system under investigation. In this context our goal is to extract molecular properties like dynamics of the inter-residue linkage, conformational changes of hydroxyl groups and last, but not least, the motion of the water molecules involved in solvating the α -glucan solute. To achieve this goal the data has to be reduced to meaningful quantities which can be related to or interpreted in terms of molecular properties. The complexity of the data calls for statistical data analysis methods where pair distribution functions and static water density calculations are approaches that can facilitate (ways of) data interpretation.

Application of molecular mechanics to α -glucan molecules

The two macromolecular constituents of starch granules are amylopectin and amylose which are both composed exclusively of glucose residues linked by only two types of bonds: α -1,4 and α -1,6 glucosidic linkages. Amylopectin (typically ~75%) is a semi-crystalline highly branched polysaccharide with an α -1,4 backbone and 4-5% α -1,6 branch points while amylose (typically ~25%) is amorphous in the native starch granule and is composed of essentially linear chains of α -1,4 linked glucose units. In spite of the simple chemistry of starch, the final starch molecules are variable and complex and much too large for an atomistic approach. Amylopectin has a molecular weight of approximately 10^{10} Dalton which can be translated into glucose units by dividing with 180 Dalton which each have 24 atoms (3 cartesian each) Using a single precision 4 bytes storage this adds up to a storage demand for one single conformation of amylopectin of 15 Giga Bytes. For the latter

reason we will in the following exemplify starch by α -glucan model compounds.

- **Glucose, the fundamental building block of starch**

Starch is a polymer of one specific monomer: α -D-glucopyranose, or for short, glucose. The glucose ring contains 12 hydrogen atoms, 6 carbon atoms and 6 oxygen atoms, linked with covalent bonding into the structure shown in **Figure 1**. The molecular structure of glucose is well known, and we will use this information to build larger α -glucan fragments.

The cyclic glucose hemiacetal ring is formed by a closing of the pentahydroxyhexanal straight chain molecule. Upon formation the ring closing can result in two possible conformers which are diastereomers, also referred to as *anomers*. The C1 atom of the ring becomes an anomeric center where the two possible conformations differ by the configuration of the hydroxyl group which can be arranged either cis or trans to the alkoxy substituent at the C5 atom. The two anomers are referred to as the *alpha* and *beta* form, where alpha corresponds to the trans and beta to the cis isomer. The conversion between the two enantiomers occurs spontaneously in aqueous solution and is commonly known as *mutarotation*. The only enantiomer found in the starch polymer is the alpha isomer. For the glucose monomer, the preferred conformation in aqueous solution is the beta isomer which accounts for about 64% due to the anomeric effect²⁵. So, when synthesizing model compounds in the laboratory to mimic starch structures it is common practice to methylate the hydroxyl on C1 in alpha position and thus prevent mutarotation.

The α -D-glucopyranose ring structure is most stable in the chair conformation denoted 4C_1 which means that C4 is situated at the top of the chair and the C1 at the lower point compared to the plane defined by O5-C5---C2-C3.

while glucose has been examined by molecular mechanics methods since the 1970's when Kildeby et al.²⁶ developed the first force field for glucose, the properties of α -D-glucopyranose in an explicit aqueous environment were first examined with molecular dynamics in 1989 by J. W. Brady²⁷. In that study Brady found that the 4C_1 crystal ring structure was stable and only slightly affected by the solvation in water.

When considering the hydration of this fundamental α -glucan building block it is striking that it is found hydrated (monohydrate) even in the crystallized form²⁸.

- The glucosidic $\alpha(1\rightarrow4)$ linkage in the starch polymer

If the starch molecule is broken down into dimers, only two different dimers will be the result. Maltose and iso-maltose, as shown in **Figure 2**, serve as primary models for the two types of glycosidic linkages found in starch. As the α -D-glucopyranose ring is usually assumed to be a fairly rigid structure, the torsional angles of the linkages between units are the primary parameters that define the overall structure of an α -glucan oligomer²⁵.

In maltose we define the torsional angles of the α -1-4 linkages as ϕ and φ , where the ϕ angle is defined by the four atoms O5-C1-O1-C4' and φ by C1-O1-C4'-C5'. The α -1-6 linkages which are found for example in iso-maltose have one additional glycosidic torsion ω that is given by O1-C6'-C5'-O5'. The preference of the ϕ torsion, the angle between substituents on C1 and O1, is strongly influenced by the exo-anomeric effect which is mainly a stereoelectronic effect caused by the polarity of both the substituent and the electron lonepair on the linkage oxygen. The φ torsion is mostly influenced by steric effects and hydrogen bonding both interresidual and with the surrounding water. The ω torsion usually prefers 3 different configurations referred to as

gauche-trans (gt, $\omega=180^\circ$), gauche-gauche (gg, $\omega=60^\circ$) and trans-gauche (tg, $\omega=-60^\circ$). The tg rotamer is rarely seen, because repulsive interactions make this conformation energetically unfavorable in comparison to gt and gg.

The parameters which have the greatest influence on the molecular conformation of the starch polymers are the torsional angles used to describe the configuration of the glycosidic linkages. The α -(1 \rightarrow 4) linkage has two angles and the α -(1 \rightarrow 6) linkage has three angles which allows some freedom for rotation. The energetics of glycosidic linkages can be explored by calculating energies for all possible rotational conformations of the two monomers on each side of the linking oxygen atom O1, and plotting these in a two-dimensional energy map as a function of Φ and Ψ will give a first hint at what the preferred torsional angles are. This can then be compared to average values from molecular dynamics trajectories and in the end evaluated by comparison to experimental data from techniques such as x-ray diffraction or NMR spectroscopy. Such a potential energy map has been calculated for maltose in the MM3 force field and is shown in **Figure 3**.

The adiabatic map reveals one global energy minimum well centered at $\Phi=100$ and $\Psi=220$ and a secondary local minimum well centered at approximately $\Phi=100$ and $\Psi=70$. The most important observations we can elucidate from this archetype α -glucan linkage is:

1. That the geometry of the α -maltose in the crystal structure²⁹ (see **Figure 3**) is found in the center of the global minimum well
2. That the global energy minimum well is favored by an intramolecular (interring) hydrogen bond between O-2. . .O-3' (see **Figure 3**)
3. And that the structure of the α -glucan linkage in the global energy well, if extrapolated to a polymer, shows a clear

preference for left-handed helical models with 6 glucose units per helical turn³⁰.

The latter observation strongly indicates that 1→4 linked α -glucans will naturally favor a helical arrangement with 6 glucose residues per helical turn in the isolated state.

Hydrogen bonding between hydroxyls on the solute, inter- or intraresidual, can change or stabilize the molecular structure, so that it assumes a conformation that is different from what one could predict from an energy minimization of an isolated molecule in vacuum. Thus, The next obvious question is whether hydration will affect these conformational preferences. Several studies have investigated the conformational preferences of α -maltose in aqueous solution^{20;31-33} and the conclusion is that the structure of the maltosidic unit is only slightly altered in aqueous solution. As an example, **Figure 4** shows the so-called population density map of methyl- α -maltoside superimposed on the outer contours of the adiabatic map of α -maltose from **Figure 3**. The figure indicates a slight shift to the left of the global energy minimum well into a region (A) which is similar in energy to minimum B, but does not promote the above-mentioned intramolecular hydrogen bond. The latter makes good sense, as the hydrogen bond now is in competition with hydrogen bonding to the water molecules. However, although it is often seen that the presence of water induces a noticeable conformational shift in the conformation of the carbohydrate³⁴, it is noteworthy that the overall structure and thus the tendency to generate a six-fold helical structure is conserved for $\alpha(1\rightarrow4)$ -glucans in aqueous solution.

A molecular dynamics study of maltohexaose³⁵ showed only small deviations from the V-amylose helical structure. Analysis of the solvent revealed that the water molecules form bridging hydrogen bonds between glucose residues which 'lock' the amylose fragment into the

helix conformation. The favored low energy conformations observed for maltose are thus preserved in maltohexaose with additional stability maintained by intermolecular hydrogen bonding with the solvent. This result corresponds to extrapolation of the most populated conformation of maltose into a helical structure, and the increased stability of maltohexaose explains the low solubility of amylose which would not be expected from the flexibility of maltose.

- The $\alpha(1\rightarrow6)$ branch point in the starch polymer

The other glucosidic linkage present in starch is the $\alpha(1\rightarrow6)$ linkage which defines the branch points in amylopectin. This linkage can be studied in exactly the same manner as the $\alpha(1\rightarrow4)$ linkage, but as it has three bonds, it is one order of magnitude more complex. The usual way to deal with the three-dimensional problem is to divide the ω -angle into three staggered slices $\omega = 300^\circ$ (gg), $\omega = 60^\circ$ (gt) and $\omega = 180^\circ$ (tg) (see **Figure 5**). All the wells have Φ in the proximity of 80° in accordance with the exo-anomeric effect and Ψ has a value of 180° in the three lowest energy wells. Again the crystal conformation, in this case from the geometry of the $\alpha(1\rightarrow6)$ linkage found in the crystal structure of α -panose³⁶ ($\Phi = 71^\circ$, $\Psi = 165^\circ$, $\omega = 75^\circ$), is positioned nearly perfectly, centered in the lowest energy minimum. And just as in the case of the $\alpha(1\rightarrow4)$ linkage of maltose, the hydration does not affect the overall structure of the iso-maltose unit. In fact, it would appear that the aqueous solvation further restricts the $\alpha(1\rightarrow6)$ to the α -panose overall geometry.

In an effort to further study and scrutinize the hydration behavior of the amylopectin branch point, a pentasaccharide was synthesized and investigated by Corzana et al.³⁷ as a model compound that could be subjected to a combined molecular dynamics and NMR spectroscopic study.

Apart from the branch point, one glucose residue was included as a representative of a linker into the amorphous layer of starch. The results in this study all suggest that the conformational flexibility of the α -(1 \rightarrow 6) branch point in water is low, especially when compared with results on a tetrasaccharide analogue. Apparently, it is the addition of the fifth glucose residue that elongates the α -(1 \rightarrow 6) branch that restricts the conformational diversity of the three-bond α -(1 \rightarrow 6) linkage in aqueous solution. The apparent rigidity of the branch point could be explained by the presence of localized water densities or water bridges across the branch point that have a determining influence on the preferred conformation.

This investigation reveals two plausible starch branch point structures, one that permits the formation of double helices and one that is adapted for interconnection of double helices. The apparent rigidity of the former is explained by the presence of water pockets/bridges in the vicinity of the branch point that lock the pentasaccharide structure into one conformational family that is able to accommodate the creation of the double-helical amylopectin structure. The conclusion of the study was that the possible structures of the α -(1 \rightarrow 6) linkage are limited to only two low energy conformations in the presence of water, and that one of these agrees with the geometry necessary for the formation of a double helix³⁸ (**Figure 6**), whereas the other is suited for the interconnection of double helices. The stability of the first-mentioned conformation was explained by the molecular dynamics simulation to be caused by a bridging water molecule. This hydration feature was discovered by calculating 2-site radial pair distribution functions. The normal 1-site radial pair distribution gives the probability of finding a pair of atoms at a distance r apart, relative to the probability for the situation at the same density as where the distribution is random:

$$g(r) = \frac{N(r)}{\rho_{\text{water}} \times \frac{4}{3} \pi \times [(r + \Delta r)^3 - r^3]}$$

Radial pair distribution is useful when studying the carbohydrate hydration by calculating the structure of water oxygen atoms around any given oxygen in the solute. The maxima and minima of the distribution will define limits of areas where the water density is higher and lower with respect to the density of bulk water, and these limits are then interpreted as divisions of space into layers which are referred to as hydration shells. These hydration shells show how the carbohydrate induces a rearrangement of the water structure with the purpose of creating an optimal hydrogen-bonding network adapted to the new environment. The hydration number is the number of water molecules in the first hydration shell of a solute. It has been shown in a force field comparison study²⁰ that different force fields give a noticeable difference for the location of hydration shells.

It has also been shown that specific water molecules can reside in a fixed position for longer periods of time due to hydrogen bonding between the water molecule and one or two oxygen atoms in the carbohydrate structure³⁹. One tool for the description of such localized waters interacting closely with the carbohydrate is so-called 2-site radial pair distributions⁴⁰. This method gives a statistical measure for the probability of finding, for example, an oxygen atom at a pair of distances (r_1 and r_2) from two given atoms in the hydrated solute, relative to the expected probability for a random distribution. The result is a contour plot where the water probability is plotted as a function of distances to the two hydration sites.

In the case of the pentasaccharide, the 2-site radial pair distribution function between O-2 (residue B) and O-5 (residue D) revealed an anisotropic water density of 8.5 (**Figure 7**) which is unusually high, but similar in magnitude to the shared water between O-2g and O-1f

reported for sucrose³⁴. In the pentasaccharide, this water bridge between O-2(B) and O-5(D) is present about 73% of the time and probably explains the reduced flexibility of the branch point of the pentasaccharide when compared to the observations made for the tetrasaccharide 6''- α -D-glucopyranosyl-maltotriose⁴¹. This result is very interesting, as it indicates that extending the α -(1 \rightarrow 6) branch (D-B-A) by one glucose unit (A) paves the way for a water bridge between O-2(B) and O-5(D) that "locks" the branch point structure into a conformation that ultimately will lead to double-helical formation between the two branches (E)-D-C and D-B-A. Moreover, it suggests that this water molecule should be present as structural water in the limiting region of the amylopectin crystalline region.

- **Double-helical structure of the amylopectin crystalline domain**

The starch granule is believed to be built from radially packed clusters arranged in a structural form similar to blocklets with a crystalline ordering of double-helical elements. Investigations of A- and B-type starches from cereals suggest that the structural elements are left-handed double helices arranged in parallel strands^{42,43}. The parallel helices are able to pack with a short distance due to a relative translation along the helical axis and thus form crystalline domains with a hexagonal or pseudo-hexagonal packing symmetry. Although the amylopectin molecule at present is inaccessible to molecular mechanics methods we can start building double-helical amylopectin fragments and begin to construct microcrystalline domains of nanocrystals (**Figure 8**). The branch point of amylopectin has been investigated with molecular modeling methods^{37,44}, and is believed to be a key stabilizer facilitating the formation, arrangement and stability of the crystalline domains.

Possibilities for the structural arrangement of adjacent amylopectin double helices including the branch point have been investigated thoroughly³⁸ using a molecular polysaccharide builder program POLYS⁴⁵ and

some of the lowest energy structures suggested by this approach were found to agree with previously reported crystal data. A thorough study of the structure of the part of the polymer that acts as a linker between the branch points starting double helices has been performed via molecular modeling³⁸, yielding the interesting result that the internal chain lengths are important for the degree of local crystallinity, because only certain chain lengths lead to parallel double helices.

This type of information is a prerequisite for future studies targeting helix-helix interactions in starch 2D nano-crystals⁴⁶ and when studies begin to concentrate on stacking and modeling of 2-D nano crystallites with realistic amorphous regions.

- Phosphorylation

Almost all starches found in different plant species are phosphorylated to some degree⁴⁷. In the highest substituted potato starches reported, one of 200 glucose residues is substituted. The most highly substituted natural starch currently known has a degree of substitution of 1%. The phosphate groups are bound as mono-esters linked to the C-3 and C-6 positions of the glucose residues. Long unit chains are more highly substituted than short ones and C-6 phosphate substitution is more pronounced in amorphous regions. The effect of phosphate groups being present in starch is an increased hydration capacity of starch pastes after gelatinization⁴⁸ which helps to prevent crystallization. Highly phosphorylated starches are generally used to control the viscosity in industrial starch products⁴⁹.

In order to study whether phosphorylation is possible in the amylopectin double-helical packing the phosphate group has been investigated by means of molecular modeling⁵⁰ and incorporated into the optimized double-helical structure⁵¹. The results of such efforts are rather unambiguous and show that a C-3 phosphorylation of the double

helices points directly out from the helix coil and is bound to disrupt double-helical hexagonal packing (**Figure 9**). In contrast, phosphorylation on C6 takes place in a surface groove and therefore does not seem to perturb the double-helical amylopectin structure, as it does not create steric problems for the hexagonal helix-helix packing.

- *ab initio* calculations

During the past decade it has been demonstrated that *ab initio* and density functional (DFT) methods can produce reliable results for small model carbohydrates. However, α -glucan molecules, even glucose, are too large to allow for highly correlated quantum mechanical methods to be applied together with large basis sets. Since hydration, and hydrogen bonding in particular, is an important factor for these systems, the α -D-glucopyranose·H₂O system (see Figure 1) is a convenient model system. While the water-water dimer can be evaluated at the highest theoretical level using coupled cluster *ab initio* methods (CCSD(T)) and large basis sets (dAUG-cc-pVQZ) to 5.4 ± 0.7 kcal/mol⁵², the interaction energy of the α -D-glucopyranose·H₂O system requires a truncated glucose molecule in order to be evaluated at the highest theoretical level. Such an approach resulted in an estimation of the glucose water interaction to be 4.9 ± 1 kcal/mol, quite similar to the water-water interaction energy.



The quantum mechanical truncation of the molecule did not have a great effect on the calculated interaction energy, indicating that the most important hydrogen-bonding interactions are represented in the smallest localized system. This is also in agreement, but on the lower limit of hydration energies of monohydrates calculated to be between -5 and -12

kcal/mol using density functional theory (DFT) method B3LYP/6-311++G**

⁵³.

Recently, it has also become possible to calculate relaxed potential energy surfaces of disaccharide analogs including maltose using the relatively low theoretical HF/6-31G* level of theory in which the glucose units are replaced with tetrahydropyran⁵⁴. While the relevance of such computational efforts may be questioned, they prove to account fairly well for conformations that are observed in crystals of the parent disaccharides, which in turn indicates that intramolecular hydrogen-bonding schemes are of minor importance. In the case of the potential energy surface of the maltose analog, significant deviations of the positions of glycosidic linkages conformation in crystal conformations from the global minimum were explained by the fact that many of these come from macrocyclic and thus strained structures (cyclodextrins). However, the general conclusion of such massive computational efforts is the quest for more highly correlated quantum mechanical methods to be applied, for hydration to be explicitly included and more realistic systems to be studied. Starch is far too complex and will in the foreseeable future require more pragmatic theoretical methods to be applied.

- Experimental methods

While simulation methods such as molecular dynamics and others like quantum mechanics calculations are the only feasible route to obtain information in such detail as described in this chapter, experimental data is always the most important source of information. Starch or carbohydrates is a class of compounds for which it is very difficult to obtain structural information by experimental methods. This is mainly due to the diverse nature of the molecules. Even for the starch polymers which are built from only one type of monomer and two types of linkages, the inherent flexibility of the inter-glycosidic bonds, the many possibilities of branching and the mobility of the glucose ring

give so much conformational freedom that the system becomes difficult to model and characterize.

For other types of biological molecules, the primary source of information on the molecular structure is without competition obtained from x-ray scattering of single crystals. In the case of carbohydrates, this method becomes unfeasible when the system grows in size beyond tetra-saccharides. The inherent flexibility of α -glucans and carbohydrates in general represents a formidable barrier to the formation of single crystals and it has therefore been necessary to look for other ways to experimentally determine structural parameters. Currently, nuclear magnetic resonance (NMR) spectroscopy is the best alternative to crystallography for determining molecular structures. The technique has its strength in the liquid phase and also provides some possible ways of obtaining dynamical parameters. With modern NMR spectrometers it is possible to perform thorough investigations on amounts as small as 1 mg of a small oligosaccharide, which is certainly within the limits of what can be produced in synthesis. A range of different NMR experiments within the area of carbohydrate NMR are available for the identification of the many different monosaccharides, but for starch, which is built exclusively from glucose and where the stereo chemistry is known, the main focus lies on information about the tertiary structure. A review of current techniques has recently been given by Duus and coworkers⁵⁵.

The identification of a certain type of glycosidic bonds is possible from the observation of chemical shift values at the glycosylation site. The difference in chemical shifts, when compared to reported standard values, can give a clear indication about which type of link is present in the oligosaccharide. The difficulties arise when a reliable estimate of the geometry is needed. This has to be derived from measurements of long-range scalar couplings across the oxygen bridge, or a sufficient number of Nuclear Overhauser Enhancement constants. Performing this type of measurements on a homo-polymer such

as α -glucans will be increasingly difficult proportional to increasing molecular weight due to spectral overlap which is going to be nearly impossible to resolve. In simple systems where assignment is possible the three bond homo-nuclear COCH scalar coupling constant ${}^3J_{\text{CH}}$ can be measured by Hadamard spectroscopy ⁵⁶ and interpreted by a Karplus-like relation ⁵⁷ to give a measure of the time-averaged geometry of the glycosidic bonds.

From molecular models to starch functionality

Using the theoretical basis explained in this chapter, we will gain new understanding of the basic molecular functionality of starches, including everything from single hydration mechanisms to macromolecular properties. With improved understanding of starch structures and improved technology in the field of genetically modifying plant to produce new starches, combined with chemical and physical modifications, we can further explore the opportunities of starch materials with tailored functionalities. At this point in time, molecular modeling gives insight into only the most basic systems. The challenge of the future will be to develop our simulation techniques to a level where we can handle larger starch models which are more realistic in size and more complex and chemical modifications of those, like different degrees of phosphorylation and variations in the distribution and size of carbohydrate side chains. An ultimate goal of molecular modeling is to predict all of the physical properties earlier described of a given starch polymer from the primary structure of the molecular system.

New functionality has already found its way into applications such as to control food texture with their gelling properties, giving an alternative to the conventional gelatine. In addition, the ability of starches to partially crystallize opens up a field of applications in the field of coating materials. Hopes are that the starches with

improved functionality can replace a range of food components used today and with the perspective of adding increased nutritional value as, for example, with high-amylose starch. The importance of research and further insight at the molecular level concerning important starch traits such as resistant starch and starch self-organising nano-crystallites cannot be underestimated.

Reference List

1. Ellis, R. P.; Cochrane, M. P.; Dale, M. F. B.; Duffus, C. M.; Lynn, A.; Morrison, I. M.; Prentice, R. D. M.; Swanston, J. S.; Tiller, S. A. *J.Sci.Food Agric.* 1998, *77*, 289-311.
2. Franks F. *Water Relationships in Food*; Plenum Press: New York, 1991; pp. 1-19.
3. Swinkels, J. J. M. *Starke* 1985, *37*, 1-5.
4. Visser, R. G. F.; Suurs, L. C. J. M.; Steeneken, P. A. M.; Jacobsen, E. *Starch-Starke* 1997, *49*, 443-448.
5. Banks, W.; Greenwood, C. T.; Muir, D. D. *Starke* 1974, *26*, 289-300.
6. Blennow, A.; Bay-Smidt, A. M.; Bauer, R. *Int.J.Biol.Macromol.* 2001, *28*, 409-420.
7. Fredriksson, H.; Silverio, J.; Andersson, R.; Eliasson, A. C.; Aman, P. *Carbohydrate Polymers* 1998, *35*, 119-134.
8. Viksø-Nielsen, A.; Blennow, A.; Jorgensen, K.; Kristensen, K. H.; Jensen, A.; Møller, B. L. *Biomacromolecules* 2001, *2*, 836-843.
9. Mackerell, A. D. *Journal of Computational Chemistry* 2004, *25*, 1584-1604.
10. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J.Comput.Chem.* 1983, *4*, 187-217.

11. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr.Res.* 1988, *180*, 207-221.
 12. Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J.Comput.Chem.* 2002, *23*, 1236-1243.
 13. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J.Comput.Chem.* 1986, *7*, 230-252.
 14. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J.Phys.Chem.* 1995, *99*, 3832-3846.
 15. van Gunsteren, W. F. and Berendsen, H. J. C. Groningen Molecular Simulation (GROMOS) Program Package. 1987. Groningen, The Netherlands, Nijenborgh 16, University of Groningen.
- Ref Type: Computer Program
16. Ott, K. H.; Meyer, B. *Journal of Computational Chemistry* 1996, *17*, 1068-1084.
 17. Jorgensen, W. L.; Tiradorives, J. *Journal of the American Chemical Society* 1988, *110*, 1657-1666.
 18. Damm, W.; Frontera, A.; Tiradorives, J.; Jorgensen, W. L. *Journal of Computational Chemistry* 1997, *18*, 1955-1970.
 19. Vandrunen, R.; Vanderspoel, D.; Berendsen, H. J. C. *Abstracts of Papers of the American Chemical Society* 1995, *209*, 49-COMP.
 20. Corzana, F.; Motawia, M. S.; Hervé du Penhoat, C.; Pérez, S.; Tschampel, S. M.; Woods, R. J.; Engelsens, S. B. *J.Comput.Chem.* 2004, *25*, 573-586.

21. Pérez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jiménez-Barbero, J.; Poveda, A.; Espinoza, J. F.; van Eijck, B. P.; Johnson, G.; French, A. D.; Kouwijzer, M. L. C. E.; Grootenhuis, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohyd.Res.* 1998, *314*, 141-155.
22. Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohyd.Res.* 2004, *339*, 937-948.
23. Verlet, L. *Phys.Rev.* 1967, *159*, 98-103.
24. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J.Chem.Phys.* 1953, *21*, 1087-1092.
25. Tvaroska, I.; Bleha, T. *Adv.Carbohydr.Chem.Biochem.* 1989, *47*, 45-123.
26. Kildeby, K.; Melberg, S.; Rasmussen, K. *Acta Chem.Scand.* 1977, *A31*, 1-13.
27. Brady, J. W. *Journal of the American Chemical Society* 1989, *111*, 5155-5165.
28. Hough, E.; Neidle, H.; Rogers, D.; Troughton, P. G. H. *Acta Cryst.* 1973, *B29*, 365-367.
29. Takusagawa, F.; Jacobsen, R. A. *Acta Crystallographica Section B - Structural Science* 1978, *34*, 213-218.
30. Imberty, A.; Buléon, A.; Tran, V.; Pérez, S. *Starch/Stärke* 1991, *43*, 375-384.
31. Brady, J. W.; Schmidt, R. K. *J.Phys.Chem.* 1993, *97*, 958-966.

32. Ott, K.-H.; Meyer, B. *Carbohyd.Res.* 1996, *281*, 11-34.
33. Best, R. B.; Jackson, G. E.; Naidoo, K. J. *J.Phys.Chem.B* 2001, *105*, 4742-4751.
34. Monteiro, C.; Hervé du Penhoat, C. *Journal of Physical Chemistry A* 2001, *105*, 9827-9833.
35. Naidoo, K. J.; Kuttel, M. *J.Comput.Chem.* 2001, *22*, 445-456.
36. Imberty, A.; Pérez, S. *Carbohyd.Res.* 1988, *181*, 41-55.
37. Corzana, F.; Motawia, M. S.; Hervé du Penhoat, C.; van den Berg, F.; Blennow, A.; Pérez, S.; Engelsen, S. B. *J.Am.Chem.Soc.* 2004, *126*, 13144-13155.
38. O'Sullivan, A. C.; Pérez, S. *Biopolymers* 1999, *50*, 381-390.
39. Engelsen, S. B.; Monteiro, C.; Hervé du Penhoat, C.; Pérez, S. *Biophysical Chemistry* 2001, *93*, 103-127.
40. Andersson, C. A.; Engelsen, S. B. *J.Mol.Graph.Model.* 1999, *17*, 101-105.
41. Best, R. B.; Jackson, G. E.; Naidoo, K. J. *Journal of Physical Chemistry B* 2002, *106*, 5091-5098.
42. Imberty, A.; Chanzy, H.; Pérez, S.; Buléon, A.; Tran, V. H. *J.Mol.Biol.* 1988, *201*, 365-378.
43. Imberty, A.; Pérez, S. *Biopolymers* 1988, *27*, 1205-1221.
44. Imberty, A.; Pérez, S. *Int.J.Biol.Macromol.* 1989, *11*, 177-185.

45. Engelsen, S. B.; Cros, S.; Mackie, W.; Pérez, S. *Biopolymers* 1996, 39, 417-433.
46. Putaux, J. L.; Molina-Boisseau, S.; Momaur, T.; Dufresne, A. *Biomacromolecules* 2003, 4, 1198-1202.
47. Blennow, A.; Engelsen, S. B.; Munck, L.; Moller, B. L. *Carbohydrate Polymers* 2000, 41, 163-174.
48. Wiesenborn, D. P.; Orr, P. H.; Casper, H. H.; Tacke, B. K. *Journal of Food Science* 1994, 59, 644-648.
49. Ellis, R. P.; Cochrane, M. P.; Dale, M. F. B.; Duffus, C. M.; Lynn, A.; Morrison, I. M.; Prentice, R. D. M.; Swanston, J. S.; Tiller, S. A. *Journal of the Science of Food and Agriculture* 1998, 77, 289-311.
50. Blennow, A.; Engelsen, S. B.; Nielsen, T. H.; Baunsgaard, L.; Mikkelsen, R. *Trends in Plant Science* 2002, 7, 445-450.
51. O'Sullivan, A. C.; Perez, S. *Biopolymers* 1999, 50, 381-390.
52. Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helgaker, T. *Theoretical Chemistry Accounts* 1997, 97, 150-157.
53. Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohyd.Res.* 2004, 339, 553-567.
54. French, A. D.; Kelterer, A. M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J.Comput.Chem.* 2001, 22, 65-78.

55. Duus, J. O.; Gotfredsen, C. H.; Bock, K. *Chemical Reviews* 2000, *100*, 4589-+.
56. Blechta, V.; del Rio-Portilla, F.; Freeman, R. J. *Magn.Reson.Chem.* 1994, *32*, 134-137.
57. Tvaroska, I.; Hricovini, M.; Petrakova, E. *Carbohydr.Res.* 1989, *189*, 359-362.

Figure Legends:

Figure 1. The structure of the alpha-D-glucopyranose unit in the 4C_1 conformation, shown together with the one water molecule which is present in the crystal structure.

Figure 2. Left: Methyl- α -maltoside including the α -1,4-linkage and the definition of Φ and Ψ angles. Right: Methyl- α -isomaltoside including the α -1,6-linkage, and the definition of Φ , Ψ and ω angles.

Figure 3. Adiabatic map of the primary dihedral rotations (Φ and Ψ) of the α -(1 \rightarrow 4)-linkage in α -maltose. Isoenergy contours are drawn at 1 kcal \cdot mol $^{-1}$ increments to 8 kcal \cdot mol $^{-1}$ above the global minimum. The geometry of the crystal structure of maltose (shown to the right) is indicated with a C. The crystal structure include an interring hydrogen bond between O-2 and O-3' (reducing end).

Figure 4. The population density map of methyl- α -maltoside superimposed on the outer contour of the adiabatic map of α -maltose. Contours are drawn at (0.1, 0.01, 0.001, and 0.0001) population levels. In this case the geometry of the glycosidic linkage is calculated for every time step in the molecular dynamics trajectory and the occurrence added to a histogram to form a mesh that can be plotted as a two dimensional landscape showing the angular population density. The figure include the position of the global minimum (+) calculated in MM3 (figure 3).

Figure 5. Adiabatic map of the primary dihedral rotations (Φ and Ψ) of the α -(1 \rightarrow 6)-linkage in α -maltose as a function of the staggered

conformations of the third primary dihedral ω . The crystal structure of α -panose is indicated with a C.

Figure 6. Optimized geometry of the amylopectin double-helical fragments constructed by the POLYSaccharide builder (POLYS) and compared to the ensemble average conformation of the branched pentasaccharide.

Figure 7. Two-site radial pair distribution function of the water structure relative to O-2(B) and O-5(D). The contour map reveals an anisotropic water density of 8.5. This water bridge between O-2(B) and O-5(D) is present about 73% of the time, having maximum and average residence times of 18.36 and 0.57 ps, respectively. The average distance O-2(B)···O-5(D) was 4.49 Å, ranging from 5.49 to 2.62 Å. The inserted figure of the pentasaccharide shows a snapshot of the relevant water bridge.

Figure 8. A schematic drawing of the layered structure of amylopectin. The parallel helices pack to form hexagonal or pseudo hexagonal packing. Alternating amorphous and crystalline lamellae are repeated with 9 nm spacing. State-of-the art modeling concern tiny double helical structure and nano-crystallite packing.

Figure 9. Models of phosphorylation sites in the double helical part of amylopectin. To the left C-3 phosphorylation sticks out from the helix coil and C-6 phosphorylation takes place in a surface groove. The latter is illustrated with a spacefilling model to the right.

Figure 10. The hydration of the an entire double helical fragment of amylopectin. With such models we can hope to begin to model excluded volume and starch swelling.

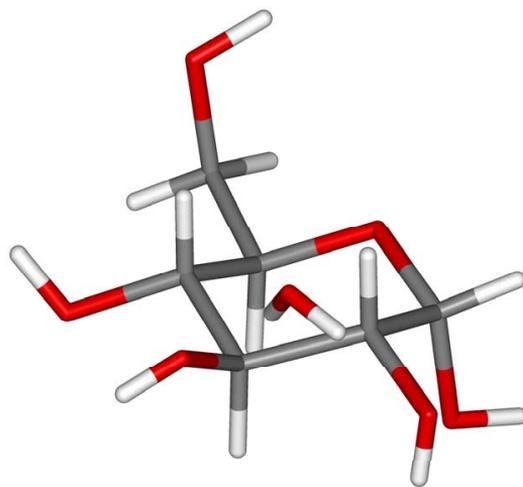


Figure 1. Hansen & Engelsen

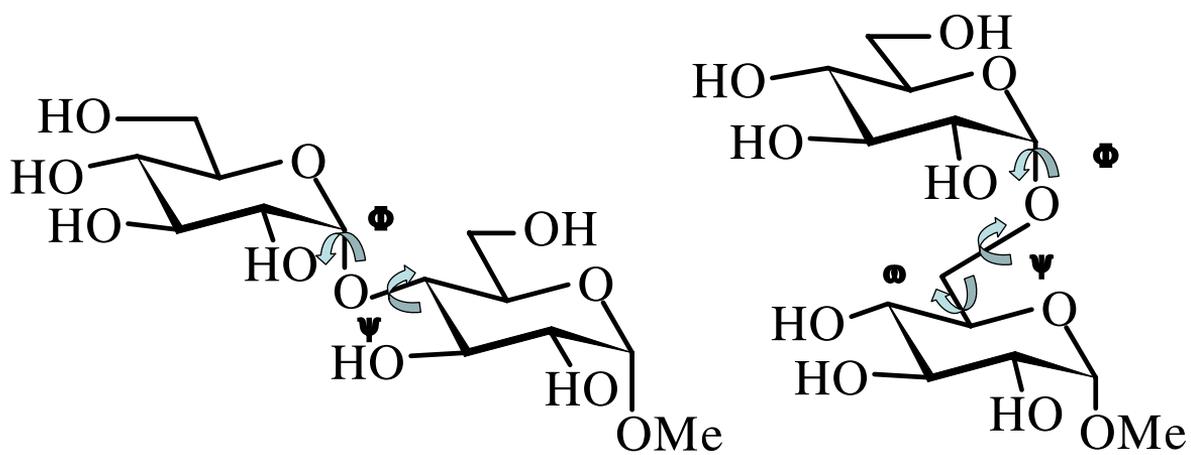


Figure 2. Hansen & Engelsen

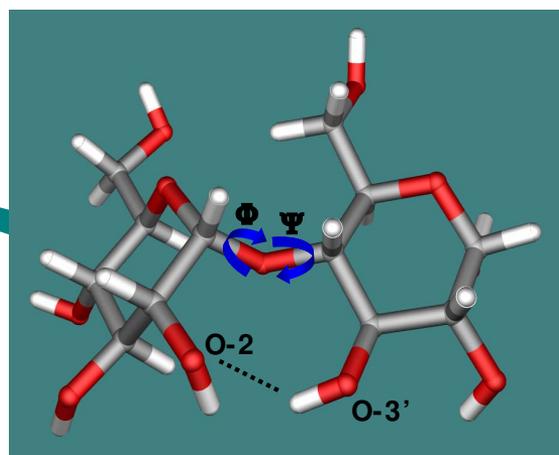
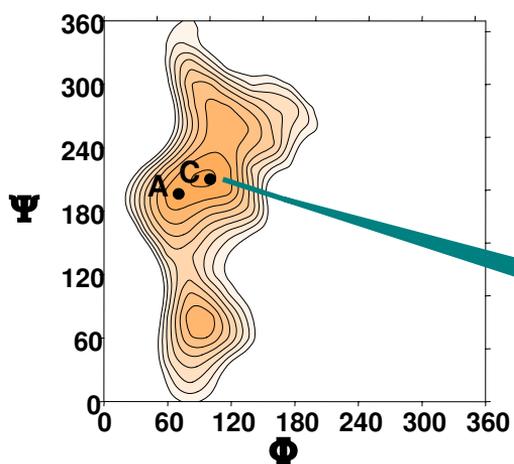


Figure 3. Hansen & Engelsen

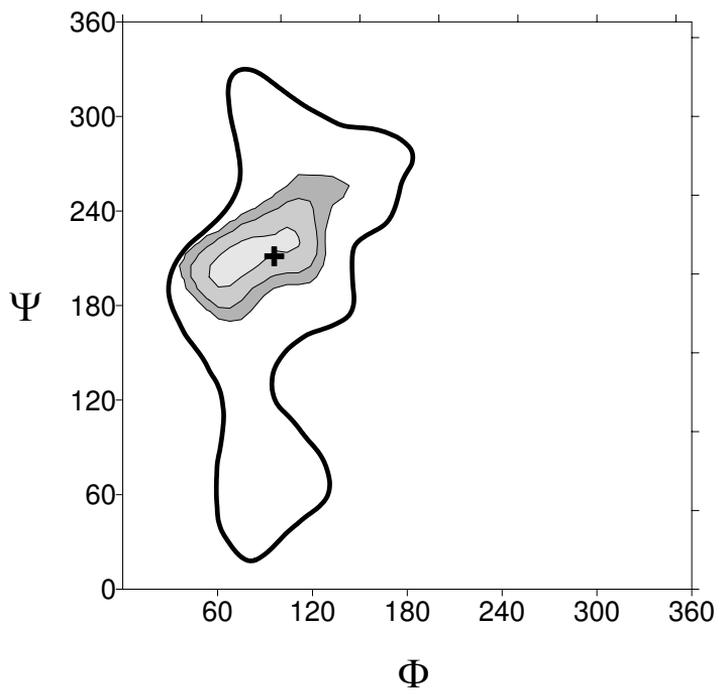


Figure 4. Hansen & Engelsen

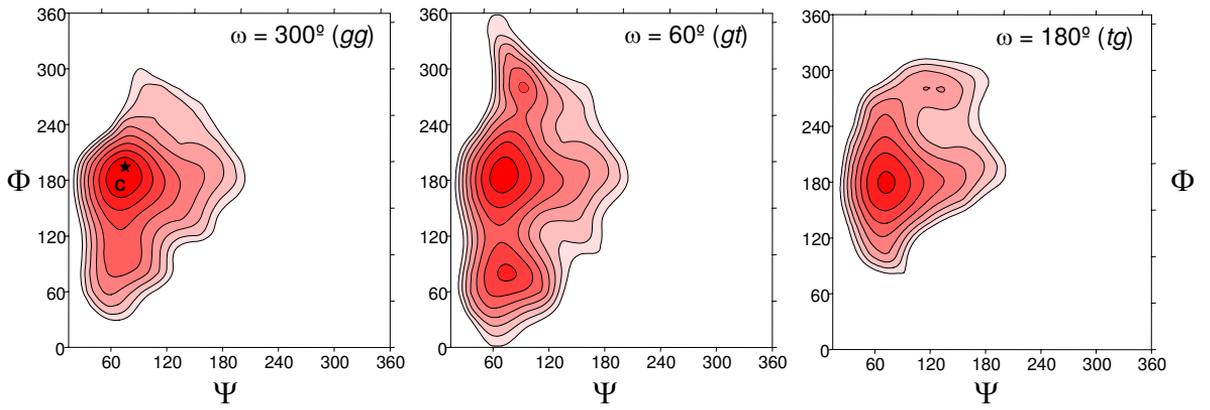


Figure 5. Hansen & Engelsen

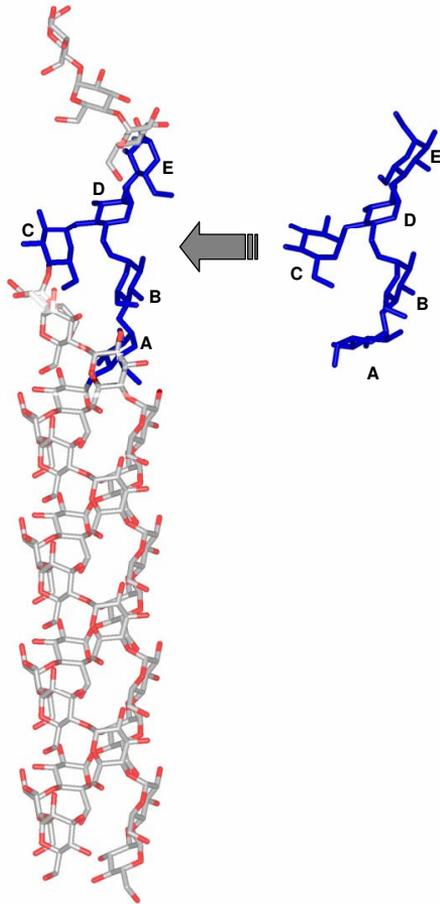


Figure 6. Hansen & Engelsen

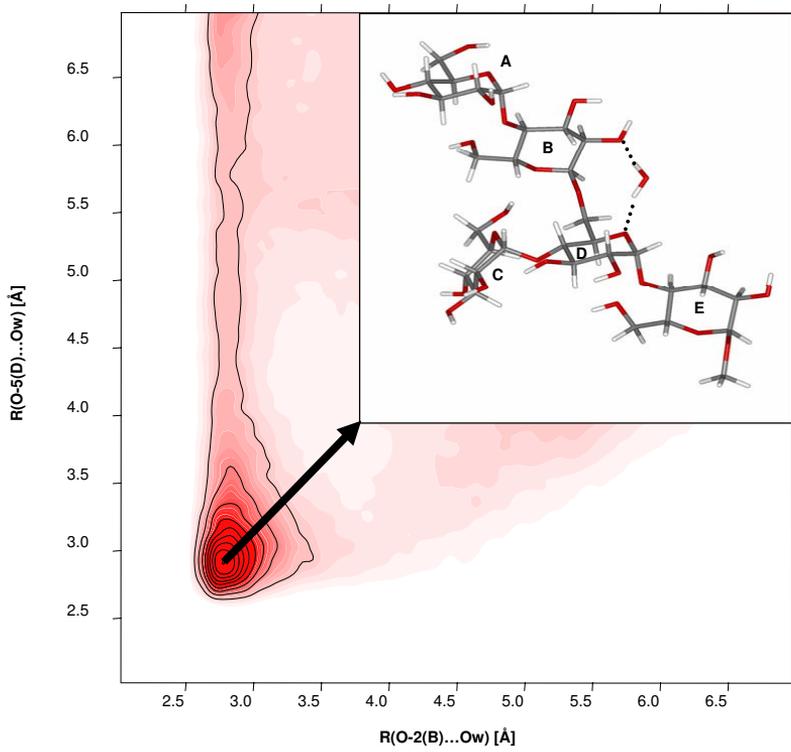


Figure 7. Hansen & Engelsen

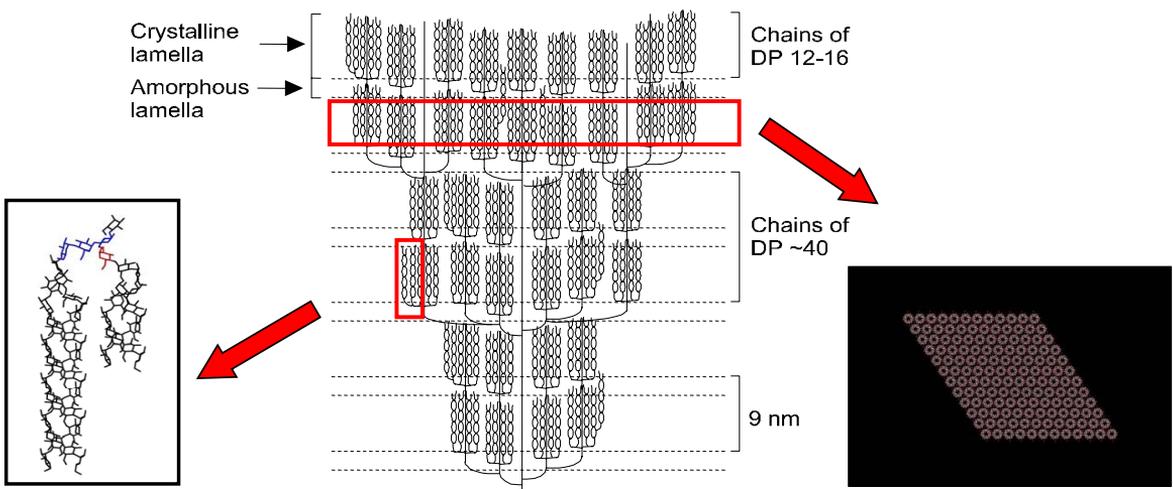


Figure 8. Hansen & Engelsen

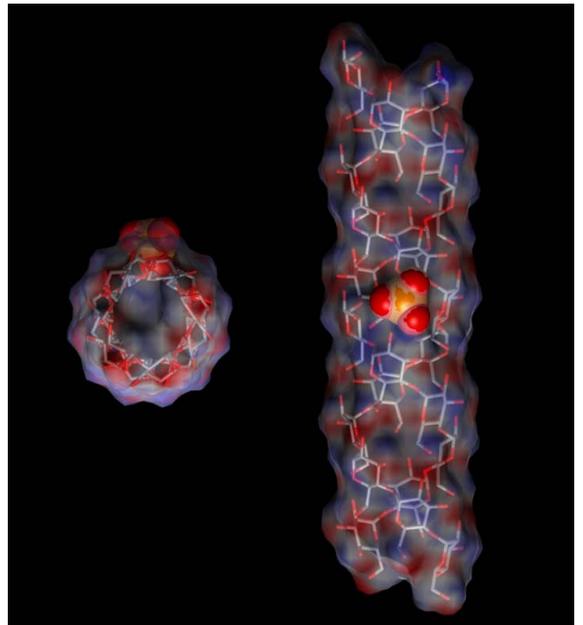
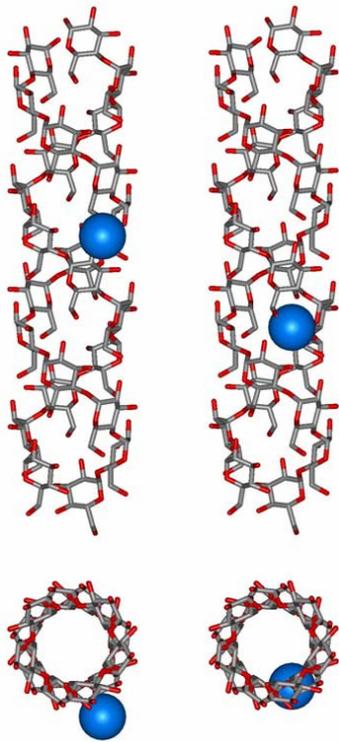


Figure 9. Hansen & Engelsen

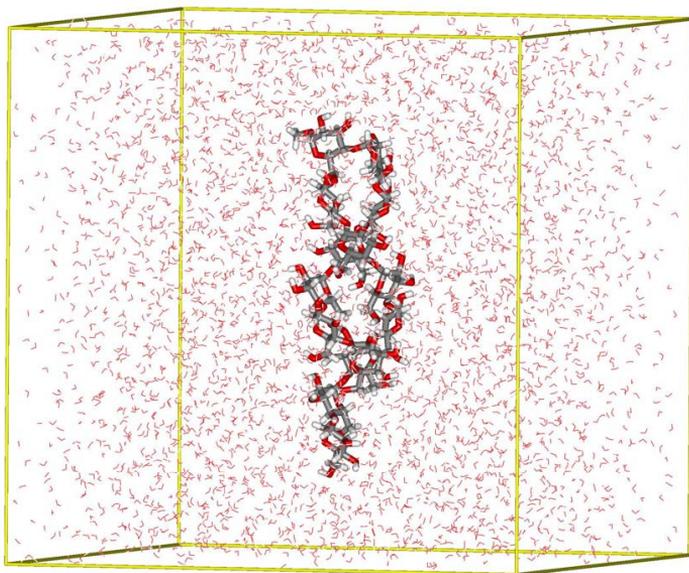


Figure 10. Hansen & Engelsen

Paper II

P. I. Hansen, F. H. Larsen, S. M. Motawia, A. Blennow, M. Spraul, P. Dvortsak and S. B. Engelsen, *Structure and hydration of the amylopectin trisaccharide building blocks – Synthesis, NMR and molecular dynamics*. ChemBioChem, **Submitted**

DOI: 10.1002/cbic.200(will be filled in by the editorial staff)

The amylopectin trisaccharides

Structure and hydration of the amylopectin trisaccharide building blocks

– Synthesis, NMR and molecular dynamics

Peter I. Hansen,^[a] Flemming H. Larsen,^[a] Saddik M. Motawia,^[b] Andreas Blennow,^[b] Manfred Spraul,^[c] Peter Dvortsak^[c] and Søren B. Engelsen*^[a]

To gain insight into the molecular details and hydration of amylopectin, the four constituting trisaccharides have been chemically synthesized as their methyl α -glycosides. The trisaccharides were subjected to 950 MHz NMR spectroscopy for complete assignment and nanosecond molecular dynamics trajectories were calculated to study the structure and dynamics of the trisaccharides in aqueous solution. Systematic analysis of the simulation data revealed several examples of bridging water molecules playing an important role in the stabilization of specific amylopectin conformations, which was also supported by the experimental NMR data such as interresidue NOE's and heteronuclear scalar couplings between nuclei from neighboring residues.

While α -maltotriose, α -iso-maltotriose, α -panose and α -isopanose are relatively well characterized structures, the study also includes one

less characterized trisaccharide with the structure α Glcp(1 \rightarrow 4) α Glcp(1 \rightarrow 6) α Glcp. This trisaccharide, tentatively labelled α -forkose, is located at the branch point of amylopectin, forking the amylopectin into two strands that align into double-helical segments.

The results show that the conformation of α -forkose takes a natural bend form which fits well into the structure of the double-helical segment of amylopectin. As the only trisaccharide in this study the structure of α -forkose is not significantly influenced by the hydration. In contrast, α -isopanose takes a restricted, but rather extended form due to an exceptionally strong localized water density. The two homo-linkage oligomers, α -maltotriose and α -iso-maltotriose, showed to be the most extended and the most flexible trimers, respectively, providing regular structure for crystalline domains and maximum linker flexibility for amorphous domains.

Introduction

Besides being the most important energy source of food and feed, starch, and thus amylopectin, is used extensively as an ingredient to manipulate the quality of our food and as an inexpensive, versatile, renewable and biodegradable polymer in a wide range of material applications such as a thickener and stabilizer for controlling consistency and as a texturizer. Many observations suggest that carbohydrate structure and dynamics are significantly influenced by localized interactions with water^[1-3] and irrespective of food, food ingredient or material application, practically all aspects of starch functionality are related to the hydration of the starch structure. As an example, significant shifts are observed in the glass transition temperature for amylopectin in the presence of water^[4]. In order to fully exploit the potential of starch material and ingredient applications, an improved understanding of starch-water interactions at the molecular level is required. Only when we understand more about the complex relationships occurring in the hydration of starch, new starch materials with well-defined functionalities can be designed, making it possible to manufacture foods as well as degradable and renewable biomaterials with new and improved functionalities^[5]. Knowledge concerning the molecular hydration

of starch may also have important nutritional implications, as the resistance of some starches towards dietary digestion, and thus the generation of dietary fiber, is directly related to their hydration properties. Further insight into this important nutritional aspect constitutes an attractive research area towards producing healthy starch-based foods to combat the rapid increase in lifestyle-related diseases such as diabetes and obesity.

The detailed structure and hydration of starch remain elusive, despite the fact that it is an apparently simple bi-molecular

[a] *Ph.d. student, P.I. Hansen, Dr. F.H. Larsen, Prof. S.B. Engelsen
Department of Food Science
University of Copenhagen
Rolighedsvej 30, DK-1958, Frederiksberg C, Denmark
Fax: (+) 45 35 33 32 45
E-mail: se@life.ku.dk*

[b] *Dr. S.M. Motawia, Dr. A. Blennow
Department of Plant Biology
University of Copenhagen
Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark*

[c] *Director M. Spraul, senior scientist P. Dvortsak
Bruker Biospin GMBH
76287 Rheinstetten, Germany*

system consisting of amylose and amylopectin, both built by only one monomeric unit, namely glucose. The amylose molecule is a straight chain $\alpha(1\rightarrow4)$ linked glucose polymer with very few $\alpha(1\rightarrow6)$ bonds and amylopectin has extremely high molecular weight and multiple $\alpha(1\rightarrow6)$ branched glucose dominated by short chain $\alpha(1\rightarrow4)$ linked side chains. Evidence from X-ray crystallography clearly shows that the amylopectin molecule is further structured in layers of alternating crystalline double-helical regions and amorphous linker regions which are arranged with high regularity^[6, 7]. While the crystalline regions contain only structural water, the amorphous regions are dominated by 'mobile' water molecules. One way to gain new insight into the detailed structure and hydration of the amylopectin molecule is by analyzing particularly interesting sub-structural features in smaller fragments of starch. This study represents the bottom-up strategy of synthesizing well-defined starch nano-scale structural motifs and subsequently carrying out a thorough characterization of their hydration and dynamic behavior using molecular dynamics simulations and ultra-high field high-resolution NMR spectroscopy. Thus far, a number of key structural motifs of amylopectin have been synthesized and characterized^[2]. In order to investigate the properties of the two possible glycosidic linkages in amylopectin, $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$, and the importance of the surrounding aqueous solvent, maltose and iso-maltose have previously been thoroughly studied^[8, 10] as the simplest possible models, including the very same disaccharides which have been blocked for mutarotation by α -methoxy derivatization^[11]. This study expands these basic studies to include the amylopectin trisaccharides (see Figure 1) which entail all possible combinations of the three structural elements of amylopectin, i.e. the monomer α -D-glucopyranose and the two glycosidic linkages $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$. One trisaccharide is special, since C1 of the middle glucose unit is blocked as its methyl α -glycoside, while both C4 and C6 of that unit are endpoints of linkages. This is equivalent to the composition found in the amylopectin branch point forking the amylopectin into two strands that align into double-helical segments. We tentatively call this essential amylopectin trisaccharide α -*forkose*.

Being small, α -glucan trisaccharide models have a high ratio of reducing end groups which in aqueous solutions will equilibrate to an equilibrium mixture of the two anomer forms due to mutarotation. The α - and β -anomers of carbohydrates are typically stable solids, but in aqueous solution the equilibration occurs via the ring opening of the cyclic sugar at the anomeric center with the acyclic form as the intermediate. The result is 'ambiguous' molecules that exhibit both alpha and beta anomer chemistry and physics. In order to mimic alpha-glucans as closely as possible and thus to hinder mutarotation we synthesized these amylopectin trisaccharides with blocked anomeric center as their methyl α -glycosides. The α -anomer configuration was chosen to fit with the natural constitution of the glycosidic linkages of amylopectin.

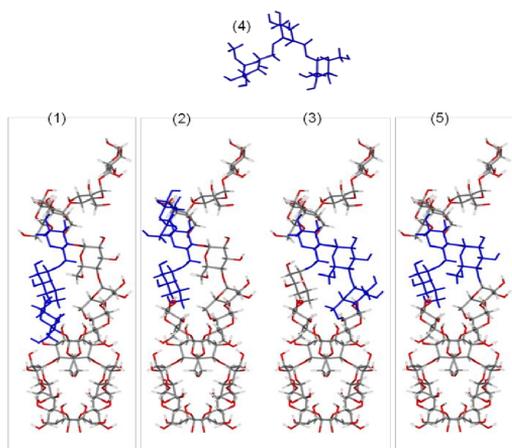


Figure 1. Schematic view of the five amylopectin trisaccharides inserted into the proposed model of the double-helical structure a: methyl α -isopanositide (1), b: methyl α -panositide (2), c: methyl α -maltotriose (3), d: Methyl α -iso-maltotriose (4) and e: methyl α -forkose (5)

Even though the trisaccharides are relatively small molecules with a molecular weight around 600 Da, the similarity of the three glucopyranose rings makes it extremely difficult to gain sufficient NMR spectral resolution at commonly used field strengths. In order to acquire well-resolved homo- and heteronuclear spectra for assignment, ultra-high magnetic field strengths are required, generating adequate spectral resolution for unambiguous and unique assignments. While the branch point amylopectin trisaccharide, α -forkose, previously has been investigated by NMR with focus on chemical shift assignments^[12], this study employs ultra high field NMR (950 MHz for ^1H) for complete assignment of the five trisaccharides, whereas experimental interresidue NOE's and heteronuclear scalar couplings between nuclei from neighboring residues have been acquired at lower fields. While NMR provides only time-averaged data concerning the hydrated structure of the trisaccharides, molecular dynamics simulations in explicit water were undertaken in order to study localized water interactions with the trisaccharides. For each trisaccharide a 15 nanosecond molecular dynamics trajectory was carried out using a second generation carbohydrate force field^[13]. From these simulations intra-molecular distances, heteronuclear coupling constants as well as translational and rotational coefficients were calculated and validated against the NMR measurements. The number of molecular dynamics studies of the hydration of larger α -glucan oligosaccharides is rather limited and includes the study of the linear α -glucans maltohexaose^[14] and maltodecaose^[15] and the branched tetrasaccharide (6'- α -D-glucopyranosyl-maltotriose)^[16] and pentasaccharide (methyl 6'- α -maltosyl- α -maltotriose)^[17]

Subheading

Five trisaccharides are studied in this work. These are methyl 6-O-(α -maltosyl)- α -D-glucopyranoside [methyl α -isopanositide (1)], methyl 6'- α -D-glucopyranosyl- α -maltoside [methyl α -panositide (2)],

methyl α -D-glucopyranosyl-(1 \rightarrow 4)- α -D-glucopyranosyl-(1 \rightarrow 4)- α -D-glucopyranoside [methyl α -maltotrioside (3)], methyl α -D-glucopyranosyl-(1 \rightarrow 6)- α -D-glucopyranosyl-(1 \rightarrow 6)- α -D-glucopyranoside [methyl α -iso-maltotrioside (4)] and methyl 4,6-di-O-(α -D-glucopyranosyl)- α -D-glucopyranoside [methyl-6- α -glucopyranosyl- α -maltoside or methyl α -forkoside (5)], and they will be referred to by their number or trivial name where appropriate. Each trisaccharide contains three glucose units, **a**, **b** and **c**, the order of which is defined by the chemical shift of the anomeric protons (H1). Thus, 'unit a' is the one where H1 has the highest chemical shift and 'unit c' the lowest chemical shift in the anomeric region. The actual order in terms of chemical composition is determined by the heteronuclear correlation across the glycosidic linkage observed in the HMBC-NMR spectra.

The conformation of a α (1 \rightarrow 4) glycosidic linkage is given by two torsional angles ϕ and φ . The dihedral angle ϕ is defined as the angle between the vectors O5-C1 and O1-C4', and the dihedral angle φ is defined as the angle between C1-O1 and C4'-C5'. For α (1 \rightarrow 6) linkages one additional torsional angle, ω , is needed to describe the linkage geometry and it is defined as the angle between O1-C6' and C5'-O5'.

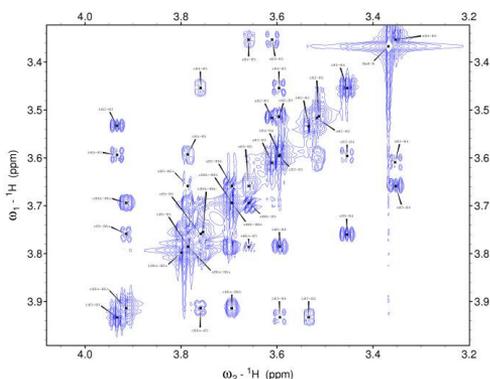


Figure 2. Section of the ^1H - ^1H COSY spectrum recorded at 22.3 T for compound (1), showing the region containing protons in the pyranose rings. Signals overlaps are severe at the diagonal, but resolvable in the cross-peaks due to the high magnetic field strength

Results and Discussion

NMR assignments

All carbon and proton resonances for all of the five trisaccharides could be readily assigned (see Table 1) using our combined strategy, including homonuclear (^1H - ^1H) COSY (Figure 2) and TOCSY and heteronuclear (^1H - ^{13}C) HSQC (Fig. 3) and HMBC NMR experiments conducted at ultra-high field (22.3 T). A subsection of the COSY spectrum (Figure 2) shows how massive overlaps of signals on the diagonal are barely resolved, which illustrates the necessity for ultra-high NMR field strengths for the

analysis of these relatively small saccharides. In addition, the HMBC spectra allowed for determination of the order of the rings linked together in each compound. The glucose monomer order in the trisaccharides was found to be (1): **a**(1 \rightarrow 4)**b**(1 \rightarrow 6)**c**^M, (2): **b**(1 \rightarrow 6)**a**(1 \rightarrow 4)**c**^M, (3): **b**(1 \rightarrow 4)**a**(1 \rightarrow 4)**c**^M, (4): **b**(1 \rightarrow 6)**a**(1 \rightarrow 6)**c**^M, and (5) **a**(1 \rightarrow 4)**c**^M(6 \rightarrow 1)**b**, with the methylation (^M) always at the c monomer.

Table 1. The complete NMR assignment obtained at 22.3 T for all ^1H and ^{13}C chemical shifts in compounds 1, 2, 3, 4 and 5)

	(1) ^[a]		(2) ^[b]		(3) ^[c]		(4) ^[d]		(5) ^[e]	
ring	^1H	^{13}C								
a	5.33	99.70	5.32	99.74	5.33	99.7	4.91	97.78	5.35	99.73
2	3.52	71.77	3.52	71.74	3.55	71.54	3.52	71.32	3.58	71.86
3	3.61	72.87	3.62	73.04	3.89	73.35	3.66	73.31	3.70	72.90
4	3.35	69.33	3.43	69.42	3.59	76.9	3.46	69.47	4.42	69.36
5	3.66	72.68	3.85	71.37	3.78	71.21	3.85	70.28	3.77	72.78
6a	3.79	60.45	3.92	65.90	3.78	60.47	3.91	65.60	3.87	60.49
6b	3.69	60.45	3.67	65.90	3.76	60.47	3.69	65.60	3.78	60.49
b	4.89	97.60	4.89	98.04	4.89	98.08	4.90	97.78	5.00	98.49
2	3.53	71.28	3.50	71.48	3.52	71.72	3.49	71.53	3.55	71.47
3	3.93	73.46	3.68	73.09	3.62	72.88	3.66	73.10	3.73	72.98
4	3.59	77.02	3.37	69.53	3.36	69.32	3.37	69.49	3.45	69.47
5	3.79	70.31	3.66	71.84	3.66	72.68	3.65	71.84	3.76	71.84
6a	3.80	60.45	3.79	60.53	3.79	60.47	3.78	60.51	3.86	60.49
6b	3.76	60.45	3.70	60.53	3.71	60.47	3.70	60.51	3.79	60.49
c	4.75	99.30	4.75	99.06	4.75	99.1	4.76	99.33	4.81	98.99
2	3.51	71.18	3.54	71.01	3.54	71.04	3.51	71.32	3.63	70.91
3	3.60	73.31	3.87	73.52	3.87	73.51	3.60	73.39	3.95	73.25
4	3.45	69.42	3.58	77.25	3.58	76.7	3.45	69.47	3.68	78.02
5	3.76	70.10	3.71	70.12	3.70	70.05	3.77	69.95	3.94	68.96
6a	3.91	65.63	3.84	60.57	3.82	60.47	3.94	65.60	3.99	66.57
6b	3.69	65.63	3.75	60.57	3.75	60.47	3.69	65.60	3.89	66.57
Meth	3.37	55.20	3.36	55.06	3.36	55.07	3.37	55.21	3.44	55.19

[a] methyl α -isopanose; [b]: methyl α -panoside; [c]: methyl α -maltotrioside; [d]: Methyl α -iso-maltotrioside and [e]: methyl α -forkoside

Measurement of $^3\text{J}_{\text{CH}}$ - couplings

Couplings across the glycosidic linkages were assigned accurately by determination of heteronuclear $^3\text{J}_{\text{CH}}$ from the doublet splitting of cross-peaks in J-HMBC spectra at 18.8 T (Figure 3). By averaging the time series of corresponding dihedral angles from the molecular dynamics trajectories, theoretical $^3\text{J}_{\text{CH}}$ coupling constants can be derived through a Karplus type parameterization optimized specially for carbohydrates^[18]. The results from these calculations are listed in Table 2 along with the experimental values.

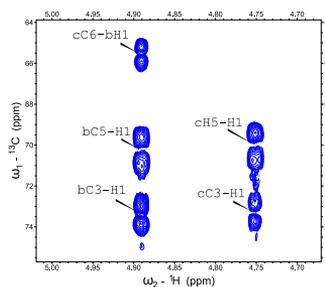


Figure 3. Section of the ^1H - ^{13}C J-HMBC spectrum acquired at 18.8 T for compound (1), showing how signals are split in the F1 dimension due to heteronuclear scalar coupling through 3 covalent bonds. The scaling of the J-coupling in the pulse sequence was set at a factor of $k=40$

In Table 2 it is noticed that the experimental $^3J_{CH}$ couplings from H1 to C4 are always smaller than the corresponding $^3J_{CH}$ coupling from H4 to C1 as seen for compounds (1), (2), (3) and (5). This relation was also obtained by the MD simulations except for (2). As for the $^3J_{CH}$ couplings across $\alpha(1\rightarrow6)$ glycosidic bonds the experimental values are of similar magnitudes ranging from 3.3 to 3.74 Hz independent on whether the couplings is from H1 to C6 or H6 to C1. The $^3J_{CH}$ couplings obtained from MD simulations for H1 to C6 range between 2.58 to 2.74 Hz whereas the experimental values are between 3.49 and 3.74 Hz. This underestimation of around 1 Hz indicates that force fields etc. could be improved to reach the correct level.

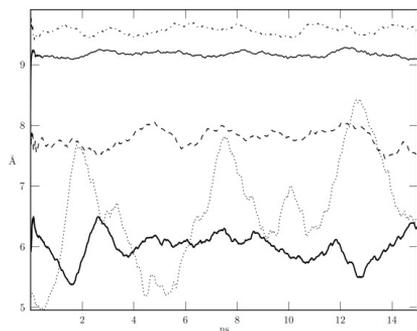


Figure 4. The calculated molecular extension for all five tri-saccharides calculated as the distance between the end ring O5 atom pair in each structure. (—) methyl α -isopanose (1), (---) methyl α -panoside (2), (.....) methyl α -maltotriose (3), (-.-.-) methyl α -iso-maltotriose (4) and (—) methyl α -forkoside (5)

MD analysis

Molecular Extension. As a measure of the overall mobility or stability of the trisaccharide conformations the time series of the O5-O5' in each of the trisaccharides was calculated. The time-evolution of the molecular extensions are displayed in Figure 4. These distances include information about the extension of the molecule, including the rotations of the glucopyranoside rings which can give significant contributions to the maximum extension of more than 1 Å. From the figure it is evident that the five compounds show not only very different extensions, but also very different overall dynamic behavior.

Two trisaccharides are significantly and consistently longer than the three other amylopectin trisaccharides: isopanose (1) and maltotriose (3). Maltotriose (3), the linear maltooligomer omnipresent in amylopectin, and the basic structural trimer of amylose, is the most extended molecule with an average extension of about 9.57 Å. Isopanose (1), the trimer that links the linear maltooligomer chain to the branch point, is only about 0.4 Å shorter on average. α -panose (2), the continuation of the branch point into a new amylosic chain, is considerably shorter than the two more extended trisaccharides with an average extension just below 7.8 Å and almost never extends beyond 8.0 Å. The trimer representing the center of the branch point, forkose (5), is on

average the least extended, or most folded, with a more or less stable maximum extension of about 6.0 Å on average. The homolinkage $\alpha(1\rightarrow6)$ trisaccharide, iso-maltotriose (4), is rather short, 6.6 Å on average, but exhibits very large fluctuations. This trisaccharide has not previously been detected in amylopectin but may possibly occur in amorphous linker regions. Iso-maltotriose is unique among the amylopectin trisaccharides by showing fluctuations in the maximum extension of more than 1.5 Å between 5.0 Å and 6.6 Å interchanging between different conformations.

Molecular diffusion. The translational diffusion coefficient, D, was estimated using the Einstein relation, valid at long times, by calculating the mean square displacement autocorrelation function for the solute center of mass as a function of Δt ^[19, 20]. The autocorrelation functions approach asymptotically the value of the translational self-diffusion coefficient. For α -maltose this value is calculated to $3.2 \cdot 10^{-6} \text{ cm}^2 \cdot \text{s}^{-1}$ compared to an experimental value of $4.0 \cdot 10^{-6} \text{ cm}^2 \cdot \text{s}^{-1}$.^[2] The translational diffusion coefficients calculated for the different trisaccharides are all centered around $2.2 \cdot 10^{-6} \text{ cm}^2 \cdot \text{s}^{-1}$ and display very small variations. The calculated diffusion coefficients of the trisaccharides appear to be in the correct range when compared to the disaccharide, albeit on the low side when compared to larger oligosaccharides such as the branch-point pentasaccharide with an experimental value of $2.2 \cdot 10^{-6} \text{ cm}^2 \cdot \text{s}^{-1}$.^[17]

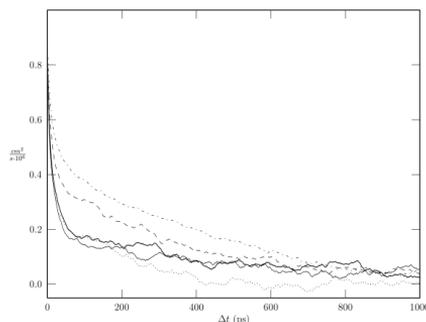


Figure 5. Rotational diffusion for all trisaccharides calculated as angular displacement of the dipolar moment vector, revealing noticeable differences to be quantified by fitting to a sum of exponential decays. (—) methyl α -isopanose (1), (---) methyl α -panoside (2), (.....) methyl α -maltotriose (3), (-.-.-) methyl α -iso-maltotriose (4) and (—) methyl α -forkoside (5)

The rotational diffusion, or the overall molecular tumbling time, for the solute was assessed from the angular evolution of the solute dipole moment, as expressed by the autocorrelation function of the angular displacement of the dipole moment vector as a function of Δt .^[20] The rotational diffusion is roughly estimated by the time when the average angular displacement becomes uncorrelated (orthogonal), but in order to get an exact estimate of the characteristic tumbling time the autocorrelation function was subsequently fitted to an exponential function. In the case of methyl- α -D-maltoside and methyl- α -D-isomaltoside this approach led to characteristic autocorrelation times of approximately 140 ps for both compounds to be compared to the experimental value of approximately 150 ps.^[11] In the case of the trisaccharides this

evaluation led to large differences as can be seen in Figure 5. Iso-maltotriose (4) exhibits the fastest tumbling time (uncorrelated motion) of 170 ps; in the same range as the maltosides. The calculated characteristic tumbling times for the four other trisaccharides range from 380 ps to 570 ps with α -forkose (5) having the longest characteristic tumbling time. This may indicate that α -forkose (5) fits well into the 3D pseudo lattice of water molecules and thus is reluctant to tumble freely – as is the case for the disaccharide α,α -trehalose.^[21]

Glycosidic linkage geometries. Trisaccharide conformational equilibrium preference and flexibility were assessed from torsional angles Φ , Ψ , and ω for each glycosidic linkage as calculated for every time-frame of the five 15 ns trajectories. Possible equilibrium structures were calculated from only statistical (ϕ , φ) population density maps in order to summarize the data.

The time series for the ω dihedral angles revealed that this dihedral was very stable throughout the trajectories and never changes its overall conformation. Except for the case of isopanositide (1) ($\omega=300^\circ$) this dihedral angle adopted a $\omega=60^\circ$ geometry for all $\alpha(1\rightarrow6)$ linkages throughout all simulations (not shown). While the flexibility of the ω torsion in the $\alpha(1\rightarrow6)$ linkages previously have been observed to be limited, the strong preference of $\omega=60^\circ$ conformation of the $\alpha(1\rightarrow6)$ linkages is presumably an artifact of the CSFF force field^[11] as the experimental evidences points to a rotamer distribution with approximately equal amounts of $\omega=60^\circ$ and $\omega=300^\circ$. Nonetheless, this observation reduces the number of descriptive population maps to those containing only combinations of Φ and Ψ .

In order to show how the presence of water influences the conformational equilibrium of the trisaccharides the population distribution of conformations was superimposed onto the MM3^[22] adiabatic energy maps of either methyl α -maltoside or methyl α -iso-maltoside (Figure 6)^[11] calculated for the isolated molecule.

In the case of iso-panositide (1), the (Φ,Ψ) pair describing the $\alpha(1\rightarrow4)$ link is positioned perfectly in the global energy minimum well ($\Phi=100^\circ$, $\Psi=220^\circ$) with the flexibility limited to the Φ in the range from 60° to 130° and Ψ in the range from 180° to 240° . Apparently this linkage is not affected by the hydration. A closer inspection of the map reveals that the distribution for this linkage is split into two major populations: ($\Phi=80^\circ$, $\Psi=214^\circ$) and ($\Phi=106^\circ$, $\Psi=222^\circ$), the latter being the most populated is in good agreement with the exo-anomeric effect and with the global MM3 minimum. For the $\alpha(1\rightarrow6)$ link, the (Φ,Ψ) pair is distributed around a single minimum ($\Phi=70^\circ$, $\Psi=187^\circ$) almost perfectly residing in the global minimum in the energy map, corresponding to the conformation where $\omega=300^\circ$ shows that the hydration does not significantly influence the structure of iso-panositide (1), but also that the molecule is kept rather rigid by the explicit hydration.

In the case of α -panositide (2), the population distribution of its $\alpha(1\rightarrow4)$ linkage is pushed slightly towards higher values of Ψ and only distributed around one minimum at ($\Phi=85^\circ$, $\Psi=216^\circ$), in contrast to the two close minima of iso-panositide (1). This linkage geometry is in good agreement with the X-ray structure of α -panose by Imberty and Perez, in which the $\alpha(1\rightarrow4)$ linkage has the geometry ($\Phi=92.9^\circ$, $\Psi=228.7^\circ$)^[23]. In the case of the $\alpha(1\rightarrow6)$ linkage with $\omega=60^\circ$ three minima along Ψ are clearly defined on the in vacuo adiabatic map, but only the vicinity of one is visited. Apparently, the flexibility along Ψ is increased, as the $\alpha(1\rightarrow6)$

linkage exhibits two almost equally populated conformations, ($\Phi=68^\circ$, $\Psi=170^\circ$) and ($\Phi=72^\circ$, $\Psi=190^\circ$), distributed along Ψ . In the X-ray structure of α -panose the former is chosen for the $\alpha(1\rightarrow6)$ linkage ($\Phi=71.4^\circ$, $\Psi=165.2^\circ$, $\Omega=75.7^\circ$)^[23].

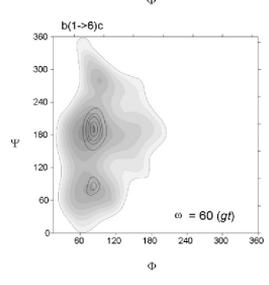
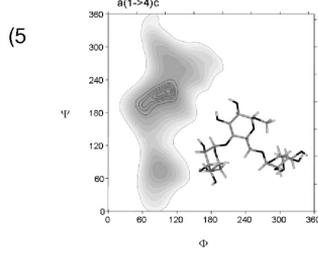
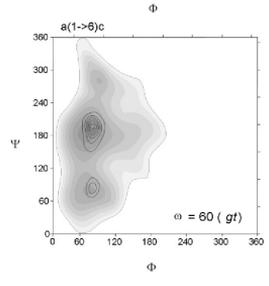
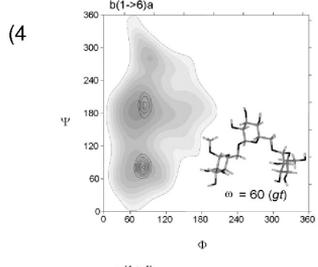
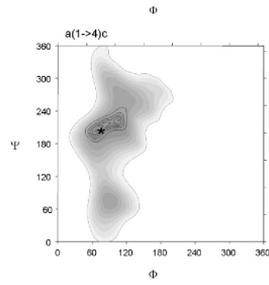
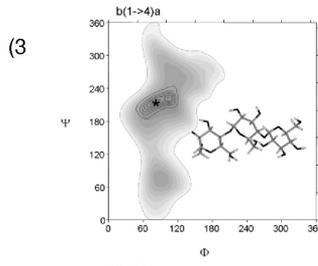
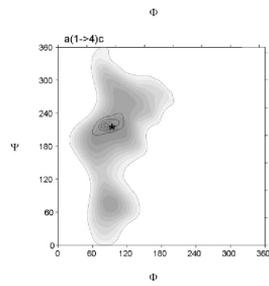
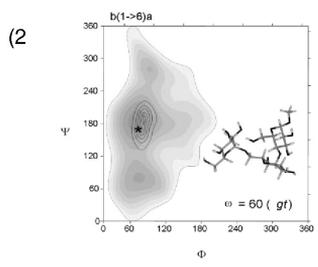
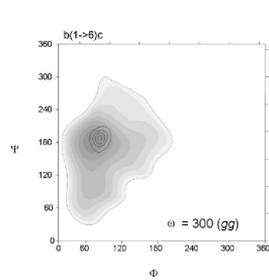
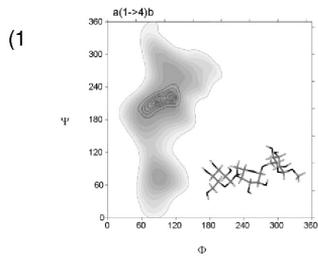
For maltotriose (3), the two $\alpha(1\rightarrow4)$ linkage distribution (Φ , Ψ) pairs are split into two very closely favored conformations. The $b(1\rightarrow4)a$ linkage exhibits the following two favored conformations: ($\Phi=76^\circ$, $\Psi=212^\circ$) and ($\Phi=102^\circ$, $\Psi=223^\circ$), whereas the $a(1\rightarrow4)c^M$ exhibits the following two only marginally changed population maxima: ($\Phi=67^\circ$, $\Psi=205^\circ$) and ($\Phi=106^\circ$, $\Psi=222^\circ$). In the crystal structure of methyl α -maltotriose tetrahydrate (3) by Pangborn and Langs^[24] the geometry of the two $\alpha(1\rightarrow4)$ linkages are found in practically the same conformation ($\Phi=82.2^\circ$, $\Psi=211.1^\circ$) and ($\Phi=82.8^\circ$, $\Psi=208.2^\circ$) for the $a(1\rightarrow4)c^M$ linkage, resembling the first $b(1\rightarrow4)a$ linkage. However, inspection of experimental and theoretical heteronuclear coupling constants (Table 2) shows very good agreement and clearly prove the presence of two different average glycosidic geometries in aqueous solution for the two $\alpha(1\rightarrow4)$ linkages.

Table 2. Heteronuclear coupling constants ($^3J_{CH}$) in Hz measured at 18.8 T and corresponding coupling constants calculated from the molecular dynamics simulation trajectory (MD)

(1) ^[a]	$^3J_{CH}$	MD	(2) ^[b]	$^3J_{CH}$	MD
aH1 -> bC4	4.00	3.8	aH1 -> cC4	3.66	4.04
aC1 -> bH4	4.94	4.14	aC1 -> cH4	4.35	3.06
bH1 -> cC6	3.74	2.65	bH1 -> aC6	3.49	2.64
cH6a -> bC1	3.38	0.99 4.64	aH6a -> bC1	3.40	2.07 2.35
cH6b -> bC1	3.37	0.99 4.64	aH6b -> bC1	3.30	2.07 2.35
(3) ^[c]	$^3J_{CH}$	MD	(4) ^[d]	$^3J_{CH}$	MD
aH1 -> cC4	3.35	3.5	aH1 -> cC6a	3.63	2.58
aC1 -> cH4	4.24	3.97	cH6a -> aC1	3.53	1.10 4.07
bH1 -> aC4	3.76	3.74	cH6b -> aC1	3.31	1.10 4.07
bC1 -> aH4	4.53	4.15	bH1 -> aC6	3.49	2.66
			aH6a -> bC1	3.31	3.92 2.13
			aH6b -> bC1	3.38	3.92 2.13
(5) ^[e]	$^3J_{CH}$	MD			
aH1 -> cC4	3.89	3.92			
aC1 -> cH4	4.73	4.1			
bH1 -> cC6	3.56	2.74			
cH6a -> bC1	3.72	2.94 1.78			
cH6b -> bC1	3.74	2.94 1.78			

[a] methyl α -isopanositide; [b]: methyl α -panositide; [c]: methyl α -maltotriose; [d]: Methyl α -iso-maltotriose and [e]: methyl α -forkoside

Figure 6 – next page. Overlay of the conformational density map on top of the corresponding maltoside and isomaltoside adiabatic energy maps, for the five trisaccharides: methyl α -isopanositide (1), methyl α -panositide (2), methyl α -maltotriose (3), methyl α -iso-maltotriose (4) and methyl α -forkoside (5).



For iso-maltotriose (4), as the only trisaccharide, two different potential energy wells are populated. Both $\alpha(1\rightarrow6)$ linkages display stable $\omega=60^\circ$ conformation, but both (Φ, Ψ) pairs also display two separate population maxima. The $b(1\rightarrow6)a$ linkage exhibits the following two favored conformations: $(\Phi=74^\circ, \Psi=88^\circ)$ and $(\Phi=74^\circ, \Psi=196^\circ)$ and the $a(1\rightarrow6)c^M$ exhibits the following two favored population maxima: $(\Phi=70^\circ, \Psi=86^\circ)$ and $(\Phi=70^\circ, \Psi=197^\circ)$. All four conformations fit well with two of the energetic minima found in the adiabatic energy map, but the two linkages in iso-maltotriose (4) adopt different conformations, since the most populated state of ϕ in the $b(1\rightarrow6)a$ linkage is $\Psi=88^\circ$ and in the $a(1\rightarrow6)c^M$ linkage it is $\Psi=197^\circ$. In particular, the $b(1\rightarrow6)a$ linkage appear to be affected by the hydration. However, the heteronuclear couplings over these two linkages do not exhibit the expected significant differences (Table 2).

In the case of α -forkose (5), both the population maximum of the $\alpha(1\rightarrow4)$ linkage: $(\Phi=104^\circ, \Psi=223^\circ)$ and the population maximum for the $\alpha(1\rightarrow6)$ linkage: $(\Phi=72^\circ, \Psi=191^\circ)$ are found in the global energy minimum wells. However, the latter also rarely visited a secondary energy minimum well with the average $(\Phi=72^\circ, \Psi=88^\circ)$ conformation. The calculated heteronuclear coupling constants for the $\alpha(1\rightarrow4)$ linkage show good agreement with experiments (Table 2) and together these observations confirm that α -forkose (5) is not significantly affected by the hydration. In the case of α -forkose, experimental evidence has indicated that for the beta anomer the ω dihedral exhibit an equal distribution of $\omega=60^\circ$ and $\omega=300^\circ$ rotamers, just like in unsubstituted glucose^[25, 26].

Overall, the glycosidic linkage geometries found in the molecular dynamics simulations only display small, but consistent effects of the explicit hydration. The relatively good agreement between the calculated and the experimental coupling constants supports the hydration models used in the simulations (Table 2).

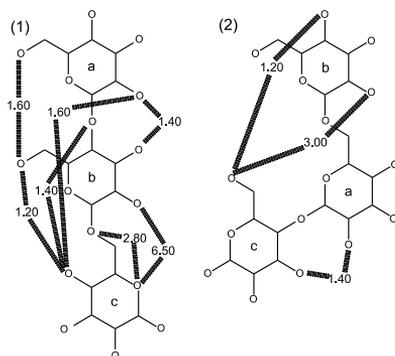


Figure 7. Schematic structures of the five trisaccharides showing occurrences of inter-ring bridging waters represented by the shared water density

Analysis of structural water molecules

Besides the solvents influence on the trisaccharide structure, the detailed influence of an explicit solvent on the carbohydrate structure was examined. Exhaustive calculations were performed to search for extraordinary localized water densities inside the first hydration shell surrounding the solute. For all trisaccharides,

every combination of two oxygen atoms was considered as a site capable of binding a water molecule for long periods of time. For each of these oxygen pairs, the normalized two-dimensional radial pair distribution^[27] was calculated for distances shorter than 3.5 Å from each of the oxygen atoms in the pair. When the radial pair distribution had a peak showing a water density of 1.0 or more this was considered to be a site with special affinity towards binding of water. The bridging water molecules found were either intra-ring, when a pair of hydroxyl groups and/or O5 share a water molecule between them, or inter-ring when the bridging water crosses a glycosidic linkage. Of these two groups, the inter-ring water bridges have the potential to significantly influence the trisaccharide conformation as a semi-structural element. All calculated inter-ring water bridges for the trisaccharides are illustrated schematically in Figure 7.

The trisaccharide with the highest number of water bridges is the highly extended trisaccharide methyl α -isopanose (1) (Figure 8). Both glycosidic linkages seem to be significantly stabilized by bridging water molecules, in particular the $\alpha(1\rightarrow6)$ linkage where bO2 and cO5 exhibit a water bridge with the very high shared density of 6.50 (a snapshot of this situation, which has a maximum residence time of 18 ps, is shown in Figure 8a) and bO1 and cO5 a shared water density of 2.80. It is also observed that cO4 shares resident waters bound to aO1, aO2 and bO6 with densities 1.60, 1.40 and 1.20 respectively. For the $\alpha(1\rightarrow4)$ linkage, the bridges from aO2 to bO3 with density 1.40 and the one from aO6 to bO6 with density 1.60 also give indications of a significant stabilization of this linkage by the explicit hydration, although to a lesser degree. The dense hydration of the $\alpha(1\rightarrow6)$ linkage is in agreement with the dihedral population density map in Figure 6 where only the proximity of the global minimum energy well is visited. The more ambiguous structure of water bridges around the $\alpha(1\rightarrow4)$, where two bridges even stretch across the $\alpha(1\rightarrow6)$ linkage, can be explained by the presence of two stable conformations, which is also expected from the distribution of dihedrals. The stability of both linkages is supported by the NOESY data. In the case of the $\alpha(1\rightarrow4)$ linkage, three NOE's were observed, a strong NOE from aH1 to bH4 and two weak ones from aH1 to bH3 and bH5. In the case of the $\alpha(1\rightarrow6)$ linkage, a strong NOE from bH1 to cH6b and a medium NOE from bH1 to cH6a were observed. A medium intensity NOE was also observed from bH3 to cH4, which is further support to the claim that the dynamics of this link are highly restricted.

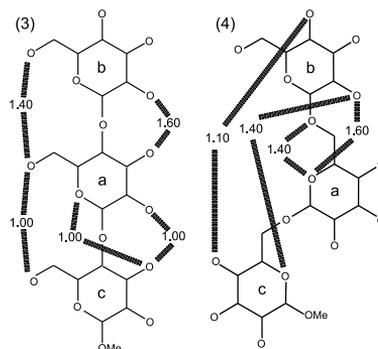


Figure 7. Continued.

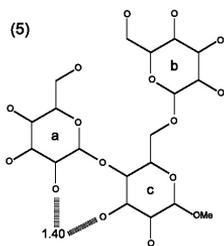


Figure 7. Continued.

The medium folded trisaccharide methyl α -panoside (**2**) (Figure 8), shows an almost reverse behavior to methyl α -isopanoside (**1**). It displays one stable conformation for the $\alpha(1\rightarrow4)$ linkage which must be stabilized by the water bridge from aO2 to cO3 with a shared water density of 1.40. This is supported by the observation of a strong NOE between aH1 and cH4. In the X-ray structure of α -panose aO2 and cO3 is hydrogen-bonded^[23], whereas in aqueous solution an equilibrium between the direct hydrogen bond and the presence of a bridging water molecule exists. The $\alpha(1\rightarrow6)$ linkage, on the other hand, only shares water across the $\alpha(1\rightarrow4)$ linkage through the bridges from bO2 to cO6 and bO4 to cO6 with shared water densities of 3.0 and 1.2, respectively. Figure 8b shows a snapshot of a water molecule bridging bO2 to cO6 which are, in the X-ray structure of α -panose, hydrogen-bonded^[23]. In this particular hydration site, the water is competing with the direct hydrogen bond between cO6 and bO2, and the site is hydrated constantly for 36 ps which is the longest observed in the five trisaccharide trajectories. Medium intensity NOE's were observed between bH1 and the two aH6 protons and from aH1 to bH6b. Calculations of the two aH1-bH6 distance time-series show that this NOE is possible for only one of the bH6 protons and that the peak is expected to be weak in intensity. One very weak NOE was observed between bH1 and cH3, which is noticeable, as this extends across both glycosidic linkages. From distance calculations this NOE is barely possible, since the distance between the two protons is only below 5 Å in a few time frames. Hence, the NOE could arise from a conformation which is not well represented by the simulation, or it could be evidence of our detection limit being better than 5 Å for this sample. Nevertheless, the presence of this NOE supports the more folded nature of methyl α -panoside (**2**) when compared to methyl α -isopanoside (**1**).

In the case of the most extended trisaccharide methyl α -maltotrioside (**3**), all stabilization by bridging waters occurs solely across a single glycosidic linkage. For the first $\alpha(1\rightarrow4)$ linkage two bridges were found, one from bO2 to aO3 with density 1.6 and one from bO6 to aO6 with a shared water density of 1.4. For the second $\alpha(1\rightarrow4)$ linkage three bridges were found, aO2 to cO3, aO5 to cO3 and aO6 to cO6, all with a density of 1.00. In addition to this, a medium intensity NOE was observed from aH1 to cH3. While these observations support the presence of a certain linkage anisotropy for methyl α -maltotrioside in aqueous solution, the shared water density network suggests very small tendency for the structure to fold in good agreement with the molecular extension results shown in Figure 4. No confirmation of the water bridges found in the simulation can be found by inspection of the crystal structure^[24] as it contain only intermolecular water associations.

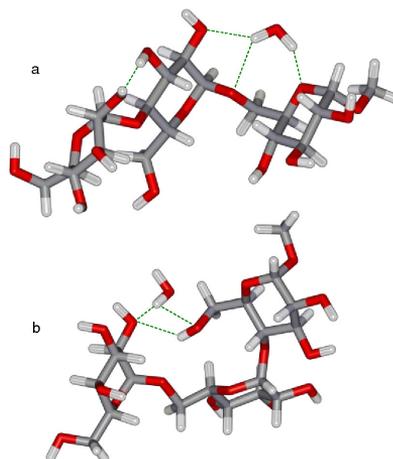


Figure 8. Snapshot showing shared bridging water molecules for a: methyl α -isopanoside (**1**), b: methyl α -panoside (**2**)

In contrast, the much more variable molecular extension of methyl α -iso-maltotrioside (**4**) (Figure 8) indicates a much less stable structure. Nevertheless, four significant shared water densities were found. Two shared water densities are located across the first $\alpha(1\rightarrow6)$ linkage: one from bO2 to aO5 with a shared water density of 1.60 and one from bO1b to aO5a with a shared water density 1.40. The other two shared water densities go across both of the glycosidic linkages from bO4 to cO4 and from bO2 to cO5 with a shared water density of 1.1 and 1.4, respectively. From inspection of the dihedral population density map, the hydration pattern seems to be able to constrain the structure into one main conformation with $\Psi_1=88^\circ$ or $\Psi_2=197^\circ$. However, a minor variation of Ψ is observed around these maxima. Since only intra-ring NOE's were observed for this system, we suggest that two other less stable conformations of the molecule also exist and the equilibrium between these and the main stable conformation is governed by fast dynamics.

Only one plausible water bridge was found for α -forkose (**5**) (Fig. 8). This shared water density crosses the $\alpha(1\rightarrow4)$ linkage from aO2 to cO3 and has a density of 1.40. This may indicate that the $\alpha(1\rightarrow4)$ linkage, but not the $\alpha(1\rightarrow6)$ linkage, is somewhat stabilized, leaving the structure to be guided mainly by the adiabatic energy minima. The stability of the linkages is confirmed experimentally by three NOE's. Two NOE's were found across the $\alpha(1\rightarrow4)$ linkage, a medium intensity NOE from aH1 to cH3 and a strong intensity NOE from aH1 to cH4. Across the $\alpha(1\rightarrow6)$ linkage a strong NOE was observed between bH1 and cH6b. Furthermore, a weak NOE was observed between bH1 and aH3. This is in agreement with the population map displayed in Figure 6 which shows that two different structures are populated. In the most populated structure with $\psi=192^\circ$ the distance from bH1 to aH3 is 6.8 Å, which is too long to give an observable NOE, but in

the lesser populated structure with $\psi = 88^\circ$, the bH1-aH3 distance is 4.5 Å which is short enough to explain the weak NOE observed.

When comparing these five structures and their hydration, it seems reasonable to postulate that sites with a shared water density of around 3.0 or more are required, if a dihedral is to be locked into one single conformation (as observed by NMR).

Conclusion

The structure of the hydration network found in this study is qualitatively mostly similar to what has previously been reported in a study where the types of glycosidic linkages were modeled by maltose for the $\alpha(1\rightarrow4)$ and iso-maltose for the $\alpha(1\rightarrow6)$ linkage^[11]. However, additional bridging water molecules were also found to stabilize the $\alpha(1\rightarrow4)$ bond by linking consecutive non-bonded O6 atoms and also neighboring O2-O3' pairs in the direction of the glycosidic link. The water that is linking O2 to O5' over the $\alpha(1\rightarrow6)$ bond was found to be exceptionally strongly bound in iso-panose and to be present in one of the two possible cases of iso-maltotriose. This could indicate that the effect of water on the more flexible 3-bond $\alpha(1\rightarrow6)$ linkages seems to be more dependent on influences of adjacent parts of the molecule. The strong water density found in iso-panose that is linking O2 to O5' over the $\alpha(1\rightarrow6)$ bond was also found in a study of a pentasaccharide model compound of the branch point of amylopectin^[17] showing that water has a decisive role in locking the conformation. However, the present work shows that this water density does not necessarily exist for shorter branch chains present in the trisaccharide models. Furthermore, only the $\alpha(1\rightarrow4)$ linkage of the branch point trisaccharide α -forkose is restricted by a water bridge, and the molecule is otherwise left to be governed by intra-molecular forces. Yet the flexibility of the molecule is still very restricted, as seen in the distribution of dihedrals, and we propose that this particular structure fits well into the natural network of the surrounding water. This lack of localized water molecules is in agreement with a study of a tetrasaccharide model of the amylopectin branch point^[16] where the branching structure was found to be much less hydrated and also more flexible in comparison to α -panose, which led to the suggestion that the branch residue induces a disorder of the surrounding solvent. The data from the present study, on the other hand, points toward another possibility, namely that the structure of the branched trisaccharide brings a foundation which facilitate the formation of the double-helical arrangement suggested by O'Sullivan *et al.*^[26], while the stability of the larger branch chain is caused by the ability of iso-panose and α -panose to accept localized waters as an additional structural element.

Experimental Section

Chemical synthesis of methyl α -triosides

The strategy for chemical synthesis of all possible $\alpha(1\rightarrow4)$ -, $\alpha(1\rightarrow6)$ - and a combination of $\alpha(1\rightarrow6)/\alpha(1\rightarrow4)$ -linkages of a methyl α -maltotriose **1-5** is outlined in Charts **1** and **2**. As apparent from Chart **1**, regioselective modifications of commercially available D-glucose (**6**), methyl α -D-glucopyranoside (**9**), maltose (**11**) and maltotriose (**15**) provided a route to the necessary synthons **7**, **8**, **10**,

12, **13**, **14** and **16** and are the crucial steps in the synthesis. We have already reported the synthesis of most of these synthons: e.g. **7**^[29], **8**^[30], **10**^[31], **12**^[32] and **16**^[33]. Chart **2** shows the approach towards obtaining the target molecules **1-5** based on the use of synthons **7**, **8**, **10**, **12**, **13**, **14** and **16** via a number of chemical manipulations. These include O-glycosidations, removal of the phenylthio group from the anomeric center using the N-bromosuccinimide method^[30] to allow methyl glycosidation and a final step to remove the benzyl protecting groups by catalytic hydrogenation. Thus, for the synthesis of the α -methyl glycoside derivative of isopanose **1**, (Chart **2A**) compound **12** was used as glycosyl donor and compound **10** as the glycosyl acceptor. The phenylthio function of **12** was activated with NIS / TrIOH combination to afford the fully protected methyl α -trisaccharide-glycoside to obtain **1** after hydrogenolysis and chromatographic purification. For the synthesis of the α -methyl glycoside derivative of panose **2** (Chart **2B**), compound **7** was activated by removal of its phenylthio function. The product was converted into its trichloroacetimidate derivative to be used as the glycosyl donor in the glycosidation reaction with the glycosyl acceptor **13** using TMSOFT as promoter. The phenylthio function was removed from the resulting protected phenylthio-trisaccharide-glycoside followed by methyl glycosidation, chromatographic purification and hydrogenolysis to obtain **2**. The desired methyl α -maltotriose (**3**) (Chart **2C**) was obtained from compound **16** which was easily converted to its hemiacetal using N-bromosuccinimide (NBS)^[33] followed by methyl glycosidation and subsequent hydrogenolysis to afford **3**. To synthesize the target α -methyl glycoside derivative **3**, (Chart **2D**), compound **8** was used as glycosyl donor and compound **10** as glycosyl acceptor where the phenylthio function of **8** was activated with NIS / TrIOH combination to afford the fully protected methyl α -disaccharide-glycoside where its 6'-OPG was regioselectively deprotected and used as a glycosyl acceptor in a repeated glycosidation reaction with the glycosyl donor **8** to afford the fully protected methyl α -trisaccharide-glycoside to obtain **4** after hydrogenolysis and chromatographic purification. Using the same strategy, the target molecule **5** was synthesized from the glycosyl donor **7** and the glycosyl acceptor **14** (Chart **2E**). Detailed synthetic data will be published elsewhere.

Molecular dynamics

The starting structures for the five trisaccharides were generated using the polysaccharide-builder POLYS^[34]. Molecular dynamics was simulated using CHARMM^[35] version 28b2 with the CSFF^[36] carbohydrate force-field. A time-step of 1 fs was used in which the two-step Verlet algorithm^[37] was used to integrate Newton's equations of motion for each atom in the system. The simulations explicitly include all hydrogen atoms, though all covalent bonds involving hydrogen were kept constant using the SHAKE constraint algorithm^[38]. Minimum convention boundary conditions were used during the simulations, and a cut-off distance of 12 Å was used to truncate interactions between atoms further apart and switching functions were used to smoothly turn off interactions between 10 and 11 Å.

Each solute was dissolved in a well-equilibrated cubic box of 512 waters modeled by the TIP3P potential energy function^[39] by superimposition of the coordinates, which led to deletion of 31, 30, 30, 31 and 32 waters due to overlap with solutes **1**, **2**, **3**, **4** and **5** respectively. Next, to relax both the solute structure and steric conflicts in the new environment the five systems were each energy-minimized using the steepest descend method in 50 iterations. After the relaxation, the side length of each box was adjusted to reassign the density of the solution to 1.00 (g/cm³). Velocities for all atoms were assigned during 5 ps from a Boltzmann distribution to represent a temperature of 300 K. The systems were then equilibrated for an additional 100 ps to further relax the new solution, with rescaling of the velocities to counter deviations in temperature by more than 3 K.

After this final equilibration, the integration of the equations of motion was continued for 15 ns for all five systems. Complete phase coordinates were saved every 20 fs for subsequent analysis.

NMR spectroscopy

To evaluate the molecular structures generated from molecular dynamics simulation the 3J heteronuclear C-O-C-H coupling constants across the inter-glycosidic linkages were measured and compared to the corresponding constants derived from the atomic coordinates via a Karplus-type relationship parameterized by Tvaroska *et al.*^[18]

Homonuclear (^1H - ^1H) COSY, TOCSY^[40] (80 ms mixing time), heteronuclear (^1H - ^{13}C) HSQC^[41] and HMBC NMR experiments were performed on a Bruker Avance-III 950 spectrometer (22.3 T) operating at Larmor frequencies of 950.13 MHz for ^1H and 238.91 MHz for ^{13}C . In addition, J-HMBC^[42] spectra were acquired on a Bruker Avance-III 800 spectrometer (18.8 T) operating at Larmor frequencies of 800.13 MHz and 201.193 MHz for ^1H and ^{13}C , respectively. Homonuclear NOESY spectra (300 ms mixing time) were acquired on a Bruker Avance 400 spectrometer (9.4 T) operating at a ^1H Larmor frequency of 400.13 MHz. All experiments were performed in broadband inverse-detection probes (except 800 and 950 MHz, where a Triple inverse probe TXI was utilized) using 5 mm (o.d.) tubes and acquired at room temperature (298 K). Chemical shifts were referenced to TSP-d4. Recycle periods of 1.8-2 s were employed for all experiments.

Samples were prepared by mixing 12.5, 11.9, 12.3, 12.3 and 12.5 mg respectively ((1): 48.2 mM, (2): 45.9 mM, (3): 47.4 mM, (4): 47.4 mM, (5): 48.2 mM) of trisaccharide and 600 μl of water with 10 % (vol.) D_2O and 5.8 mM TSP-d4 (the solvent was 99% D_2O).

Spectral assignments were performed using the Sparky program (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

Acknowledgements

The authors gratefully acknowledge the Faculty of Life Sciences, University of Copenhagen, for the first author's Ph.D. scholarship as well as the Danish Dairy Foundation, Bruker Biospin and the Danish Ministry for Food, Agriculture and Agri Business (FFS05-9 Build your Food) for generous support. We thank Gilda Kischinovsky for proofreading the manuscript.

Keywords: amylopectin trisaccharides, molecular dynamics, NMR, starch hydration, forkose

- [1] S. B. Engelsen, S. Pérez. *Carbohydr. Res.* **1996**, 292(1), 21-38.
- [2] S. B. Engelsen, C. Monteiro, C. Hervé du Penhoat, S. Pérez. *Biophysical Chemistry* **2001**, 93(2-3), 103-127.
- [3] K. N. Kirschner, R. J. Woods. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98(19), 10541-10545.
- [4] M. Di Bari, F. Cavatorta, A. Deriu, G. Albanese. *Biophys. J.* **2001**, 81(2), 1190-1194.
- [5] F. Franks. Hydration phenomena: an update and implications for the food processing industry, In: (Eds.: H. Levine and L. Slade), *Water Relationships in Food*, New York, Plenum Press, 1991, pp. 1-19.
- [6] A. Sarko, H.-C. H. Wu. *Starch/Stärke* **1978**, 30, 73-78.
- [7] A. Imberty, H. Chanzy, S. Pérez, A. Buléon, V. H. Tran. *J. Mol. Biol.* **1988**, 201(2), 365-378.
- [8] M. S. Motawia, I. Damager, C. E. Olsen, B. L. Møller, S. B. Engelsen, S. Hansen, L. H. Øgdenal, R. Bauer. *Biomacromolecules* **2005**, 6, 143-151.
- [9] J. W. Brady, R. K. Schmidt. *J. Phys. Chem.* **1993**, 97(4), 958-966.
- [10] R. B. Best, G. E. Jackson, K. J. Naidoo. *J. Phys. Chem. B* **2001**, 105(20), 4742-4751.
- [11] F. Corzana, M. S. Motawia, C. Hervé du Penhoat, S. Pérez, S. M. Tschampel, R. J. Woods, S. B. Engelsen. *J. Comput. Chem.* **2004**, 25(4), 573-586.
- [12] K. Bock, H. Pedersen. *Carbohydr. Chem.* **1984**, 3(4), 581-592.
- [13] S. Pérez, A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jiménez-Barbero, A. Poveda, J. F. Espinoza, B. P. van Eijck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenhuys, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten, K. Rasmussen. *Carbohydr. Res.* **1998**, 314(3-4), 141-155.
- [14] K. J. Naidoo, M. Kuttel. *J. Comput. Chem.* **2001**, 22(4), 445-456.
- [15] F. A. Momany, J. L. Willett. *Biopolymers* **2002**, 63(2), 99-110.
- [16] R. B. Best, G. E. Jackson, K. J. Naidoo. *J. Phys. Chem. B* **2002**, 106(19), 5091-5098.
- [17] F. Corzana, M. S. Motawia, C. Hervé du Penhoat, F. van den Berg, A. Blennow, S. Pérez, S. B. Engelsen. *J. Am. Chem. Soc.* **2004**, 126(40), 13144-13155.
- [18] I. Tvaroska, M. Hricovini, E. Petrakova. *Carbohydr. Res.* **1989**, 189, 359-362.
- [19] M. P. Allen, D. J. Tildesley. *Computer Simulations of Liquids*, Oxford, UK, Clarendon Press, 1987.
- [20] S. B. Engelsen, C. Hervé du Penhoat, S. Pérez. *J. Phys. Chem.* **1995**, 99(36), 13334-13351.
- [21] S. B. Engelsen, S. Pérez. *J. Phys. Chem. B* **2000**, 104(39), 9301-9311.
- [22] N. L. Allinger, M. Rahman, J.-H. Lii. *J. Am. Chem. Soc.* **1990**, 112, 8293-8307.
- [23] A. Imberty, S. Pérez. *Carbohydr. Res.* **1988**, 181, 41-55.
- [24] W. Pangborn, D. Langs, S. Pérez. *Int. J. Biol. Macromol.* **1985**, 7(6), 363-369.
- [25] Y. Nishida, H. Hori, H. Ohru, H. Meguro, D. Reimert, V. Sinnwell, H. Paulsen. *Tetrahedron Letters* **1988**, 29(35), 4461-4464.
- [26] K. Bock, H. Pedersen. *Acta Chem. Scand.* **1988**, 42(4), 190-195.
- [27] C. A. Andersson, S. B. Engelsen. *J. Mol. Graph. Model.* **1999**, 17(2), 101-105.
- [28] A. C. O'Sullivan, S. Pérez. *Biopolymers* **1999**, 50(4), 381-390.
- [29] I. Damager, C. E. Olsen, B. L. Møller, M. S. Motawia. *Carbohydr. Res.* **1999**, 320(1-2), 19-30.
- [30] M. S. Motawia, J. Marcussen, B. L. Møller. *Carbohydr. Chem.* **1995**, 14, 1279-1294.
- [31] S. B. Engelsen, A. O. Madsen, A. Blennow, M. S. Motawia, B. L. Møller, S. Larsen. *FEBS Letters* **2003**, 541(1-3), 137-144.
- [32] M. S. Motawia, C. E. Olsen, B. L. Møller, J. Marcussen. *Carbohydr. Res.* **1994**, 252, 69-84.
- [33] I. Damager, C. E. Olsen, B. L. Møller, M. S. Motawia. *Synthesis* **2002**, 418-426.
- [34] S. B. Engelsen, S. Cros, W. Mackie, S. Pérez. *Biopolymers* **1996**, 39, 417-433.
- [35] B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus. *J. Comput. Chem.* **1983**, 4(2), 187-217.
- [36] M. Kuttel, J. W. Brady, K. J. Naidoo. *J. Comput. Chem.* **2002**, 23(13), 1236-1243.
- [37] L. Verlet. *Phys. Rev.* **1967**, 159(1), 98-103.
- [38] J. P. Ryckaert, G. Ciccoiti, H. J. C. Berendsen. *Journal of Computational Physics* **1977**, 23(3), 327-341.
- [39] W. L. Jørgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. *J. Chem. Phys.* **1983**, 79(2), 926-935.
- [40] A. Bax, D. G. Davis. *J. Magn. Reson.* **1985**, 65(2), 355-360.
- [41] W. Willker, D. Leibfritz, R. Kerssebaum, W. Bermel. *Magn. Reson. Chem.* **1993**, 31(3), 287-292.
- [42] A. Meissner, O. W. Sørensen. *Magn. Reson. Chem.* **2001**, 39(1), 49-52.

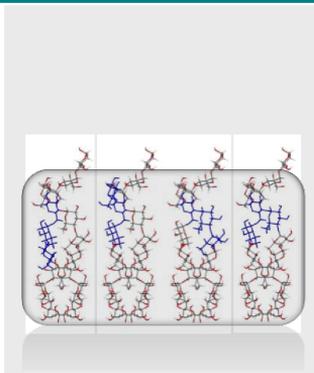
Received: ((will be filled in by the editorial staff))
Published online: ((will be filled in by the editorial staff))

Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPERS

((Text for Table of Contents))



*Peter I. Hansen, Flemming H. Larsen,
Saddik M. Motawia, Andreas Blennow,
Manfred Spraul, Peter Dvortsak and
Søren B. Engelsen*

Page No. – Page No.

**Structure and hydration of the
amylopectin trisaccharide building
blocks – Synthesis, NMR and
molecular dynamics**

SYNTHESIS CHARTS

Chart 1

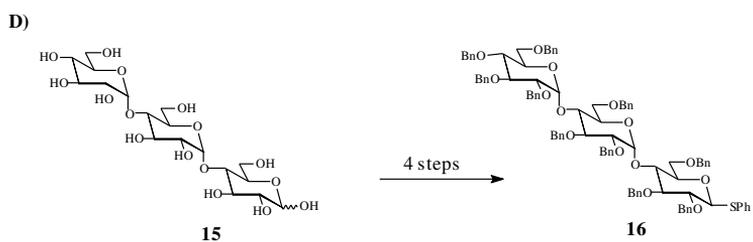
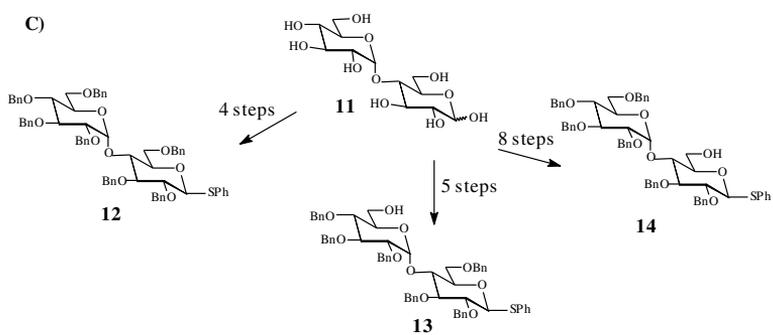
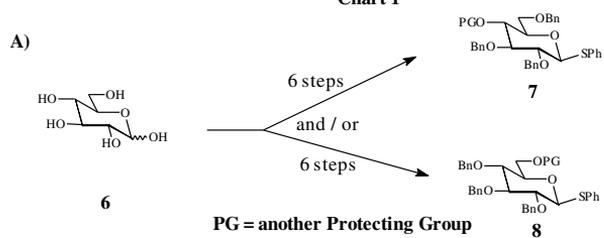
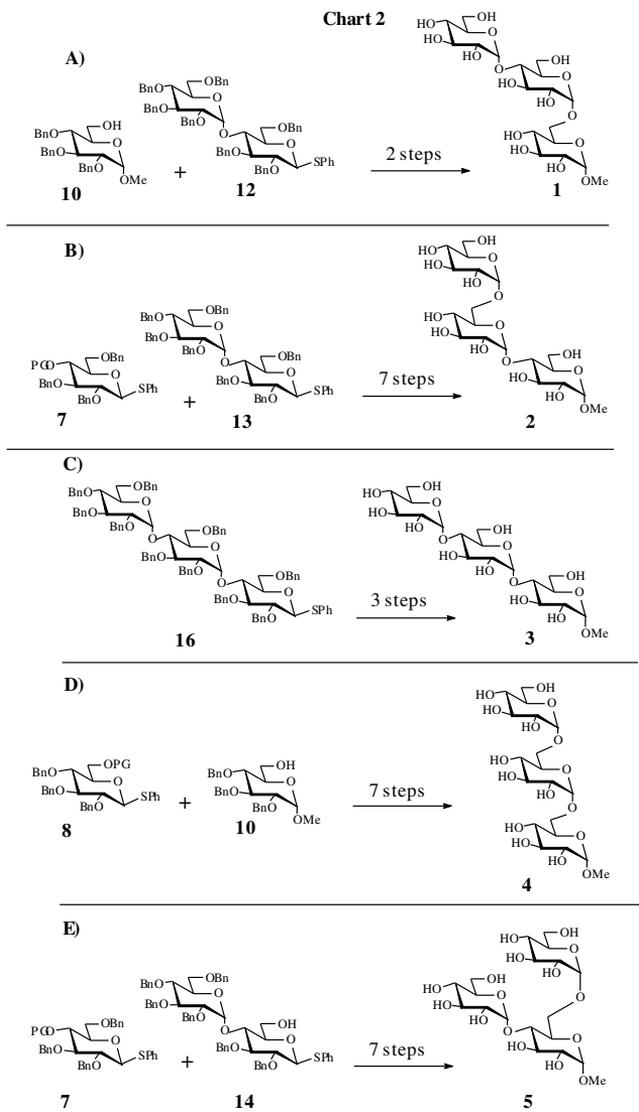


Chart 2



Paper III

P. I. Hansen, J-L. Putaux, A. Blennow and S. B. Engelsen, *Nano-crystal platelets of A- and B-type starches – A preliminary study*. Selected Investigations in the Field of Starch Science, Nova Science Publishers, Inc. 2008. Editors: Vladimir Yuryev, Piotr Tomasik, Andreas Blennow, Lyubov Wasserman, G.E. Zaikov

In press

NANO-CRYSTAL PLATELETS OF A- AND B-TYPE STARCHES – A PRELIMINARY STUDY

Peter I. Hansen[†], Jean-Luc Putaux[‡], Andreas Blennow[‡] and Søren B. Engelsen[†]

[†]*Department of Food Science, Quality and Technology, Faculty of Life Sciences,
University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.*

[‡]*Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS), BP 53,
38041 Grenoble Cedex 9, France - affiliated with Université Joseph Fourier and member
of the Institute de Chimie Moléculaire de Grenoble*

[‡]*Center for Molecular Plant Physiology (PlaCe), Molecular Plant Biology
Department of Plant Biology, Faculty of Life Sciences, University of Copenhagen
Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark*

Abstract

Two-dimensional crystal platelets of A-type waxy maize starch and B-type potato starch were produced by performing mild, long term acidic hydrolysis of the native starch granules. Transmission electron microscopy (TEM) images provided data on the size and shape of the resulting crystalline nanoplatelets. In order to account for some of these features, we have built three-dimensional models of the two types of crystallite.

For nanocrystals corresponding to the monoclinic A-type allomorph, an acute angle of approximately 60° was observed by direct measurement on the TEM images. Using theoretical molecular models taking into account the specific geometry of the crystallites, different organizations of double helices were proposed. This approach suggested that the double helices were not normal to the plane of the lamellar crystals, but has a tilt angle of 26.7°. Using the same approach, we also have explored the hexagonal honeycomb organization of double helices in the B-type allomorph, based on data from TEM images of acid hydrolyzed amylopectin-rich potato starch. In contrast to the A-type crystals the TEM images indicate that the B-type crystals has a geometry corresponding to the structure originally proposed in the starch granule in which the B-type crystal platelets are not tilted.

Introduction

Starch is an extraordinary biopolymer appearing in its native form as partly crystalline granules with properties that strongly depend on those of the crystalline domains. For example, some types of resistant starches are highly crystalline and are used as dietary fiber [1]. The basic molecular structure for the two main constituents of starch, namely amylose and amylopectin, consists of backbone of $\alpha(1\rightarrow4)$ -linked glucose monomers with a varying degree of $\alpha(1\rightarrow6)$ branching linkages. Amylopectin has approximately 5-6% $\alpha(1\rightarrow6)$ bonds resulting in a highly branched molecular structure, while amylose is mostly linear and typically contains less than 1% branches [2].

The branching structure of amylopectin is not random [3] and in the native starch granule, this results in a lamellar structure where the branch points and the linear segments of the polymer structure cluster in alternating amorphous and crystalline lamellae. The crystalline lamellae contain the linear short-chain segments of amylopectin that intertwine to form parallel double helices, while the branch points are located in the amorphous lamellae [4]. Native starch adopts two different crystalline polymorphs as determined by powder X-ray crystallography. The A-type allomorph is found in cereals and has been reported to contain left-handed, parallel-stranded double helices that crystallize into a monoclinic lattice [5]. As shown in Fig. 1A, it is a rather dense structure with room for only very four water molecules between helices. In the B-type allomorph, which is mainly found in tubers, the double helices are crystallized into a hexagonal unit cell [6], most accurately described as a honeycomb-like pattern. As can be seen in Fig. 1B, this is a much more open arrangement with room for 36 water molecules.

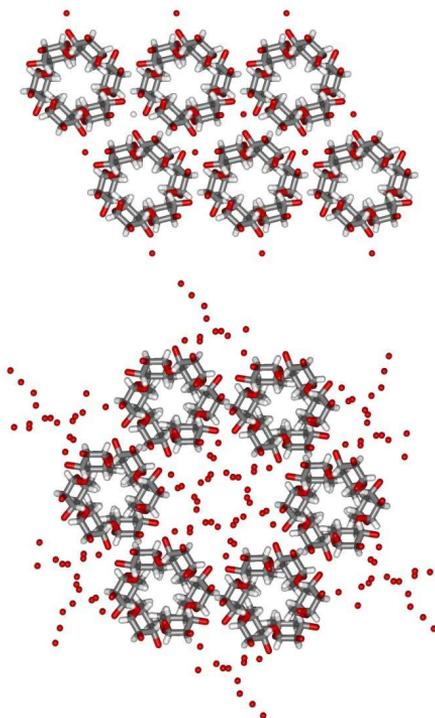


Fig. 1. The crystalline packing of A-type and B-type starch unit cells using the coordinates proposed by Imberty et al.⁵ and Imberty and Perez⁶, respectively. The structures are viewed perpendicular to the helical axis.

The alternating lamellae have been suggested to form specific supramolecular organizations like, for instance, superhelices in the case of potato starch using electro-optical tomography [7-8]. The fact that axis of the constituting double helices would remain parallel to the axis of the superhelices would result in a tilt of the lamellae. Such a tilt was also suggested from electron diffraction data recorded from fragments of acid-hydrolyzed waxy maize starch granules [9].

The current understanding of the molecular structure of starch is only partially complete from the level of the basic α -D-glucopyranose monomer to the amount of

branching linkages and the clustered organization of the polymer into double-helical structures. However, beyond the alternating layers of amorphous and crystalline domains in the amylopectin molecules, an important gap in knowledge includes the effects of different molecular and crystalline nanostructures on the microscopic and granular features. To solve these problems, approaches involving a combination of molecular models and physical observations such as small- and wide-angle X-ray scattering (SAXS and WAXS, respectively) data and electron microscopy imaging were used.

In a recent report [10], it was shown that A-type waxy maize nano-platelets could be isolated after long-term mild hydrochloric acid hydrolysis at moderate temperatures. However, considering the nanometric size of the crystallites visualized by TEM imaging, only hypotheses could be made on the organization of the double helices in the platelets.

In this study, we have built simple molecular models of A- and B-type nanocrystals in order to test different packing of double helices and account for the geometrical data observed in the TEM images. In particular, we examined the possibility of a tilt of the double helices with respect to the plane of the lamellae.

Materials and Methods

Preparation of the nanocrystals

Suspensions of so-called starch "lintners" were prepared according to the method described by Robin *et al.* [11]. Native granules from A-type amylopectin-rich (99%) waxy maize (Waxilys® kindly provided by Roquette, Lestrem, France) and B-type low-amylose cultivar *Dianella* starch were treated with 2.2 N hydrochloric acid at 36°C during 22 days. The suspensions were manually stirred four times a day to limit

sedimentation. Each insoluble residue was washed to neutrality by repeated centrifugation in distilled water. The pellet finally re-suspended in water and kept at 4°C.

Transmission electron microscopy

The lintner suspensions were diluted in water to an approximate concentration of 0.01% (w/v). After a brief sonication, a drop of suspension was deposited onto glow-discharged carbon-coated 200 mesh copper grids. The preparations were then negatively stained with 2% uranyl acetate and allowed to dry. They were observed using a Philips CM200 microscope, operating at 80 kV. Images were recorded on Kodak SO163 films.

Molecular modeling

The various crystal structures investigated were built geometrically by translating the A- and B-type unit cell coordinates as defined by Imberty *et al.*[5] and Imberty and Perez [6], respectively, in order to form a lattice defined to fulfil the symmetry operations implied by the experimentally determined space groups. A software program was made for the purpose to calculate the new coordinates for the series of suggested structures and export these in the MOL2 format as required for visual inspection of the crystal and for future investigation by additional computational techniques such as molecular dynamics or energy optimization studies [12]. In this context the primary output parameter is the angular relationships resulting from the modified crystal geometry.

Results and Discussion

A-type platelets

As already shown in the papers by Putaux *et al.*[10] and Angellier *et al.*[13], the insoluble residue obtained after a few weeks of mild acid hydrolysis of waxy maize starch granules contains platelets that exhibit specific angular features and retained the A allomorphic type of the parent granules. Acute and obtuse angles of approximately 60° and 120° , respectively, could be measured from TEM images of negatively stained preparations. We measured the acute angles of a distribution of similar negatively stained waxy maize platelets obtained after 22 days of acid hydrolysis (Fig. 2A). The average acute angle was approx. 60° which to a first approximation was expected to correspond to a projection of the A-type structure along the helical axis (see figure insert). In that case, the angles between the planes of the monoclinic unit cell are 56.5° and 123.5° (Fig. 1A), which are close but not identical to the values observed in the TEM images.

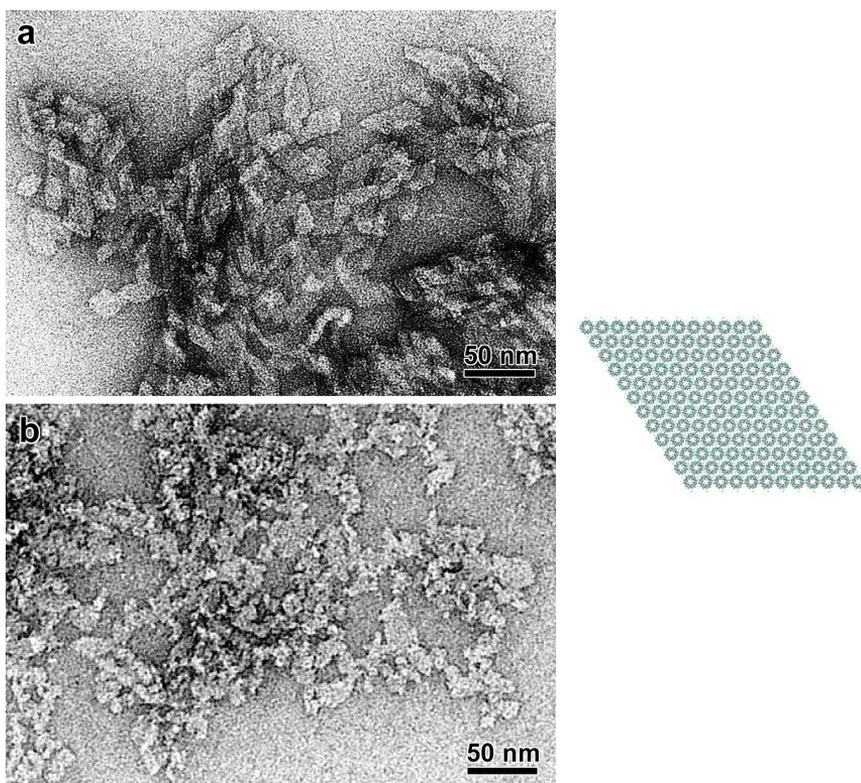


Fig.2. TEM images of negatively stained lintners from HCl-hydrolyzed waxy maize (A) and low amylose potato (B) starch granules.

This observation has led to the idea that the axis of the double helices was not necessarily normal to the plane of the lamellae. Indeed, this idea was supported by earlier results of van Breemen *et al.* [9]. In addition, the observation of feathered-like structures in oblong A-type crystals prepared by *in vitro* crystallization of short amylose chains by Pohn *et al.* lead to similar suggestions [14]. In both reports, though, the images showed populations of lamellae seen edge-on and no single lamella could be observed lying flat in the carbon film. If the double helices are not perpendicular to the lamellae plane, then, when the platelets are lying flat on the carbon film, the double helices are not parallel to

the observation axis. As a consequence the acute angle, Φ , measured on the projected shape should be different than the expected value.

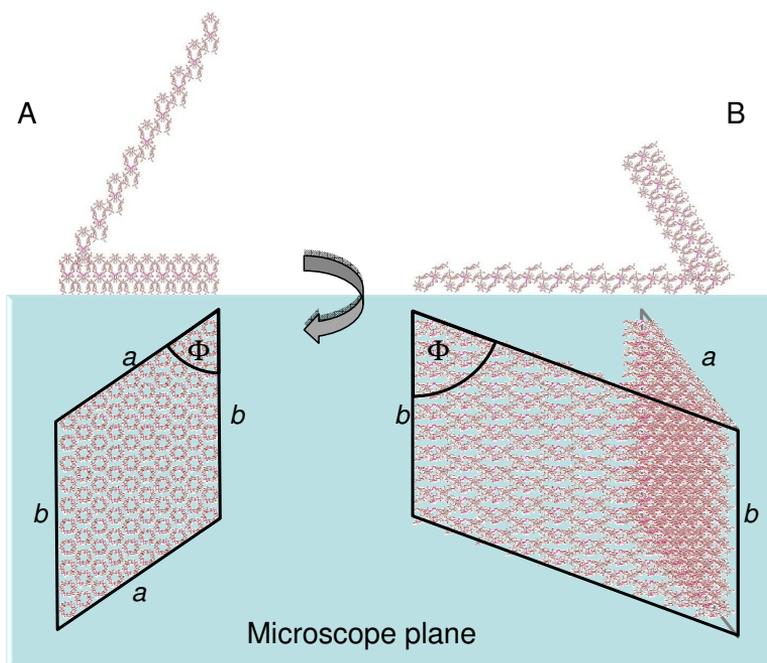


Fig.3. A schematic molecular model to visualize the shift (or tilt angle) influence on the acute angle, Φ , using A-type starch as an example. A (upper): a $(10 \times 10 \times 1)$ A-type platelet orthonormal to the microscope plane superimposed to a $(10 \times 10 \times 1)$ platelet with systematic up-up-up-up shift. A (lower): the same model as A rotated as viewed by the microscope. B (upper): the same model as A but with the shifted platelet lying flat on the microscope surface. B (lower): the same model as B but rotated as viewed by the microscope.

In order to investigate this phenomenon, series nanocrystal platelets containing 10 by 10 double helices were built *in silico* by repeating the unit cells defined by Imberty *et al.* [5] and Imberty and Perez [6] for A-type and B-type allomorphs, respectively, along the a and b and extending only one unit along the c axis ($10 \times 10 \times 1$) (Fig. 3A, bottom).

Secondly, in order to mimic a double helical tilt, rows of double helices (along the a axis) were shifted along the helical axis c resulting in a new sets of crystal platelet coordinates. In Fig.3A a crystal platelet with a systematic helical shift is superimposed to the 'original' non-shifted platelet. When projected to the microscope plane the non-shifted platelet leads to an acute angle, Φ , of 56.5° (Fig.3A, bottom) while the shifted platelet leads to a more open acute angle (Fig. 3B, bottom).

The effect on the acute angle of the platelet was calculated as a function of the shift along the helical axis c , by performing the calculation in steps of 0.1 nm. From the calculations involved in the modelling it was possible to obtain the direct relationship between the inclination of the crystal platelets and the effect of this parameter on the resulting acute angle of the platelet image is observed lying on a flat surface. In Fig. 4 the calculated acute angle, Φ , in the resulting geometry is plotted as a function of the displacement, measured as a distance, between each of the double helical units. From the figure it can be seen that an acute angle of 60° would be the result of a systematic helical displacement of 5.0 \AA . In Fig.5, the acute angle, Φ , is instead calculated as a function of the angle by which the crystal platelet is tilted. In this case the observed acute angle of 60° in A-type platelets corresponds to a helical tilt of the A-type crystals of 26.7° . This result supports the earlier suggestion by Pohu *et al.* [14] that the crystals are built with a tilt of approximately 27° , and indicate that, on average, the acute angle should be close to 60° .

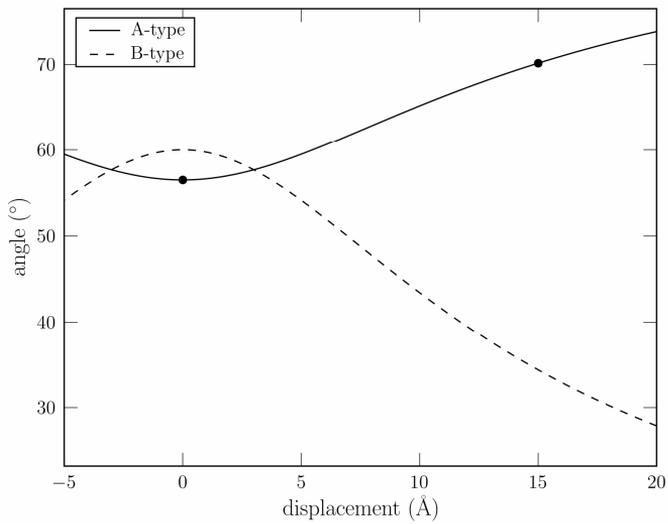


Fig.4. The acute angle, Φ , of A-type and B-type platelets as a function of the systematic displacement (up-up-up-up) of double-helices along the a/b axes.

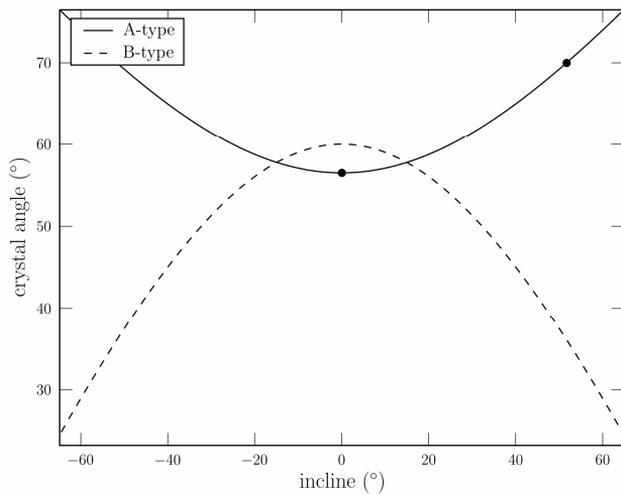


Fig.5. The acute angle, Φ , of A-type and B-type platelets as a function of the tilt angle of double-helices along the a/b axes.

While the tilt angle of approximately 27° is a most interesting result by itself, there remain a number of features to be further investigated. First of all the helical shift is not a continuous feature and in order to provide optimal helical packing the natural shift in the A-type unit cell is $c/2$ [5]. Any intermediate displacement of the helices along c is likely to be energetically unfavourable and, at least, imply a lateral expansion of the unit cell. It means that only discrete angles corresponding to discrete shifts (A-type: $c/2 = 5.35 \text{ \AA}$; B-type: $c/2 = 5.2 \text{ \AA}$) [15] are allowed in the curves shown in Fig. 4. Furthermore other patterns than up-up-up-up in the a or b plane for shifting rows of helices must be considered. The possibility of systematically shifting (up-up-up-up) the A-type platelet in the (a,b) plane can be excluded as it will lead to an even smaller acute angle, however it is possible that other linear combinations of ups ($+c/2$) and downs ($-c/2$) can lead to the same observed acute angle.

B-type platelets

The morphology of the insoluble residue obtained after acid hydrolysis of B-type *Dianella* potato granules (Fig.1B) is not as well defined as that of the A-type platelets prepared from waxy maize granules (Fig.1A). TEM images show a 'lacy' network that occasionally contains objects resembling platelet crystals. By scrutinising the TEM images (Fig.2B) it is possible to find crystal platelets which all seem to possess a 60° acute angle. In contrast to the A-type platelets the 60° acute angle is exactly what to be expected from the inherent features of hexagonal packing without applying any tilt angle. All attempts to modify the structure in a similar manner as with the A-type platelets resulted in a decrease in the acute angle. However, the limited experimental TEM evidence cannot support such a decrease in acute angle albeit this could have confirmed

previous SAXS micro-focus results by Oostergetel and Van Bruggen [8] and Waigh *et al.* [16] who suggested a model with a helical tilt angle of 17.25° . However, since beam-damage is reported among the main problems of this technique, one can speculate whether or not this contributes to the irreversible transformation from B- to A-type morphology.

Conclusion

By building simple molecular models for the determination of the acute angles of A- and B-type starch nanocrystals, it was possible to obtain a model that is in reasonably good agreement with the experimental TEM microscopic data. By shifting every double helical segment along the crystalline axes in the A-type structure 5.35 \AA (corresponding to $c/2$) the resulting crystal structure had a geometry that can explain a 60° acute angle as observed in the TEM images. This shift implies that the helices in the A-type platelets are tilted approximately 27° . For the B-type structure, the original flat layout of the platelets produced 60° acute angles as expected from hexagonal packing of non-shifted flat lamella. We cannot confirm the proposed tilt angle of 17° in B-type starches, but 2D nanocrystals are unquestionably a quite different system than a starch granule.

The quality of these molecular models strongly depends on obtaining a statistically reliable measurement of the angles in the crystal platelets. The understanding of the 2D crystals can be substantially improved by observation of a larger number of starch nanocrystals giving a better accuracy of the average geometry. Unfortunately, the generation of nanocrystals from B-type starches is still very problematic and the procedure need to be further developed to generate appropriate number of structures.

Finally, further improvement would come from an automated unbiased method like computer assisted image analysis of TEM pictures.

References

- [1] Topping, D. L.; Fukushima, M.; Bird, A. R., Resistant starch as a prebiotic and symbiotic: state of the art, *Proceeding of the Nutrition Society* **2003**, *62*, 171-176.
- [2] Buléon, A.; Colonna, P.; Planchot, V.; Ball, S., Starch granules: structure and biosynthesis, *Int.J.Biol.Macromol.* **1998**, *23*, 85-112.
- [3] Thompson, D. B., On the non-random nature of amylopectin branching, *Carbohydr.Polym.* **2000**, *43*, 223-239.
- [4] Jenkins, P. J.; Donald, A. M., The influence of amylose on starch granule structure, *Int.J.Biol.Macromol.* **1995**, *17*, 315-321.
- [5] Imberty, A.; Chanzy, H.; Pérez, S.; Buléon, A.; Tran, V. H., The double-helical nature of the crystalline part of A-starch, *J.Mol.Biol.* **1988**, *201*, 365-378.
- [6] Imberty, A.; Pérez, S., A revisit to the 3-dimensional structure of B-type starch. *Biopolymers* **1988**, *27*, 1205-1221.
- [7] Gallant, D. J.; Bouchet, B.; Baldwin, P. M., Microscopy of starch: Evidence of a new level of granule organization, *Carbohydr.Polym.* **1997**, *32*, 177-191.
- [8] Oostergetel, O. G. T.; Van Bruggen, E. F. J., The crystalline domains in potato starch are arranged in a helical fashion. *Carbohydr.Polym.* **1993**, *21*, 7-12.
- [9] van Breemen, J. F. L.; Keegstra, W.; Van Bruggen, E. F. J.; Oostergetel, G. T. The structure of crystalline amylopectin domains in potato and maize starch granules, *Proc.ICEM13* **1994**, 535-536.

- [10] Putaux, J. L.; Molina-Boisseau, S.; Momaur, T.; Dufresne, A., Platelet nanocrystals resulting from the disruption of waxy maize starch granules by acid hydrolysis, *Biomacromolecules* **2003**, *4*, 1198-1202.
- [11] Robin, J. P.; Mercier, C.; Duprat, F.; Charbonnière, R.; Guilbot, A., Lintnerized starches - chromatographic and enzymatic studies of insoluble residues from acid-hydrolysis of various cereal starches, particularly waxy maize starch, *Starke* **1975**, *27*, 36-45.
- [12] Pérez, S.; Kouwijzer, M. L. C. E.; Mazeau, K.; Engelsen, S. B., Modeling polysaccharides: Present status and challenges, *J.Mol.Graphics* **1996**, *14*, 307-321.
- [13] Angellier, H.; Putaux, J. L.; Molina-Boisseau, S.; Dupeyre, D.; Dufresne, A., Starch nanocrystal fillers in an acrylic polymer matrix, *Macromolecular Symposium* **2005**, *221*, 95-104.
- [14] Pohn, A.; Putaux, J. L.; Planchot, V.; Colonna, P.; Buléon, A., Origin of the limited alpha-amylolysis of debranched maltodextrins crystallized in the A form: A TEM study on model substrates, *Biomacromolecules* **2004**, *5*, 119-125.
- [15] Sarko, A.; Wu, H.-C. H., Crystal-structures of A-type polymorphs, B polymorphs and C-polymorphs of amylose and starch, *Starch/Stärke* **1978**, *30*, 73-78.
- [16] Waigh, T. A.; Donald, A. M.; Heidelberg, F.; Riek, C.; Gidley, M. J., Analysis of the native structure of starch granules with small angle x-ray microfocus scattering, *Biopolymers* **1999**, *49*, 91-105.

Paper IV

P. I. Hansen, M. Spraul, P. Dvortsak, F. H. Larsen, A. Blennow, M. S. Motawia and S. B. Engelsen, *Starch phosphorylation - maltosidic restrains upon 3'- and 6'-phosphorylation investigated by chemical synthesis, molecular dynamics modeling and NMR spectroscopy.*

Journal of the American Chemical Society, **Submitted**

Starch phosphorylation – maltosidic restrains upon 3'- and 6'- phosphorylation investigated by chemical synthesis, molecular dynamics and NMR spectroscopy

Peter I. Hansen[†], Manfred Spraul[‡], Peter Dvorsak[‡], Flemming H. Larsen[†], Andreas Blennow[‡],

Mohammed S. Motawia[‡] and Søren B. Engelsen^{†}*

[†] Quality & Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark. [‡] VKR Center of Excellence Pro-Active Plants, Department of Plant Biology and Biotechnology, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark. [‡] Bruker Biospin GMBH, 76287 Rheinstetten, Germany

AUTHOR EMAIL ADDRESS: se@life.ku.dk

CORRESPONDING AUTHOR FOOTNOTE: Professor Søren Balling Engelsen, Department of Food Science, Quality & Technology, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark. Phone: +45 3533 3205, Fax: +45 3533 3245, e-mail: se@life.ku.dk

Abstract

Phosphorylation is the only known *in vivo* substitution of starch, yet no structural evidence have been provided to explain its implications of the amylosidic backbone and its stimulating effects on starch degradation in plants. In this study we provide evidence for a major influence on the glucosidic bond in starch specifically induced by the 3-O-phosphate. Two phosphorylated maltose model compounds were synthesized and subjected to combined molecular dynamics studies and 950 MHz NMR studies. The two phosphorylated disaccharides represent the two possible phosphorylation sites observed in natural starches, namely maltose phosphorylated at the 3'- and 6'-position (maltose-3'-O-phosphate and maltose-6'-O-phosphate). When compared to maltose, both of the maltose-phosphates exhibit a restricted conformational space of the $\alpha(1\rightarrow4)$ glycosidic linkage. When maltose is phosphorylated in the 3'-position, molecular dynamics and NMR show that the glucosidic space is seriously restricted to one narrow potential energy well which is strongly offset from the global potential energy well of maltose and almost 50° degrees from the Φ angle of the α -maltose crystal structure. The driving force is primarily steric, but also the configuration of the structural waters is significantly altered. Also the 6'-phosphorylated maltose proved to be conformationally restricted across the glucosidic linkage when compared to maltose. Explicitly solvation and deprotonation significantly improved the agreement between theoretical and experimental data. Both the favored conformation of the maltose-3'-phosphate and the maltose-6'-phosphate align well into the 6-fold double helical structure of amylopectin when the effects on the glucosidic bond is not taken into account. However, the restrained geometry of the glucosidic linkage of maltose-3'-phosphate cannot be accommodated in the helical structure, suggesting a major local disturbing effect if present in the starch granule semi-crystalline lattice.

Keywords Starch, amylopectin, maltose, phosphorylation, hydration, NMR, molecular dynamics

Introduction

Starch is the major energy storage component in plant seeds and tubers and is the most important energy resource in the human diet. Starch is deposited in the plant as highly organized, semi crystalline granules¹. The high density and crystallinity prevents hydrolytic enzymes to efficiently attack the native starch granule. There are basically two constituents of starch: amylose which is $\alpha(1\rightarrow4)$ linked glucose linear polymer and amylopectin that is enormous biomacromolecules which are amylosidic chains interlinked by $\alpha(1\rightarrow6)$ branches. Besides from being a glucose polymer, there exists only one natural modification that occur during biosynthesis of starch, and that is phosphorylation². In the plant tuberous starch the phosphate groups have an important function where they provide molecular signals for starch degradation³. The high density and crystallinity prevents hydrolytic enzymes to efficiently attack the native starch granule. It has been suggested² that phosphate restructures the semi crystalline starch granule affecting granule degradability and a recent study provides evidence for a direct effect on hydrolytic enzymes³.

Starch phosphorylation does not only have implications for starch biosynthesis and granule degradability, it also has functional effects in starch applications. Starch is used in foods and materials to control texture, rheology, material strength and barrier performance⁴. Increased starch phosphorylation provides opportunities for generating hydrocolloid systems with new functionality⁵. Regrettably, the mechanisms by which phosphate esters affect starch molecular structure is not known, for which reason detailed information at the molecular level on the effects of phosphate esters on the α -glucan backbone of the starch is required.

Currently, the knowledge about starch phosphorylation is restricted mainly to the fact that most starches are phosphate mono-esterified to a low degree⁶ and that the phosphorylation exclusively occurs in the C-3 and C-6 position⁷⁻⁹. The occurrence of phosphorylation in the crystalline parts of the starch

granules has been evidenced by differential scanning calorimetry¹⁰ and further characterized by ³¹P NMR studies of Naegeli dextrans¹¹ and by studying the gelatinization of B-type tuber starches by ³¹P NMR¹².

In order to provide initial evidence for a structural role of phosphate in double-helical amylopectin, the structure of a methyl α -D-glucopyranoside 6-O-phosphate potassium salt was investigated¹³. The study entailed a combined X-ray crystallography and molecular modeling approach as a first step towards an understanding of such gluco-phosphate structures at the molecular level. It was concluded by extrapolation of the geometry of the phosphate group on a single glucose unit that the 6-phosphorylation does not inflict significant steric problems on the double-helical amylopectin structure. The aim of the present work is to examine the conformational effects of phosphorylation upon the torsional angles of the glucosidic bond in maltose. This provides insight into the naturally adapted conformations of the two primary structural units of phosphorylated starch^{7,8}. For this purpose we have synthesized two maltoses which are phosphorylated either at the C-3 or at the C-6 position at the reducing glucose unit of maltose which are nearest to the glucosidic linkage and thus most likely to influence the glucosidic conformation. A combined approach including NMR spectroscopy and molecular dynamics in explicit aqueous solution was employed to assess the influence of the phosphate substituents on the conformational space of the maltosidic linkage. Specific steric effects was disclosed providing evocative molecular evidence for its known stimulating effect for starch degradation in plants and its tremendous effect on starch hydration.

Nomenclature

The chemical structures of the two phosphorylated disaccharides, maltose-3'-*O*-phosphate (**1**) and maltose-6'-*O*-phosphate (**2**), are shown in Figure 1. Atoms that belong to the reducing glucose unit are marked by a prime (') in the following text.

The conformation of the $\alpha(1\rightarrow4)$ glycosidic linkage is characterized by two torsional angles, Φ and Ψ . The dihedral angle Φ is defined as the angle between the vectors O5-C1 and O1-C4', and the dihedral angle Ψ is defined as the angle between C1-O1 and C4'-C5' in accordance with the IUPAC standard notation for specifying the conformation of polysaccharide chains¹⁴.

The conformation of the phosphate group is also defined according to the IUPAC standard for exocyclic torsions. The geometry of the substituent on C3 is given by the exocyclic torsion χ_3 , defined by the angle between the vectors C2'-C3' and O3'-P' as well as χ_7 which is the angle between C3'-O3' and P'-PO1' where PO1 is a phosphate oxygen. When the phosphate is on C6, the torsions are χ_5 : defined by the vectors O5'-C5' and C6'-O6', χ_6 : defined by the vectors C5'-C6' and O6'-P' and χ_7 : defined by the vectors C6'-O6' and P'-PO1'.

Figure 1

Results and discussion

Adiabatic map of maltose. The adiabatic map of α -maltose was previously calculated by Dowd et al.¹⁵ using the MM3 force field¹⁶ which can be considered as a reference for carbohydrates in general and for the conformational energy of $\alpha(1\rightarrow4)$ glucosidic linkage^{17,18}. In this map the global energy minimum is found at ($\Phi=100^\circ$, $\Psi=223^\circ$) and a secondary, higher energy minimum is located at ($\Phi=90^\circ$, $\Psi=75^\circ$). In the anhydrous crystal structure of α -maltose¹⁹ the conformation was found to have the dihedrals ($\Phi=116^\circ$, $\Psi=242^\circ$), and this displacement of the global energy minimum has been explained by a favorable inter-ring hydrogen-bond between O2 and O3'²⁰.

Initial high temperature dynamics. To evaluate the structural effects of the phosphate substitution on maltose a set of conformational maps was generated by simulating the model systems in implicit

aqueous solvation using high temperature molecular dynamics. These data give an impression of the motional boundaries in the molecules corresponding to an adiabatic map, but with the additional inclusion of solvent effects. From the resulting maps, shown in Figure 2, it is observed that the conformational distribution for maltose-3'-phosphate (Fig. 2A) is concentrated around a single energy minimum centered at ($\Phi=70^\circ$, $\Psi=245^\circ$). The maltose-6'-phosphate is seemingly less affected by the substitution (Fig 2B). In this case, the global minimum is located at ($\Phi=88^\circ$, $\Psi=197^\circ$) and there is a second preference for a conformational family around ($\Phi=88^\circ$, $\Psi=70^\circ$). A conformational family at about ($\Phi=320^\circ$, $\Psi=200^\circ$) is also represented, the relevance of which is doubtful due to the high energy of this area in the adiabatic map.

Figure 2

Molecular dynamics simulation (MD). After the initial high temperature dynamics microcanonical molecular dynamics simulations were conducted at ambient temperature including explicit waters. Since molecular dynamics force fields at present are only available for the most simple carbohydrate structures, additional force field parameters were adapted from the CHARMM²¹ nucleic acid force field²² where both 3'- and 5'-phosphate groups are present as substituents for the deoxy-ribose ring of DNA. All three possible protonated states of the two maltose phosphates were modeled using the modified version of the CSFF²³ force field. The simulation time was 10 ns for each calculation which previously has been shown to provide an excellent model for maltose²⁴. The resulting conformational population maps for the three maltose-3'-phosphates and the three maltose-6'-phosphates are shown in Figure 3. For the maltose-3'-phosphate it is observed that the only populated state is a small circular area centered at ($\Phi=70^\circ$, $\Psi=200^\circ$). This is more than a 30° shift in Φ from the global energy minimum on the MM3 adiabatic map ($\Phi=100^\circ$, $\Psi=223^\circ$) and almost a 50° shift in Φ from the crystal structure geometry ($\Phi=116^\circ$, $\Psi=242^\circ$). This is indeed a major conformational change upon phosphate substitution and

indicates a strong steric effect. However, the restrained minimum of the maltose-3'-phosphate also shows a 45° difference in the Ψ angle compared to the high temperature dynamics data. This suggests that the restrained maltose-3'-phosphate conformation is also influenced by specific hydration events not taken into account for by the continuous solvation approach. In the case of maltose-6'-phosphate no major changes of the conformational population are found when compared to the in vacuum adiabatic map or to the high temperature dynamics conformational map.

When comparing the population density maps with the different protonation states (Figure 3) it is observed that the charge of the phosphate group does not have a strong effect on conformational preference in the case of the 3'-phosphorylation. For the maltose-6'-phosphate two sub minima exist. In the fully protonated state the most populated conformation is located around ($\Phi=65^\circ$, $\Psi=200^\circ$), and as the phosphate group is deprotonated the preferred conformation is shifted towards ($\Phi=100^\circ$, $\Psi=210^\circ$).

Figure 3

Conformation of the phosphate group. To quantify the geometry and dynamic behavior of the phosphate group in explicit water, the time-dependent variation of the phosphate dihedrals is summarized in a set of histograms showing the distribution of observed orientations. This is illustrated in Figure 4. For the 3'-phosphorylation defined by two dihedrals: χ_3 and χ_7 it is observed that the distribution of χ_3 is unimodal and centered at $\chi_3=83^\circ$, but is able to vary by $\pm 45^\circ$. The χ_7 dihedral, on the other hand, is distributed in a staggered-like manner at the 3 angles of $\chi_7=116^\circ$, $\chi_7=230^\circ$ and $\chi_7=353^\circ$, all within $\pm 30^\circ$, with a preference for the 230° orientation. This means that the main direction of the phosphate group defined by the O3'-P bond is relatively fixed with respect to the glucose ring and that the phosphate group is relatively freely rotating around this bond.

Figure 4

For the geometry of the 6'-phosphorylation it is observed that both the χ_5 and the χ_6 are unimodally distributed at $\chi_5=242^\circ \pm 50^\circ$, $\chi_6=128^\circ \pm 50^\circ$. As for the 3'-phosphate, the three conformations of χ_7 adopt the three staggered orientations which are $\chi_7=66^\circ \pm 40^\circ$, $\chi_7=186^\circ \pm 40^\circ$ and $\chi_7=304^\circ \pm 40^\circ$. Again we see that the direction of the O6'-P bond is mostly static. The rotation of the phosphate group in this case is more evenly distributed between the preferred values of the dihedral, and the angles between these three main conformations have population levels of 0.05% which is a factor 10 larger than found for the 3-phosphorylation. In the study by Engelsen et al.¹³ of glucose-6-phosphate the preferred geometry were found to be $\chi_5=63^\circ$, $\chi_6=190^\circ$ and $\chi_7=292^\circ$. The geometry observed in the X-ray data of the glucoside, as illustrated in Figure 5A, is somewhat different from the MD result, Figure 5B, where the phosphate group is closer to C6 and C5 on the non-reducing ring, which suggests that the phosphate group is influenced by interactions with other parts of the structure. This observation could also explain the low-field shift of the C5 resonance found from the NMR assignment.

Figure 5

Hydration analysis. The detailed hydration of α -maltose has previously been investigated by several authors^{20,24,25} who found that two inter-ring bridging water molecules stabilize the structure in aqueous solution. When studied in the CSFF force field both were found to have a density of 1.4 times that of the bulk solvent density. Two oxygen pairs are connected by a water bridge. Those pairs are the O6...Ow...O6' and the O2...Ow...O3' pairs, which are the nearest neighbor oxygen pairs that connects across the glycosidic bond. In order to study how the presence of the phosphate groups affect the water structure across the glucosidic linkage a thorough hydration analysis was carried out by calculating the 2-dimensional radial pair distribution²⁶ for every combination of two oxygen atoms in the solutes. The results are shown in Figure 6, where a density larger than 1.0 (bulk density) is identified as an oxygen

pair which is engaged in the sharing of a water molecule. As expected, many shared water sites were identified between intra-ring neighboring oxygen pairs.

Figure 6

Within the phosphate group, oxygen pairs all have a shared water density between 0.5 and 0.7, thus not being able to bind water for long periods of time. The radial pair distributions were calculated for all phosphate oxygens in the deprotonated form to further evaluate the water structure around the group. From the illustration in Figure 7 it is seen that there is both a clear first and second hydration shell, but that the first shell is pushed to a distance of 3.1 Å with a density of 1.3.

Figure 7

For all the protonation states of the phosphate group the highest shared water density is found at the O5...Ow...O6 site with a value in the range 2.0-2.5, and for maltose-3'-phosphate the O5'...Ow...O6' site shows nearly the same shared water density. The latter site does not have a high affinity for water in maltose-6-phosphate presumably due to the strong competition of the phosphate oxygens.

The O2...Ow...O3' water bridge present in non-substituted maltose²⁴ is practically not significant in maltose-3'-phosphate, as it here has a water density which is in the range 0.9-1.1 (increases with the de-protonation). Water bridges between the oxygens of the phosphate group and O2 are found to be a multisite interaction between the three phosphate oxygens and O1, O2 and O3. In the fully deprotonated case the density between a PO' and O2 is 0.8, 0.5 and 1.0, respectively. A snapshot from the simulation data of the most resident water in this position is shown in Figure 8A.

Figure 8

In the case of the maltose-3'-phosphate the phosphate group also has a strong preference for participating in a water bridge to the oxygens O1', O2' and O3' (Figure 8B), which are the neighboring oxygens in the same ring. For maltose-6-phosphate the hydration of this site is absent. The O6...Ow...O6' pair which was found to be a strong site for binding water in maltose does not exceed 1.0 in any of the simulations, and for maltose-6'-phosphate the water density of this site is below 0.2. This solvation interaction has been taken over by the phosphate oxygens which exhibit collective water sharing between the phosphate group and the three oxygens O4, O5 and O6, which adds up to a structure stabilizing contribution. The interaction between these two groups is in agreement with the structure shown in Figure 5B. Again we observe that this effect is increasing with the degree of deprotonation. In the fully deprotonated case the PO' to O4 densities are 0.1, 0.2 and 0.4. From the PO' to O5 densities are 0.2, 0.3 and 0.2. From the PO' to O6 densities are 0.5, 0.5 and 0.2.

NMR. A complete assignment (Table 1) of resonances was obtained for both solutes from standard 2D NMR spectra. The assignments are in agreement with previous results^{24,27}, except in the case of the H1' and C1' for which the proton chemical shift of Corzana et al. is 0.5 ppm higher and the carbon shift is 8 ppm lower due to the methylation of O1'. In the phosphorylation site of the maltoses we observe a downfield shift of both proton and carbon signals compared to the non-substituted maltose. The heteronuclear coupling constants ${}^3J_{C4'-H1}$ and ${}^3J_{C1'-H4'}$ were used as experimental measures of the dihedral angles and are listed in Table 2 along with calculated values from the simulations and previous results from literature. For maltose-3'-phosphate the coupling constants ${}^3J_{C4'-H1}$ and ${}^3J_{C1'-H4'}$ were measured as 2.3 Hz and 3.4 Hz, respectively. This is in excellent agreement with the results from the molecular dynamics in explicit solvent. The best agreement with theory is found in the case of full deprotonation where the couplings are 2.3 Hz and 3.0 Hz. The continuous solvation model strongly disagrees with experiments for ${}^3J_{C1'-H4'}$.

For maltose-6'-phosphate the heteronuclear coupling constants $^3J_{C4'-H1}$ and $^3J_{C1-H4'}$ are measured to be 4.6 Hz and 3.2 Hz, respectively, which is not in as good agreement with the explicit solvent simulation as for the 3'-phosphorylation, but again the simulation provides the better model when the phosphate group is fully deprotonated where the coupling are 3.5 Hz and 4.0 Hz, respectively. In this case the continuous solvation model is in better agreement with experiments, which is in agreement with the observation that solvation effects are not as strong for this structure as for the methyl-3'-phosphate. The experimental result for the Φ dihedral in maltose-6'-phosphate resembles the data recorded for α -maltose and β -maltose^{24,28}, whereas it is about 1 Hz too small for maltose-3'-phosphate. Also, for the Ψ dihedral in both maltose-3'-phosphate and maltose-6'-phosphate the heteronuclear $^3J_{C4'-H1}$ coupling is almost 2 Hz smaller than that of maltose. Differences of similar magnitude were found by Corzana et al.²⁴

The ROESY spectra showed several intra-ring NOEs but only one additional inter-ring NOE connecting H1 and H4'. For maltose-3'-phosphate the inter-ring NOE is of medium intensity when compared to the intra-ring signals. For maltose-6'-phosphate the NOE is strong providing qualitative support for the difference of average conformations observed in the molecular dynamics data indicating a closer positioning between the two protons in maltose-6'-phosphate.

In relation to the hydration patterns we observed an increased amount of structural waters upon decreased protonation. This is an interesting effect as both phosphate groups are expected to be almost fully deprotonated at pH 7. The change in protonation state also agrees with the molecular dynamics data being best in agreement with the NMR data when fully deprotonized.

Figure 9

In order to investigate how the favored phosphate orientations as calculated from molecular dynamics are accommodated in the double-helical structure of the amylopectin structure without taking any effects of the glucosidic bond into consideration they were inserted into the proposed structure of O'Sullivan and Pérez²⁹. The favored 6-phosphate geometry (Figure 9B) match well in the void in the double-helix as previously predicted from energy calculations¹³. The conformation of the 6-phosphate found here is different from the previous one predicted using force field calculations and X-ray study¹³ on glucose-6-phosphate, now leaning more toward the preceding residue, which can be explained by interactions with functional groups across the glucosidic linkage. The 3-phosphate (Figure 9A) protrudes more from the helical structure than the 6-phosphate. However, the effect is less pronounced than previously predicted.

Figure 10

While the phosphate geometry as confirmed by this study, is rather flexible and able to adapt when substituted into a helical structure, the maltosidic $\alpha(1\rightarrow4)$ linkage of the maltose-3'-phosphate is not. In Figure 10 the adiabatic map of MM3 for the maltosidic $\alpha(1\rightarrow4)$ linkage is superimposed with the most populated potential energy well of the glucosidic linkage of the maltose-3'-phosphate (A) and maltose-6'-phosphate (B) predicted from molecular dynamics simulations (and validated by NMR) as well as with the helical parameter contours. In case of the maltose-6'-phosphate (Figure 10B) the n=6-fold contour eclipse the global energy minimum of maltose as well as the most populated region predicted from the molecular dynamics simulation. This provides strong evidence for the fact that the 6-fold helical structure can form naturally and undisturbed despite 6-phosphorylation and the abundant presence of water. This is however not the case for the maltose-3'-phosphate (Figure 10A). From the figure it is evident that the only populated region of the glucosidic linkage cannot accommodate the n=6-fold helical structure inherent to crystalline amylopectin³⁰. This most interesting evidence clearly contradicts the presence of phosphate in the crystalline regions of amylopectin. However, the

demonstrated presence of 3-phosphate in Nageli dextrans¹¹, suggesting 3-substitution in crystalline lamellae of the starch granule, indicates the presence of minor and local defects within the crystallites. A recent combined thermodynamic and X-ray investigations on transgenic starches³¹ actually indicates that phosphate can induce local amorphisation in the granule resulting in decreased free surface energy of the starch crystallite sides decreased enthalpy of melting without changing order of the crystalline parts as judged by unchanged melting temperature of the starch crystallites. Such effects are thus likely to be induced by 3-phosphorylation and can explain the physiological role of phosphate in stimulating catalysis of specific hydrolases³ by decreasing crystallinity and weaken specific sites of the starch granule.

Concluding remarks

The data generated in the present study provide new basic molecular information on conformational changes induced by phosphorylation at C-6' and at C-3' positions of α -maltose. While the 6-phosphorylation does not change the overall structure of the disaccharide, the 3'-phosphate group shifts the conformational equilibrium of the maltogenic linkage to an Φ angle which is around 30° lower than observed for the unsubstituted α -maltose and around 50° lower than the geometry observed in the crystal structure for the unsubstituted α -maltose. The shift in conformation induced by 3'-phosphorylation is mainly attributed to the steric effects of the large substituent.

The study shows a good agreement between experimental NMR heteronuclear couplings and molecular dynamics simulations in an explicit solvent. This supports the validity of the adapted force-field parameters for the phosphorylated maltoses and adds evidence to the many observations that suggest that carbohydrate structure and dynamics are significantly influenced by localized interactions with water³²⁻³⁴.

Structural preferences for the maltose-phosphates provide insight into functional properties of starch polymers and give circumstantial evidence for stimulatory effects of phosphate esters on starch degradation in the plant. The data suggest that phosphorylation at the 3-position induces significant structural changes in the glucan backbone that can disrupt the crystalline segments of the starch granule and thus provide increased access of hydrolytic enzyme activities to the semi-crystalline starch granule matrix.

METHODS

Chemical synthesis of the phosphorylated maltosides

The chemical synthesis of both maltose-3-phosphate [**Mal3P**] (**4**) and maltose-6-phosphate [**Mal6P**] (**5**) in the form of their disodium salt is outlined in Chart 1. Maltose (**1**) was converted to 1,2,2',3',4',6,6'-hepta-*O*-acetyl- β -Maltose, (**2**)³⁵ and phenyl 2,3,2',3',4',6'-hexa-*O*-acetyl-1-thio- β -maltoside (**3**)³⁶, as described in the literature. Maltose derivatives (**2**) and (**3**) were then phosphorylated at 3-*O*- and 6-*O*-positions using the 2-diethylamino-1,2,3-benzodioxaphosphepane-tetrazole system³⁷. Subsequent oxidation, de-protection and purification provided the desired products (**4**) and (**5**). Detailed synthetic data will be published elsewhere.

Molecular dynamics

Conformational sampling. The phi-psi energy surfaces for the glucosidic linkages were generated using the high temperature molecular dynamics approach of Bruccoleri et al.³⁸. Frank et al.³⁹ showed that 10 ns simulations at 1000K provided an effective representation of visited conformations for small carbohydrates. The 10 ns timeframe has previously been shown to be adequate for establishing a conformational equilibrium in saccharide linkages. To take solvent effects implicitly into account, a generalized Born/surface area (GBSA) algorithm^{40,41} was added to the dynamics calculation. These

calculations were performed for all protonation states of the two maltose-phosphates. Complete phase coordinates were saved every 20 fs in order to describe the conformational distribution.

Explicit solvation. The starting structures for the two phosphorylated carbohydrates were generated using the polysaccharide-builder POLYS⁴², for which the two new monomer structures, α -D-glucose-3-phosphate and α -D-glucose-6-phosphate, was built from the basic α -D-glucose ring structure. Molecular dynamics was simulated using the TINKER Molecular Modeling Package version 4.2 implemented with a modified CSFF²³ carbohydrate force-field including parameters for phosphorylated carbohydrates. A time-step of 1 fs was used in the two-step Verlet algorithm⁴³ which was used to integrate Newton's equations of motion for each atom in the systems. The simulations explicitly include all hydrogen atoms, although all covalent bonds involving hydrogen were kept constant using the SHAKE⁴⁴ constraint algorithm. Minimum convention boundary conditions were used during the simulations, and a cut-off distance of 12 Å was used to truncate interactions between atoms further apart and switching functions were used to smoothly turn off interactions between 10 and 11 Å.

Each solute was dissolved in a well-equilibrated cubic box of 500 waters modeled by the TIP3P⁴⁵ potential energy function by superimposition of the coordinates, which led to deletion of 20 and 19 waters due to overlap with the two solutes maltose-3'-phosphate and maltose-6'-phosphate respectively. In the case of the negatively charged solutes, one or two hydrogens were removed and the corresponding number of sodium ions was placed randomly in the solution box. Next, in order to relax both the solute structure and steric conflicts in the new environment the six systems were each energy-minimized using the Tinker optimization routine until reaching a 0.001 kcal/mol/Å root mean square criterion. Subsequently the side length of each box was adjusted to reassign the density of the solution to 1.00 (g/cm³). Velocities for all atoms were assigned during 5 ps from a Boltzmann distribution to represent a temperature of 300 K. The systems were then equilibrated for an additional 100 ps to further relax the new solution, with rescaling of the velocities to counter deviations in temperature by more than 3 K.

After this final equilibration, the integration of the equations of motion was continued for 10 ns for all 6 systems. As for the conformational sampling technique above, complete phase coordinates were saved every 20 fs for subsequent analysis.

Phosphate force field parameters

Force-field parameters for building a phosphate group onto the α -D-glucopyranose ring were adapted from the existing CHARMM²² parameter set using an approach similar to the work of Crouzy et al.⁴⁶ who adapted the phosphate group to simulations of a cyclodextrin derivative.

The *ab initio* distribution of charges in the phosphate group was calculated for 3'-deoxyribose, 5'-deoxyribose, α -D-glucose-3-phosphate and α -D-glucose-6-phosphate to provide a basic validation of the assumption that the electronic structure of the atoms in the glucose ring is not dramatically affected when the 3- and 6-positions of the sugar ring are substituted by a glycosidic linkage. The charge distribution of the phosphate functional group was also compared when moving from the deoxyribose 5-ring structure to the glucose 6-ring. In both cases the assumption was supported by a close similarity in electronic structure in the vicinity of the phosphate group.

The quantum mechanical calculations were performed using the Massively Parallel Quantum Chemistry Program⁴⁷ (MPQC), at the B3LYP, 6-31G* level of theory.

Helical parameters

Polysaccharides form helical structures and the helical arrangements can be described in terms of helical parameters (n, h) where n is the number of repeating units per turn of the helix and h is the translation along the helical axis. The helical parameters presented in Figure 10, was calculated for all combinations of the torsions in the glucosidic linkage of an amylose strand using the POLYS⁴² programme implementing the procedure described by Pérez et al.⁴⁸

NMR spectroscopy

Molecular structures generated from molecular dynamics simulation were evaluated using the $^3J_{\text{C-H}}$ heteronuclear C-O-C-H coupling constants across the inter-glycosidic linkages measured and compared to their corresponding dihedral angles derived from the atomic coordinates via a Karplus-type relationship parameterized by Tvaroska *et al.*⁴⁹

Homonuclear (^1H - ^1H) COSY, ROESY, TOCSY⁵⁰ (80 ms mixing time), heteronuclear (^1H - ^{13}C) HSQC⁵¹ and HMBC NMR experiments were performed on a Bruker Avance-III 950 spectrometer (22.3 T) operating at Larmor frequencies of 950.13 MHz for ^1H and 238.91 MHz for ^{13}C . In addition, J-HMBC⁵² spectra were acquired on a Bruker Avance-III 800 spectrometer (18.8 T) operating at Larmor frequencies of 800.13 MHz and 201.193 MHz for ^1H and ^{13}C , respectively. All experiments were performed with a triple broadband inverse probe TXI using 5 mm (o.d.) tubes and acquired at room temperature (300 K). Chemical shifts were referenced to TSP-d4. Recycle periods of 1.8-2 s were employed for all experiments.

Samples were prepared by mixing 28.1 mg and 26.3 mg respectively of the disaccharides, and 600 μl of water with 10 % (vol.) D_2O and 5.8 mM TSP-d4 (the solvent was 99% D_2O).

The data was processed with NMRPipe⁵³, except for the J-HMBC which was processed in TopSpin and spectral assignments were performed using the Sparky program version 3.113 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Faculty of Life Sciences, University of Copenhagen, for the first author's Ph.D. scholarship as well as the Danish Dairy Foundation, Bruker Biospin and the Danish Ministry for Food, Agriculture and Agri Business (FFS05-9 Build your Food) for generous support. We thank Gilda Kischinovsky for proofreading the manuscript. We would also like to thank Curtis Janssen for the helpful conversations about features of the MPQC program, and thanks go to Dr. Martin Frank for providing the CAT (Conformational Analysis Tool) software.

Reference List

1. Buléon, A.; Colonna, P.; Planchot, V.; Ball, S. Starch granules: structure and biosynthesis. *Int. J. Biol. Macromol.* **1998**, *23* (2), 85-112.
2. Blennow, A.; Engelsen, S. B.; Nielsen T.H.; Baunsgaard, L.; Mikkelsen, R. Starch phosphorylation: a new front line in starch research. *Trends in Plant Science* **2002**, *7* (10), 445-450.
3. Edner, C.; Li, J.; Albrecht, T.; Mahlow, S.; Hejazi, M.; Hussain, H.; Kaplan, F.; Guy, C.; Smith, S. M.; Steup, M.; Ritte, G. Glucan, water dikinase activity stimulates breakdown of starch granules by plastidial beta-amylases. *Plant Physiol.* **2007**, *145* (1), 17-28.
4. Ellis, R. P.; Cochrane, M. P.; Dale, M. F. B.; Duffus, C. M.; Lynn, A.; Morrison, I. M.; Prentice, R. D. M.; Swanston, J. S.; Tiller, S. A. Starch production and industrial use. *J. Sci. Food Agric.* **1998**, *77* (3), 289-311.
5. Blennow, A.; Wischmann, B.; Houborg, K.; Ahmt, T.; Jørgensen, K.; Engelsen, S. B.; Bandsholm, O.; Poulsen, P. Structure function relationships of transgenic starches with engineered phosphate substitution and starch branching. *Int. J. Biol. Macromol.* **2005**, *36* (3), 159-168.
6. Blennow, A.; Engelsen, S. B.; Munck, L.; Møller, B. L. Starch molecular structure and phosphorylation investigated by a combined chromatographic and chemometric approach. *Carbohydr. Polym.* **2000**, *41* (2), 163-174.
7. Hizukuri, S.; Tabata, S.; Nikuni, Z. Studies on starch phosphate. 1. Estimation of glucose-6-phosphate residues in starch and presence of other bound phosphate(s). *Starch/Stärke* **1970**, *22* (10), 338.
8. Tabata, S.; Hizukuri, S. Starch phosphate. Part 2. Isolation of glucose-3-phosphate and maltose phosphate by acid hydrolysis of potato starch. *Starch/Stärke* **1971**, *23*, 267-272.
9. Noda, T.; Takigawa, S.; Matsuura-Endo, C.; Kim, S. J.; Hashimoto, N.; Yamauchi, H.; Hanashiro, I.; Takeda, Y. Physicochemical properties and amylopectin structures of large, small, and extremely small potato starch granules. *Carbohydr. Polym.* **2005**, *60* (2), 245-251.
10. Muhrbeck, P.; Eliasson, A. C. Influence of the naturally-occurring phosphate-esters on the crystallinity of potato starch. *J. Sci. Food Agric.* **1991**, *55* (1), 13-18.
11. Blennow, A.; Bay-Smidt, A. M.; Olsen, C. E.; Møller, B. L. The distribution of covalently bound phosphate in the starch granule in relation to starch crystallinity. *Int. J. Biol. Macromol.* **2000**, *27* (3), 211-218.
12. Larsen, F. H.; Blennow, A.; Engelsen, S. B. Starch granule hydration - A MAS NMR investigation. *Food Biophysics* **2008**, *3* (1), 25-32.
13. Engelsen, S. B.; Madsen, A. O.; Blennow, A.; Motawia, M. S.; Møller, B. L.; Larsen, S. The phosphorylation site in double helical amylopectin as investigated by a combined approach using chemical synthesis, crystallography and molecular modeling. *FEBS Letters* **2003**, *541* (1-3), 137-144.

14. IUPAC-IUB Nomenclature of Carbohydrates. *Carbohydr. Res.* **1997**, *297*, 1-92.
15. Dowd, M. K.; Zeng, J.; French, A. D.; Reilly, P. J. Conformational analysis of the anomeric forms of kojibiose, nigerose and maltose using MM3. *Carbohydr. Res.* **1992**, *230* (2), 223-244.
16. Allinger, N. L.; Rahman, M.; Lii, J.-H. A Molecular Mechanics Force Field (MM3) for Alcohols and Ethers. *J. Am. Chem. Soc.* **1990**, *112*, 8293-8307.
17. Pérez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jiménez-Barbero, J.; Poveda, A.; Espinoza, J. F.; van Eijck, B. P.; Johnson, G.; French, A. D.; Kouwijzer, M. L. C. E.; Grootenhuis, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. A Comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr. Res.* **1998**, *314* (3-4), 141-155.
18. Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. Evaluation of Carbohydrate Molecular Mechanical Force Fields by Quantum Mechanical Calculations. *Carbohydr. Res.* **2004**, *339* (5), 937-948.
19. Takusagawa, F.; Jacobsen, R. A. Crystal and molecular-structure of alpha-maltose. *Acta Crystallographica Section B - Structural Science* **1978**, *34*, 213-218.
20. Brady, J. W.; Schmidt, R. K. The Role of Hydrogen Bonding in Carbohydrates: Molecular Dynamics Simulations of Maltose in Aqueous Solution. *J. Phys. Chem.* **1993**, *97* (4), 958-966.
21. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187-217.
22. MacKerell, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21* (2), 105-120.
23. Kuttel, M.; Brady, J. W.; Naidoo, K. J. Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J. Comput. Chem.* **2002**, *23* (13), 1236-1243.
24. Corzana, F.; Motawia, M. S.; Hervé du Penhoat, C.; Pérez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B. A hydration study of (1-4) and (1-6) linked α -glucans by comparative 10 ns molecular dynamics simulations and 500 MHz NMR. *J. Comput. Chem.* **2004**, *25* (4), 573-586.
25. Ott, K.-H.; Meyer, B. Molecular dynamics simulations of maltose in water. *Carbohydr. Res.* **1996**, *281*, 11-34.
26. Andersson, C. A.; Engelsen, S. B. The mean hydration of carbohydrates as studied by normalized two-dimensional radial pair distributions. *J. Mol. Graph. Model.* **1999**, *17* (2), 101-105.
27. Hoffman, R. E.; Levy, G. C. Modern methods of NMR data-processing and data evaluation. *Progr. NMR Spectroscopy* **1991**, *23*, 211-258.

28. Pérez, S.; Taravel, F. R.; Vergelati, C. Experimental evidences of solvent induced conformational changes in maltose. *New J. Chem.* **1985**, *9* (8-9), 561-564.
29. O'Sullivan, A. C.; Pérez, S. The relationship between internal chain length of amylopectin and crystallinity in starch. *Biopolymers* **1999**, *50* (4), 381-390.
30. Imberty, A.; Chanzy, H.; Pérez, S.; Buléon, A.; Tran, V. H. New three-dimensional structures for A-type starch. *J. Mol. Biol.* **1988**, *201*, 365-378.
31. Kozlov, S. S.; Blennow, A.; Krivandin, A. V.; Yuryev, V. P. Structural and thermodynamic properties of starches extracted from GBSS and GWD suppressed potato lines. *Int. J. Biol. Macromol.* **2007**, *40* (5), 449-460.
32. Engelsen, S. B.; Pérez, S. The Hydration of Sucrose. *Carbohydr. Res.* **1996**, *292* (1), 21-38.
33. Engelsen, S. B.; Monteiro, C.; Hervé du Penhoat, C.; Pérez, S. The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy. *Biophysical Chemistry* **2001**, *93* (2-3), 103-127.
34. Kirschner, K. N.; Woods, R. J. Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (19), 10541-10545.
35. Dick, W. E.; Baker, B. G.; Hodge, J. E. Preparation and characterization of 1,2,6,2',3',4',6'-hepta-O-acetyl-beta-maltose. *Carbohydr. Res.* **1968**, *6* (1), 52-62.
36. Motawia, M. S.; Olsen, C. E.; Møller, B. L.; Marcussen, J. Chemical synthesis and NMR-spectra of a protected branched-tetrasaccharide thioglycoside, a useful intermediate for the synthesis of branched oligosaccharides. *Carbohydr. Res.* **1994**, *252*, 69-84.
37. Watanabe, Y.; Komoda, Y.; Ebisuya, K.; Ozaki, S. An efficient phosphorylation method using a new phosphitylating agent, 2-diethylamino-1,3,2-benzodioxaphosphepane. *Tetrahedron Letters* **1990**, *31* (2), 255-256.
38. Bruccoleri, R. E.; Karplus, M. Conformational Sampling Using High-Temperature Molecular Dynamics. *Biopolymers* **1990**, *29*, 1847-1862.
39. Frank, M. *In Silico Biology* **2002**, *2*, 427-439.
40. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytic treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127-6129.
41. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry B* **1997**, *101* (16), 3005-3014.
42. Engelsen, S. B.; Cros, S.; Mackie, W.; Pérez, S. A Molecular Builder for Carbohydrates: Application to Polysaccharides and Complex Carbohydrates. *Biopolymers* **1996**, *39*, 417-433.
43. Verlet, L. Computer Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98-103.

44. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327-341.
45. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
46. Crouzy, S.; Fauvelle, F.; Debouzy, J. C.; Goschi, M.; Chapron, Y. Investigation of the alpha-cyclodextrin-myo-inositol phosphate inclusion complex by NMR spectroscopy and molecular modeling. *Carbohydr. Res.* **1996**, *287*, 21-35.
47. *MPQC*, Sandia National Laboratories, Livermore, CA, USA, 2004
48. Pérez, S.; Kouwijzer, M. L. C. E.; Mazeau, K.; Engelsens, S. B. Modeling polysaccharides: Present status and challenges. *J. Mol. Graphics* **1996**, *14* (6), 307-321.
49. Tvaroska, I.; Hricovini, M.; Petrakova, E. An attempt to derive a new Karplus-type equation of vicinal proton-carbon coupling constants for C-O-C-H segments of bonded atoms. *Carbohydr. Res.* **1989**, *189*, 359-362.
50. Bax, A.; Davis, D. G. MLEV-17-Based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **1985**, *65* (2), 355-360.
51. Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **1993**, *31* (3), 287-292.
52. Meissner, A.; Sørensen, O. W. Measurement of J(H,H) and long-range J(X,H) coupling constants in small molecules. Broadband XLOC and J-HMBC. *Magn. Reson. Chem.* **2001**, *39* (1), 49-52.
53. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPIPE - A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995**, *6* (3), 277-293.

Table 1. Resonance assignment table for α -maltose-3'-phosphate (1) and α -maltose-6'-phosphate (2)

non. red.	(1)		(2)	
	δ_H	δ_C	δ_H	δ_C
1	5,56	98,2	5,48	98,9
2	3,51	72,6	3,56	72,1
3	3,58	72,6	3,71	73,3
4	3,36	69,2	3,36	69,8
5	3,61	72,6	3,78	76,3
6a	3,77	60,5	3,91	60,8
6b	3,70	60,5	3,73	60,8
red.				
1	5,16	91,5	5,22	91,8
2	3,64	71,8	3,58	71,4
3	4,36	77,3	3,96	73,5
4	3,73	73,9	3,70	76,1
5	3,89	69,9	4,05	69,1
6a	3,85	60,9	4.16 4.01	64,2
6b	3,71	60,9	4.10 4.06	64,2

Table 2. Experimental values of the heteronuclear $^3J_{C1-H4'}$ and $^3J_{C4'-H1}$ coupling constants for α -maltose-3'-phosphate and α -maltose-6'-phosphate measured using ^{13}C J-HMBC spectra recorded at 18.8 T, along with the corresponding couplings calculated from simulations, both in continuous (GBSA) and explicit (TIP3P) solvent, and also data from previous X-ray and adiabatic mapping studies of α -maltose

maltose-3-phosphate	$^3J_{H1-C4'}$	$^3J_{C1-H4'}$
experimental	2.3	3.4
CSFF-GBSA	3.0	5.5
CSFF-TIP3P (charge 0)	2.5	3.1
CSFF-TIP3P (charge -1)	2.5	3.1
CSFF-TIP3P (charge -2)	2.3	3.0
maltose-6-phosphate		
experimental	4.6	3.2
CSFF-GBSA	4.1	2.9
CSFF-TIP3P (charge 0)	3.2	3.9
CSFF-TIP3P (charge -1)	3.3	3.9
CSFF-TIP3P (charge -2)	3.5	4.0
maltose		
Experimental (Corzana et al.)	4.1	4.9
CSFF-TIP3P (Corzana et al.)	3.4	3.9
X-ray (Takusagawa et al.)	5.6	5.6
MM3 (Dowd et al.)	5.0	5.2

Chart 1: Chemical synthesis of maltose-3'-phosphate **4** and maltose-6'-phosphate **5** in the form of their disodium salts

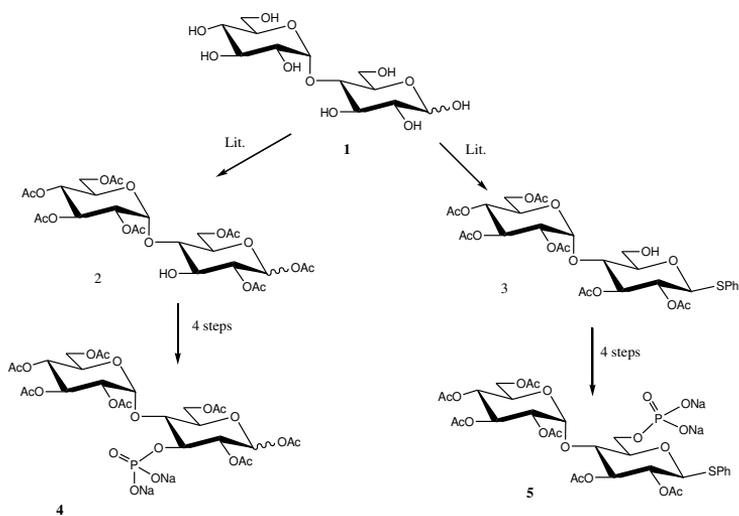


Figure 1. Molecular structure of α -maltoside 3'-O-phosphate (A) and α -maltoside 6'-O-phosphate (B) in their fully protonated state, including the torsions defined in the nomenclature section

Figure 2. Population distribution maps calculated from high temperature molecular dynamics of (A) maltose-3'-phosphate and (B) maltose-6'-phosphate. Conformational statistics is calculated from counting visitations in a grid with a field size of 2x2 degrees, and the conformational energy is derived from a Boltzman relationship.

Figure 3. Population distribution maps calculated from molecular dynamics in explicit solvation of maltose-3'-phosphate (left column) and maltose-6'-phosphate (right column). The upper row is fully protonated (charge 0), the middle row is charge -1, and the lower row is charge -2. Contours are drawn at 0.1, 0.01, 0.001, and 0.0001 population levels. The crystal structure is marked by a filled circle.

Figure 4. Cumulative histogram of the torsional angle distribution, defining the geometries of the substituted phosphates. The data is normalized so that the sum of all bars equals 1. Left column from the top and down shows the χ_3 and χ_7 torsions in maltose-3'-phosphate. Right column from the top and down shows the χ_5 , χ_6 , and χ_7 torsions in maltose-6'-phosphate

Figure 5. Molecular structure of maltose-6'-phosphate with the phosphate group in the conformation found from X-ray data on glucose-6-phosphate (A) and this study (B)

Figure 6. 2D Hydration maps of maltose-3'-phosphate (left column) and maltose-6'-phosphate (right column). First row fully protonated (charge 0), second row is charge -1, and third row is charge -2. Values represented as shaded squares are maximum localized water density of shared water relative to the bulk water density calculated using 2D pair distribution functions.

Figure 7. Radial pair distributions for all phosphate oxygens in the deprotonated form of maltose-3'-phosphate (A) and maltose-6'-phosphate (B). Red curves are PO1', blue PO2', and green PO3'

Figure 8. Snapshot of most populated water bridge found in maltose-3'-phosphate (A), where the distance from HO2 to the water oxygen is 2.7Å and the distance from each of the two PO' to the water hydrogen is 2.8Å, and a shared water between PO1' and O2' (B) with a distance of 2.7 Å from both these oxygens to the water oxygen

Figure 9. The structure of double helical amylopectin with an attached phosphate group, ignoring the effects of the glucosidic bond, with phosphate in the 3-position (left) adapting the most populated conformation $\chi_3=83^\circ$, $\chi_7=230^\circ$ and 6-position (right) again adapting the most populated conformation using the optimal exocyclic torsions derived from molecular dynamics simulation data, $\chi_3=242^\circ$, $\chi_5=128^\circ$, and $\chi_7=66^\circ$.

Figure 10. Superimposition of the adiabatic map of MM3 for the maltosidic $\alpha(1\rightarrow4)$ linkage with the most populated potential energy well of (A) maltose-3'-phosphate and (B) maltose-6'-phosphate predicted from molecular dynamics simulations in explicit water and with the helical parameter contours. The number n is the number of repeats per helical turn, and h is the helical repeat advancement distance in Ångström. (---) shows left hand chirality, and (—) right hand chirality.

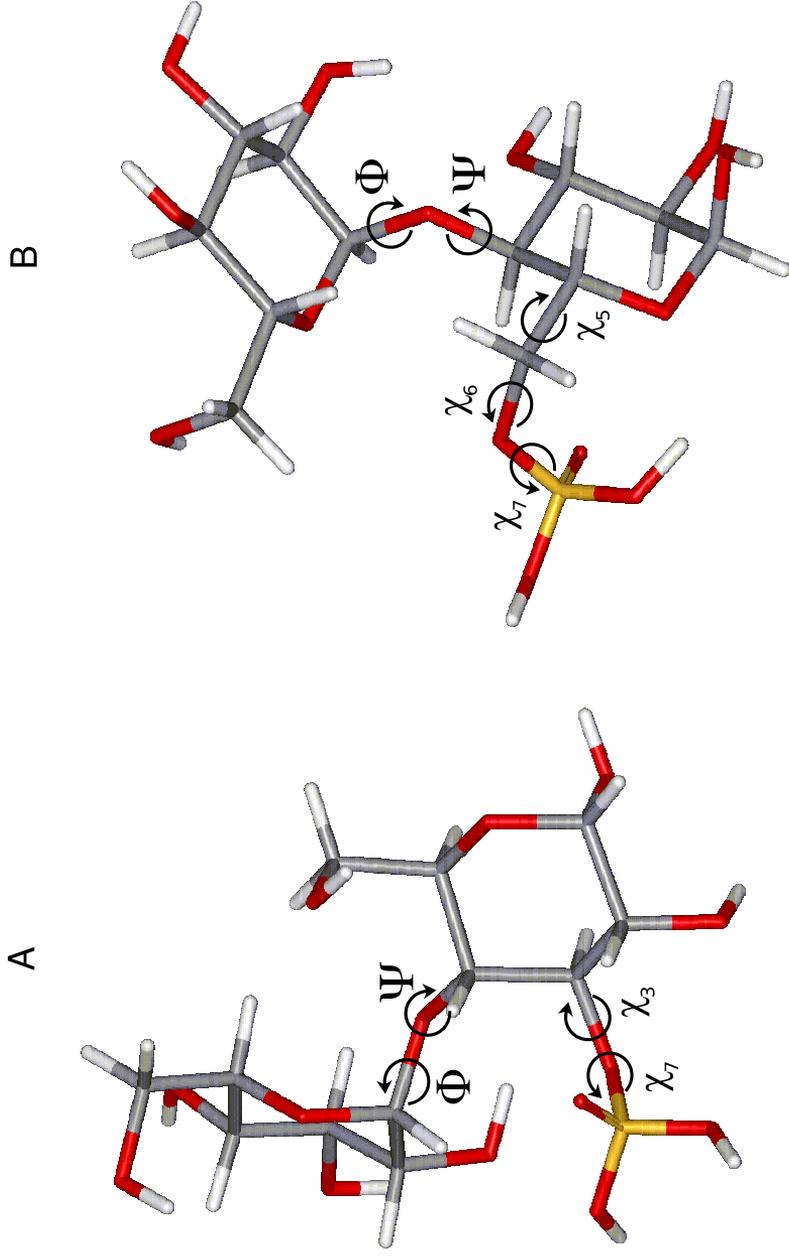


Figure 1
Hansen et al.

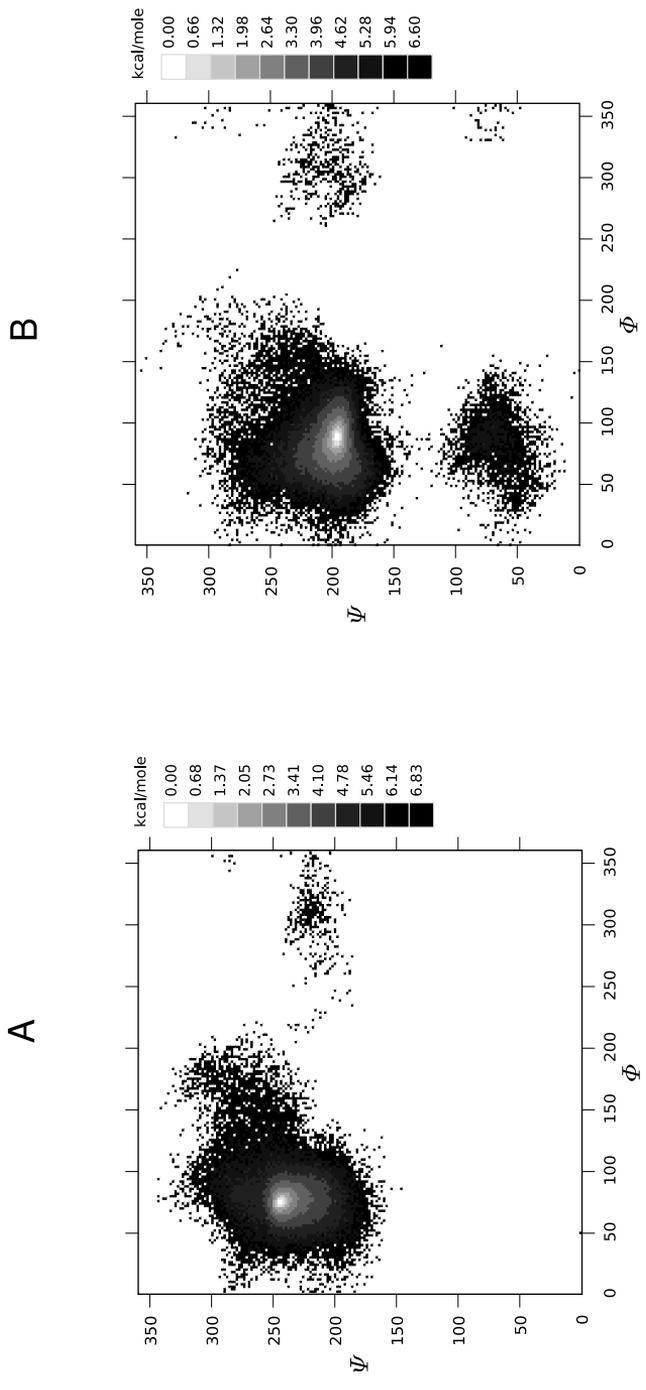


Figure 2
Hansen et al.

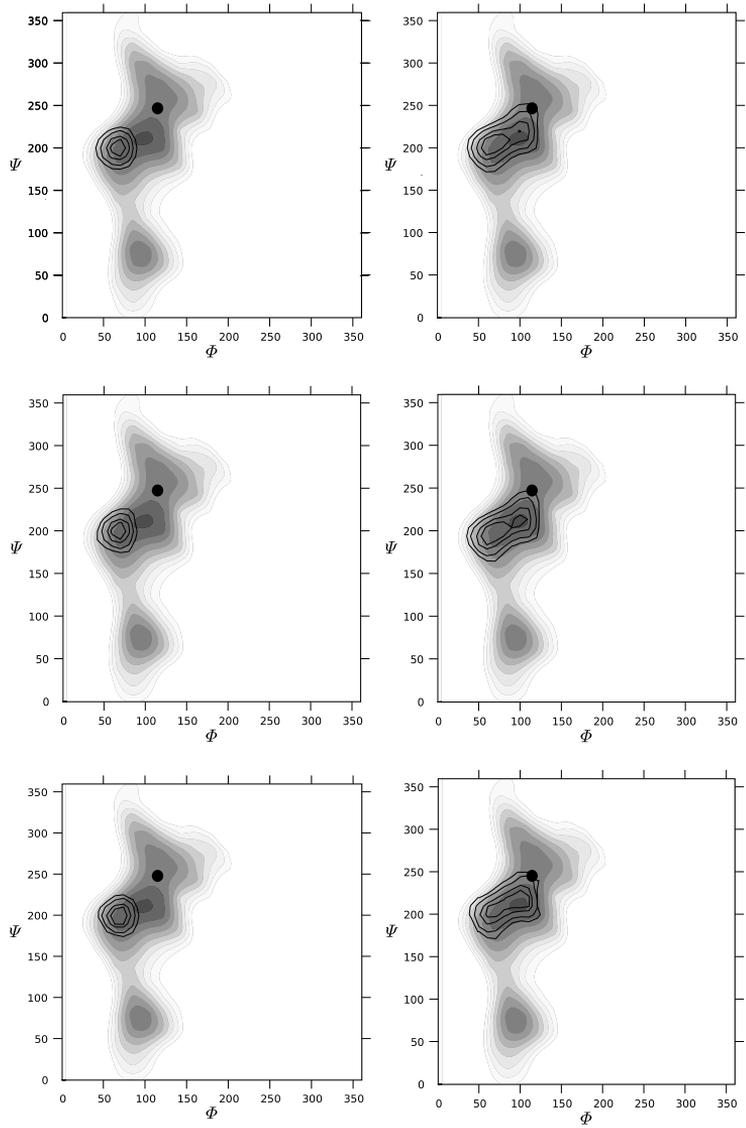


Figure 3
Hansen et al.

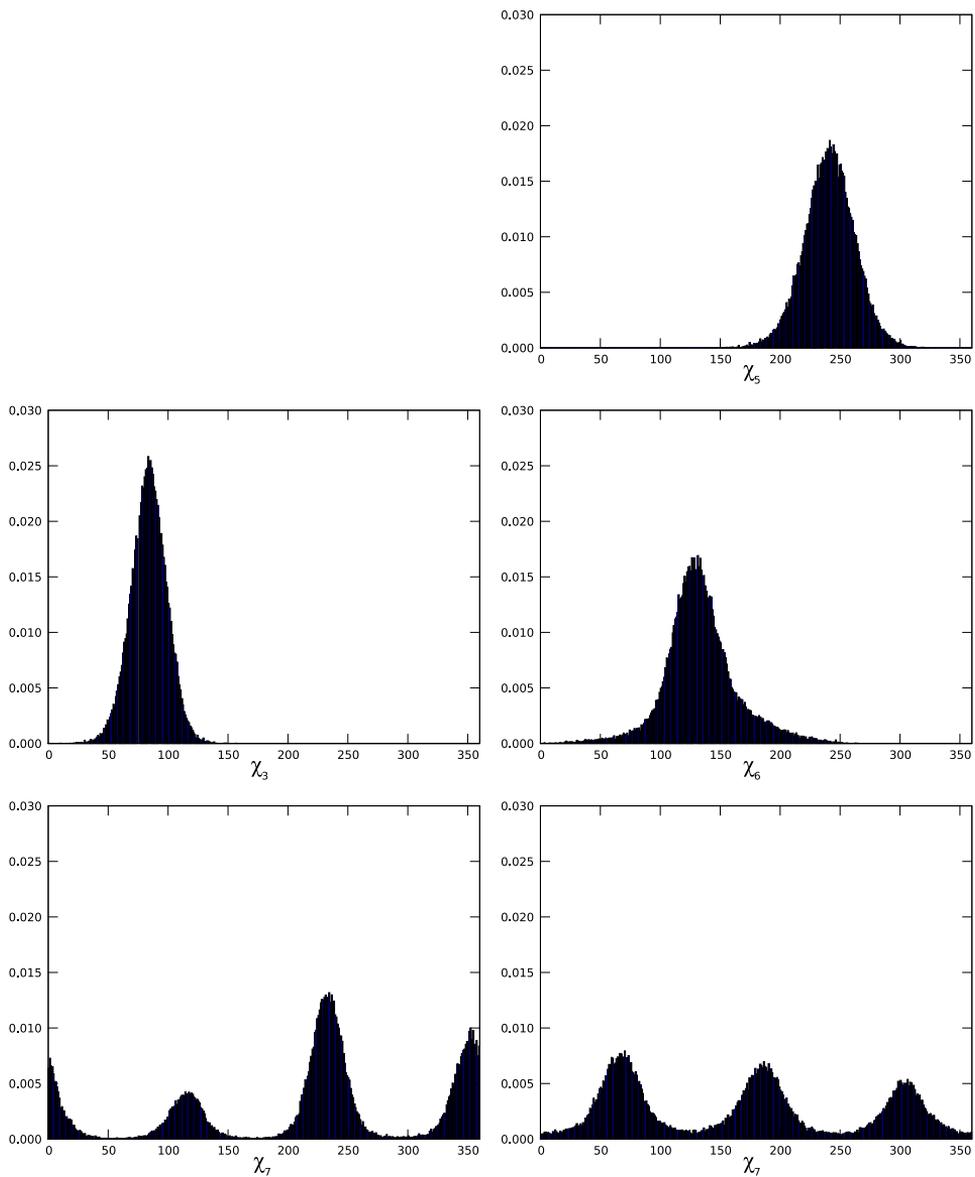


Figure 4
Hansen et al.

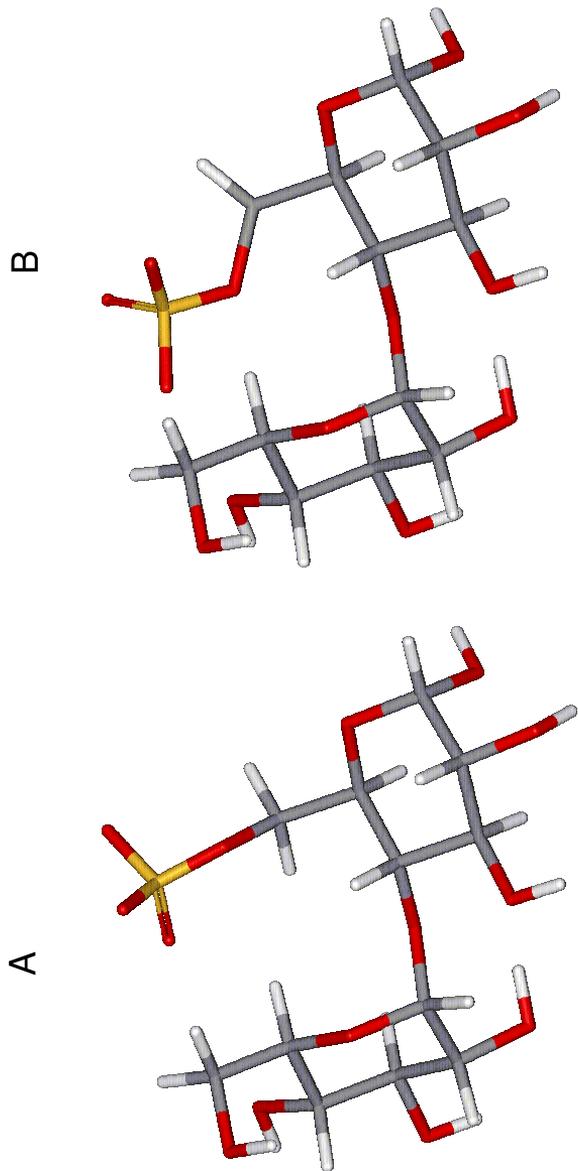


Figure 5
Hansen et al.

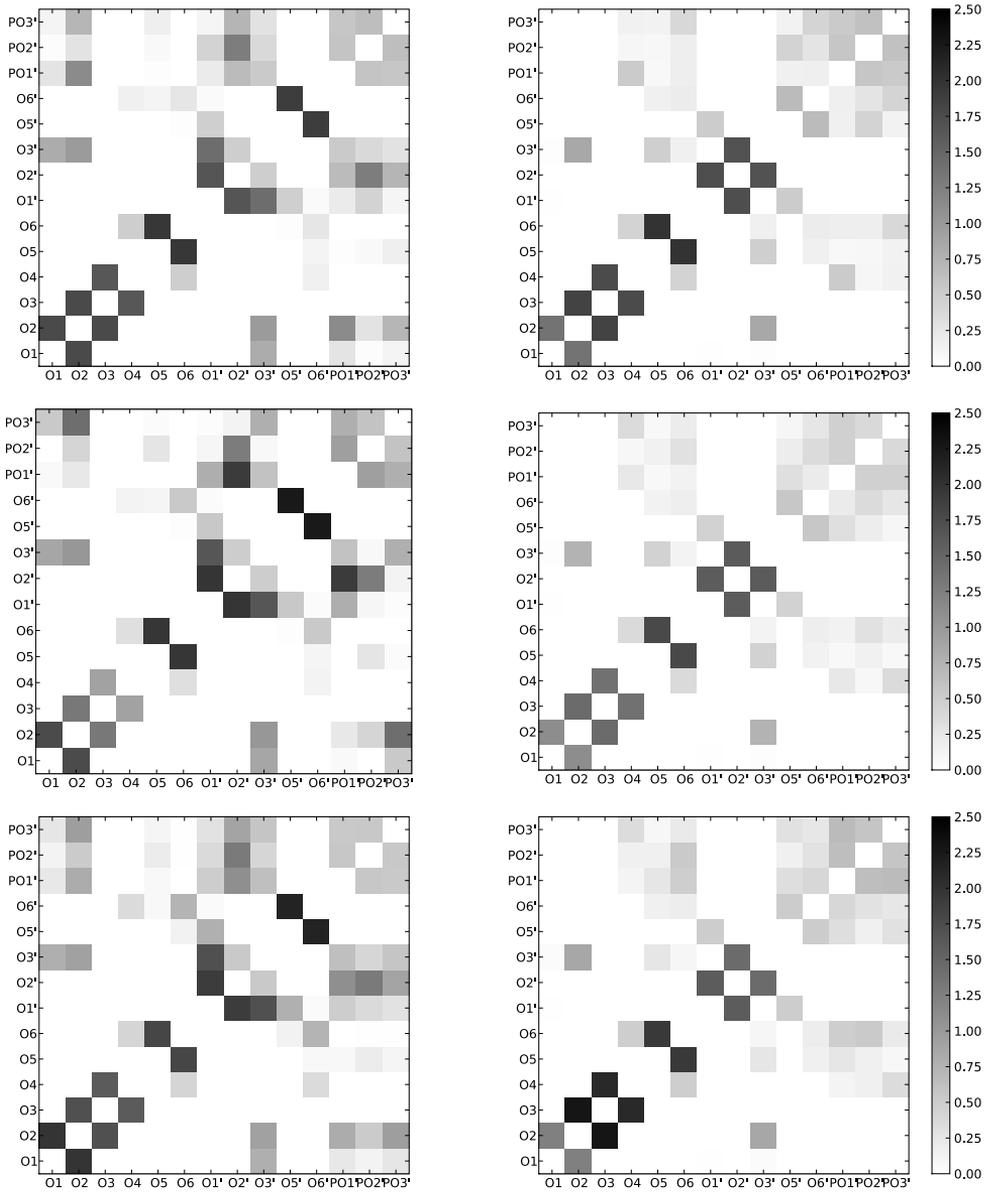


Figure 6
Hansen et al.

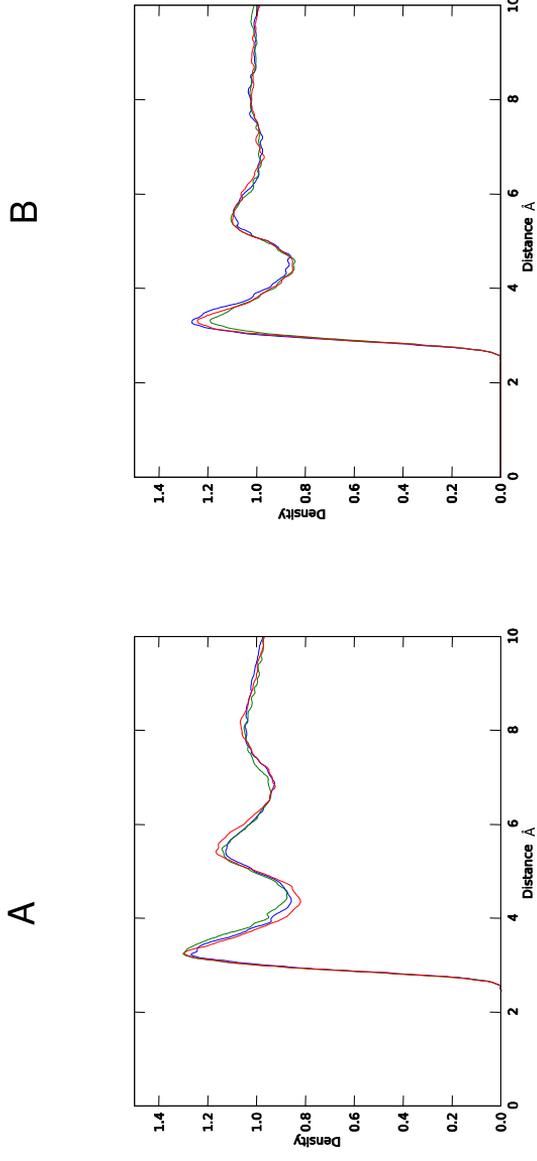


Figure 7
Hansen et al.

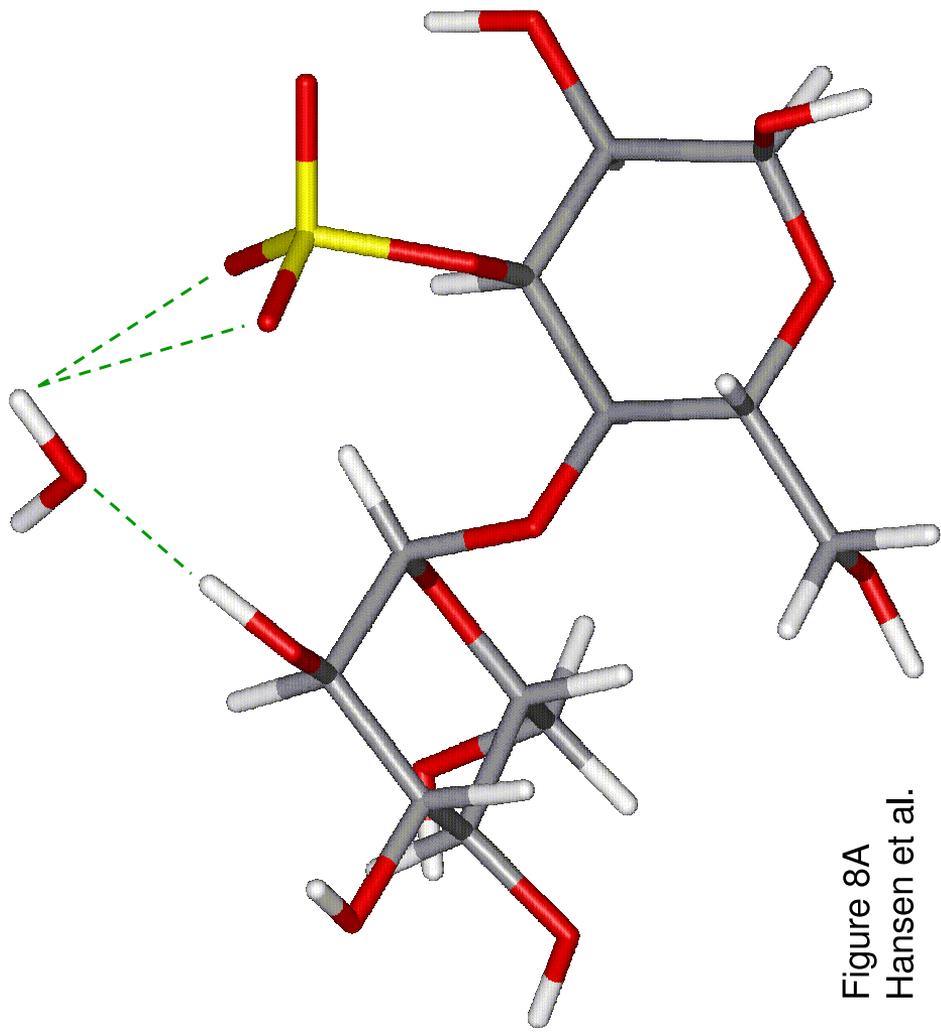


Figure 8A
Hansen et al.

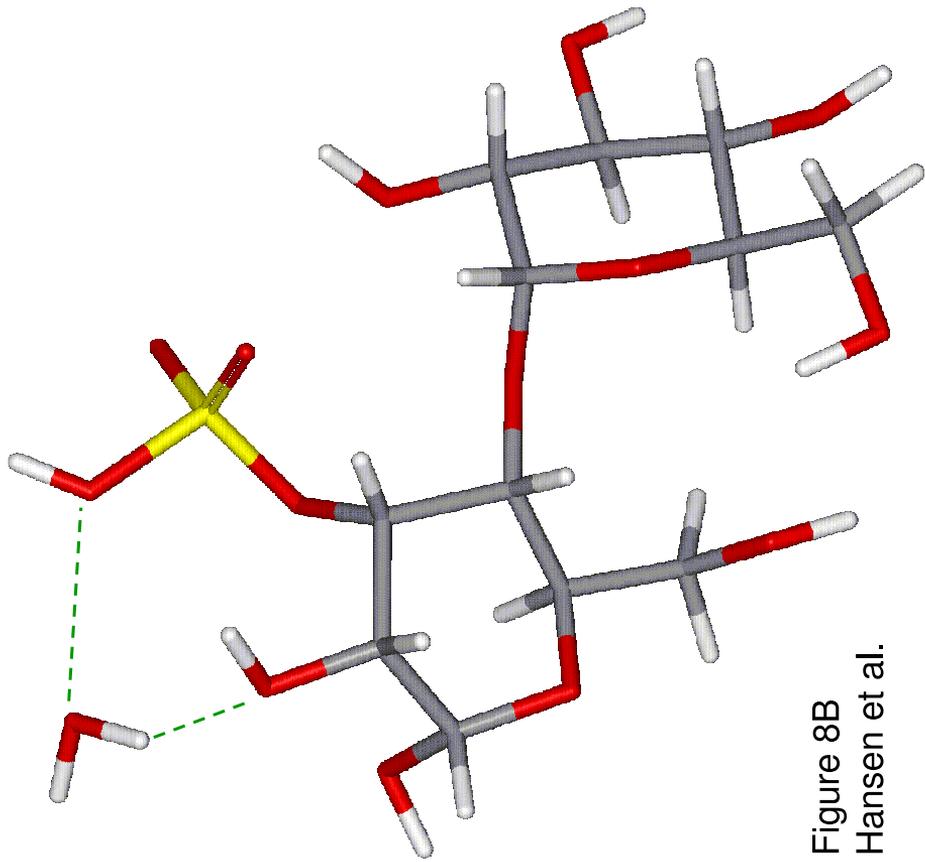
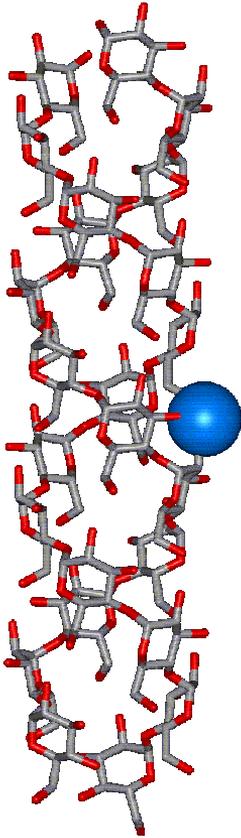


Figure 8B
Hansen et al.

A



B

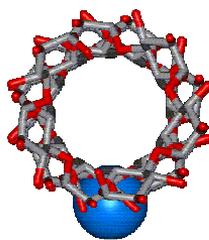
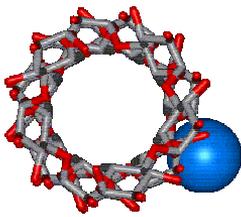
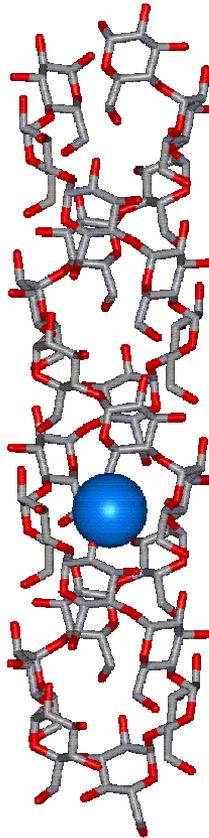


Figure 9
Hansen et al.

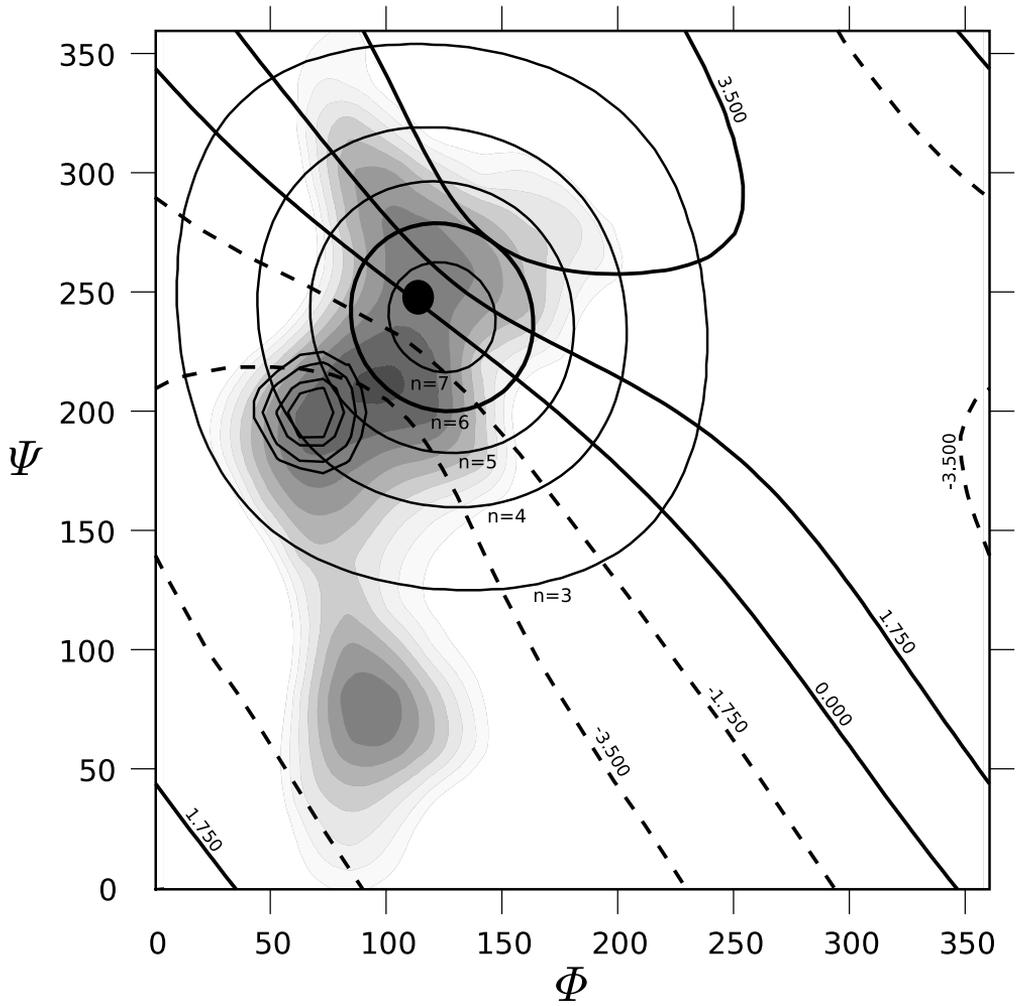


Figure 10A
Hansen et al.

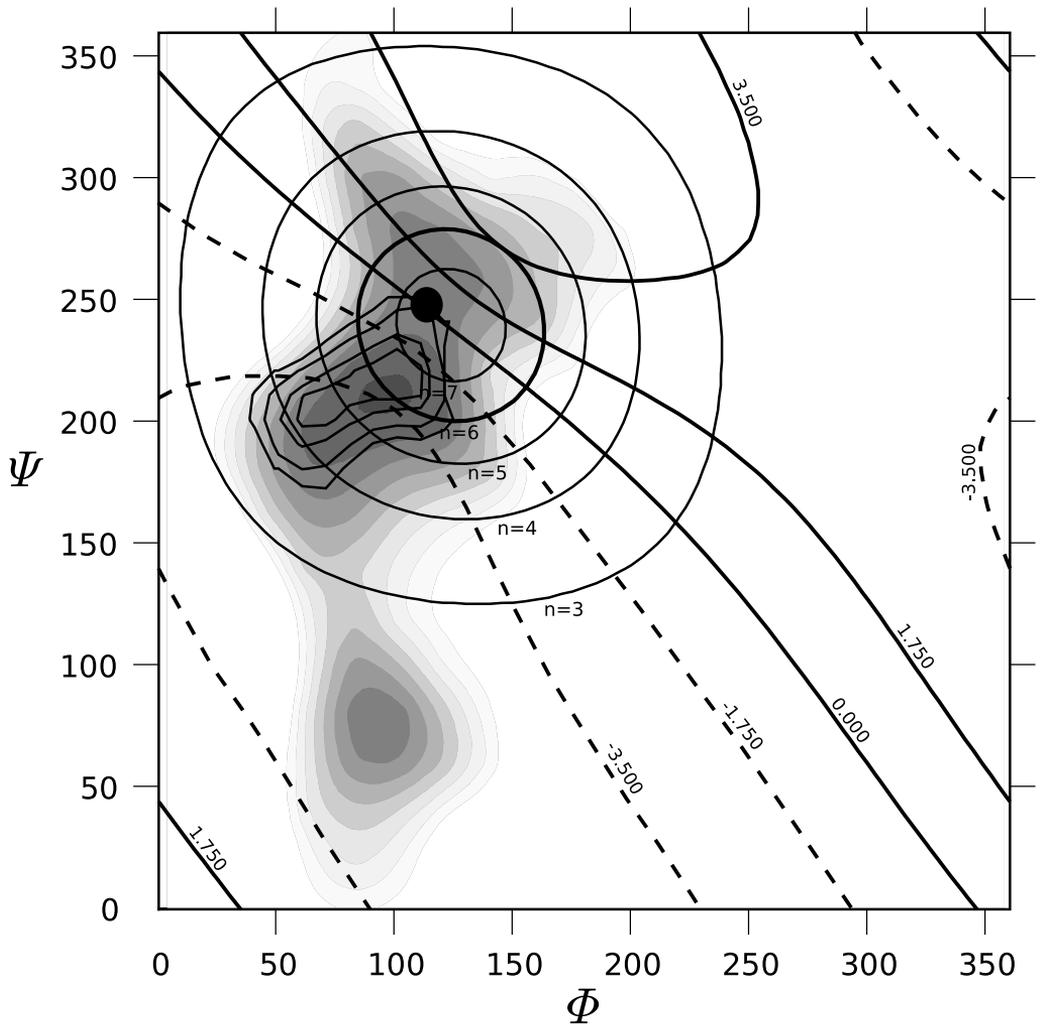


Figure 10B
Hansen et al.

Paper V

R. Bro, N. Viereck, M. Toft, H. Toft, P. I. Hansen and S. B. Engelsen,
Solving the nuclear magnetic cocktail party effect in systems biology
Angewandte Chemie International Edition, **Submitted**

Solving the Nuclear Magnetic Cocktail Party Effect in Systems Biology

Rasmus Bro*, Nanna Viereck, Marianne Toft, Henrik Toft, Peter I. Hansen, and Søren B. Engelsen

Nuclear magnetic resonance (NMR) spectroscopy has become an indispensable analytical technique for characterization of complex biological samples. The versatility and high resolution of NMR is exploited in diverse applications such as medical diagnostic body scanning, biomacromolecular structure determination and metabolite mapping in tissue and body fluids.

The extraction of quantitative and qualitative information from NMR spectra of complex biological samples necessitates appropriate data analytical tools. The first report on the use of multivariate data analysis in the form of principal component analysis (PCA)^[1] on NMR spectra appeared in 1983^[2], but it was not before the advent of high-throughput and hyphenated NMR methods such as LC-NMR and metabonomics^[3] that quantitative NMR became a focal point of international research. The future perspectives in phenotyping for personalized nutrition and personalized drug treatment has only just begun^[4].

The mathematical basis for the metabonomic advances has been the use of multivariate models such as PCA. While powerful, such data analytical tools *cannot* provide direct information on the underlying chemistry due to the inherent ‘rotational ambiguity’ explained below. This has important implications, for example in the search for biomarkers which requires a causal interpretation to avoid spurious results.

In quantitative NMR spectroscopy, a set of I samples is measured as a function of J chemical shifts. Each spectrum may be contained in one row of an $I \times J$ matrix, \mathbf{X} . If each spectrum is a sum of contributions from F underlying spectra (i.e. F chemical analytes) held in a matrix \mathbf{B} ($J \times F$), then \mathbf{X} can be modeled as a bilinear model as the product of the sample-specific concentrations, a_{if} (held in \mathbf{A} of size $I \times F$) and the pure analyte spectra [Eq. (1)]

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T + \mathbf{E} \quad (1)$$

where \mathbf{E} contains measurement noise. The problem, in practice, is to determine \mathbf{A} and \mathbf{B} from the measured \mathbf{X} . A natural proposition is to estimate these parameters from \mathbf{X} using a least squares or similar

fitting procedure. However, because [Eq. (2)]

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{B}^T = \mathbf{A}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{B}^T = \mathbf{C}\mathbf{D}^T \quad (2)$$

for any non-singular $F \times F$ matrix \mathbf{Q} , then any estimate may equally well lead to estimates of \mathbf{C} and \mathbf{D} instead of the true \mathbf{A} and \mathbf{B} , i.e. the solution may be any one of a class of alternative solutions due to this mathematical rotational ambiguity. As a result, bilinear models such as PCA cannot provide pure analyte NMR spectra.

In 1944 Cattell^[5] described a fundamental way to solve the rotational indeterminacy. This idea, called parallel proportional profiles, became operational in 1970 when Harshman^[6] developed the PARAFAC (parallel factor analysis) model. The principle in PARAFAC is to use several sets of matrices which provide information on the same basic entities (e.g. concentrations and spectral profiles), but do so in different proportions^[7]. Such different proportions may arise by measuring the basic matrix data, \mathbf{X} , as a function of a new dependent variable, for example analyte diffusion in NMR spectroscopy. The basic two-way data tables are thereby extended to three-way tables. A three-way table may be described as an $I \times J \times K$ array $\underline{\mathbf{X}}$ with typical elements x_{ijk} containing the measurement of sample i at variable j on occasion k . The PARAFAC model decomposes such three-way data into a set of trilinear terms and residuals. The algebraic model structure is [Eq. (3)]

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (3)$$

For a given choice of F , the number of chemical analytes, the three-way array is decomposed into so-called scores \mathbf{A} ($I \times F$) with elements a_{if} , loadings \mathbf{B} ($J \times F$) with elements b_{jf} and loadings \mathbf{C} ($K \times F$) with elements c_{kf} . Unlike PCA, the PARAFAC solution is *unique* and directly estimates the underlying physical features of the data, if these approximately follow the model^[8,9]. Hence, there is no rotational ambiguity.

Using diffusion-edited 2D NMR spectroscopy we can obtain data that follow the PARAFAC model^[10]. The signal intensity is recorded as a function of chemical shift and of the squared field gradient strength following a low-rank trilinear structure [Eq. (4)]:

$$I_{\Omega,\delta,k} = \sum_{f=1}^F S_{\Omega f} \cdot A_{\delta f} \cdot C_{kf} \quad (4)$$

where $S_{\Omega f}$ denotes the spectral intensity at chemical shift Ω for compound f , and where $A_{\delta f} = \exp[-D_f \delta^2 \Delta^2 - R_f]$ denotes the attenuation due to diffusion at gradient strength δ for compound f . The parameter C_{kf} denotes the concentration in sample k of compound f . When comparing the above model with the PARAFAC model, it follows that the diffusion-edited 2D NMR spectra can be separated into individual constituent spectra, diffusion profiles and relative concentrations.

[*] Prof. R. Bro, Dr. N. Viereck, Dr. P. I. Hansen, Prof. S. B. Engelsen
Department of Food Science, Faculty of Life Sciences
University of Copenhagen
Rolighedsvej 30, 1958 Frederiksberg C, Denmark
Fax: (+45) 35 33 32 45
E-mail: rb@life.ku.dk
Homepage (optional):

Dr. M. Toft
Umetrics AB
Stortorget 21, 211 34 Malmö, Sweden

Dr. H. Toft
Protein Structure and Biophysics
Novo Nordisk A/S
Novo Nordisk Park, 2760 Måløv, Denmark

Supporting information for this article is available on the WWW under <http://www.models.life.ku.dk/research/data/>

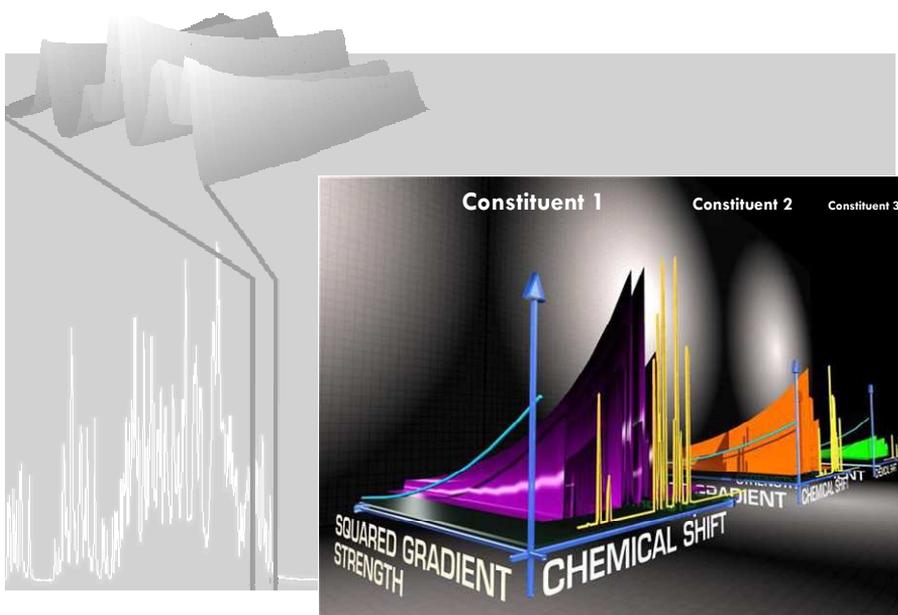


Figure 1. NMR data and hypothetical PARAFAC model. In the background to the left is an example of a diffusion edited 2D NMR spectrum recorded as a function of chemical shift and as a function of squared gradient strength (shown for part of the spectrum) suitable for decomposition by PARAFAC. To the right, a schematic view of a three-component PARAFAC model resolving the data from several samples into estimated diffusion profiles and estimated spectra of three constituents/analytes. For each sample, a set of estimated scores (relative concentrations) determines the amount of each of the three analytes.

As a simple example on the power of PARAFAC analysis, eight mixtures of glucose, maltose and maltotriose originating from a fractional factorial design were measured using a 2D diffusion edited pulse experiment (Figure 1). The self-diffusion of the three compounds is quite similar (approx. $7, 5$ and $4 \times 10^{-6} \text{ cm}^2/\text{s}$, respectively^[11]) and the individual analyte spectra are practically similar with highly overlapping spectra especially in the pyranose proton range. The three-component PARAFAC solution is shown in Figure 2 for one of the three recovered components (glucose). As can be seen, the measurements of the mixtures are mathematically

separated into pure analyte information. For example, the estimated relative concentrations of the analytes are in good agreement with the actual concentrations as shown for glucose. This PARAFAC solution has been achieved without assumptions on NMR line shape or distribution of analyte concentrations; only the number of components had to be decided *a priori*. Normally, in complex biological matrices the number of components is not known and has to be determined from the data using mathematical diagnostic tools.^[12]

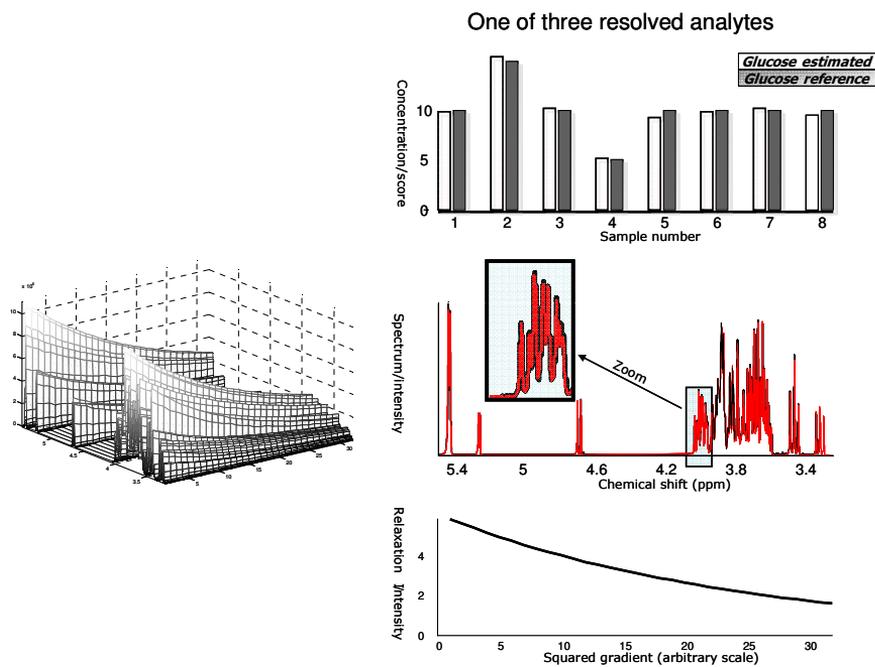


Figure 2. The PARAFAC solution for the designed NMR data. Mixtures of glucose, maltose and maltotriose were made according to a reduced design using only eight samples. To the left, the raw diffusion-edited 2D NMR spectrum of a mixture is shown and to the right, the PARAFAC solution for glucose as an example. The solution is given as (top) the estimated relative concentrations compared to the actual, (middle) estimated relative spectrum as well as a reference spectrum of glucose and (bottom) the estimated diffusion profile.

A more complex set of samples will help illustrate the complexity of metabolomic data that can be handled. People suffering from Type II diabetes or dyslipidemia have an increased risk of coronary heart disease (CHD). The individual risk of CHD is related to the distribution of cholesterol and triglyceride in different types of lipoproteins^[13;14] in the blood. It is therefore important to be able to monitor the lipoprotein profile in, for example, medical diagnostics and human nutrition studies. Lipoproteins are micellar triglyceride and cholesterol transport vehicles in the blood and can be divided into the main fractions: Very-Low-Density-Lipoproteins (VLDL), Intermediate-Density-Lipoproteins (IDL), Low-Density-Lipoproteins (LDL) and High-Density-Lipoproteins (HDL). Lipoproteins in blood can be measured using high-resolution NMR spectroscopy,^[15] but due to their similar composition the lipoprotein fractions have very similar signals. The signals are minimally shifted in frequency due to the different densities of the lipoproteins, giving rise to different local magnetic fields. Diffusion-edited 2D NMR for analyzing lipoproteins takes advantage of the fact that

lipoproteins vary in size, which according to the Stokes-Einstein relationship will lead to different diffusion properties.

We demonstrate an example using 20 blood plasma samples and 2D diffusion-edited NMR data represented by the methylene region indicative of the triglycerides and cholesterol of the lipoproteins (Figure 3A). The three-way data are resolved by a four-component PARAFAC model yielding four estimated spectra and diffusion profiles as well as four concentrations for each sample. The spectra look similar, reflecting a shift in the methylene peak from the lipid fractions of the lipoproteins. The diffusion profiles are similar (Figure 3C), but result in molecular diameters of 42 nm, 26 nm, 18 nm and 8 nm when using the Stokes-Einstein relationship and assuming spherical geometry. This is in good agreement with experimental range estimations of the four lipoprotein main fractions: 30-80 nm, 25-35 nm, 18-25 nm and 5-12 nm for the VLDL, IDL, LDL and HDL fractions, respectively^[16]. Thus, despite the continuous nature of the distribution of lipoproteins, PARAFAC is able to model the NMR data in what may be termed quasi-particles, representing the VLDL, IDL, LDL and HDL fractions.

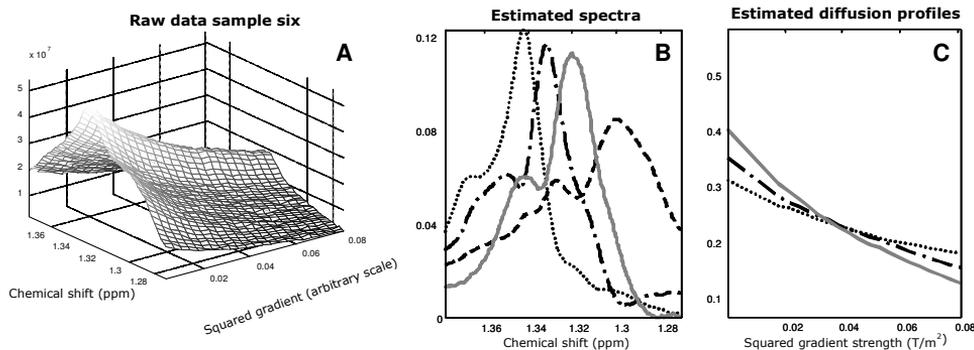


Figure 3. The PARAFAC solution for lipoprotein 2D NMR data. To the left (A), the raw data of one sample is shown, in the middle (B), the spectral estimates are shown and to the right (C), the estimated diffusion profiles are shown.

These applications illustrate the potential for multi-way analysis of complex 2D-NMR data in systems biology. Separation of analyte signals can be accomplished, despite the fact that the chemical analytes display near identical NMR spectra and have self-diffusion coefficients in the same order of magnitude. It is the combined modulation of the chemical shift spectra and the diffusion profiles that allow PARAFAC to perform *mathematical chromatography*.

Received: ((will be filled in by the editorial staff))

Published online on ((will be filled in by the editorial staff))

Keywords: Metabolism · Metabonomics · NMR spectroscopy · PARAFAC · Tensor analysis

- [1] H. Hotelling, *J. Educ. Psychol.* **1933**, *24*, 417-441.
 [2] D. Johnels, U. Edlund, H. Grahn, S. Hellberg, M. Sjoström, S. Wold, S. Clementi, W. J. Dunn, *J. Chem. Soc. Perkin Trans. 2* **1983**, 863-871.
 [3] J. K. Nicholson, J. Connelly, J. C. Lindon, E. Holmes, *Nat. Rev. Drug Discov.* **2002**, *1*, 153-161.

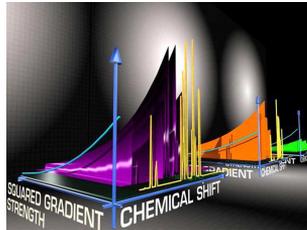
- [4] T. A. Clayton, J. C. Lindon, O. Cloarec, H. Antti, C. Charuel, G. Hanton, J. P. Provost, J. L. Le Net, D. Baker, R. J. Walley, J. R. Everett, J. K. Nicholson, *Nature* **2006**, *440*, 1073-1077.
 [5] R. B. Cattell, *Psychometrika* **1944**, *9*, 267-283.
 [6] R. A. Harshman, *UCLA working papers in phonetics* **1970**, *16*, 1-84.
 [7] R. Bro, *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149-171.
 [8] J. B. Kruskal, *Lin. Algebra Appl.* **1977**, *18*, 95-138.
 [9] N. D. Sidiropoulos, R. Bro, *J. Chemom.* **2000**, *14*, 229-239.
 [10] D. S. Johnson, *Progr. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203-256.
 [11] J. W. Lee, T. O. Kwon, I. S. Moon, *Carbon* **2004**, *42*, 371-380.
 [12] H. A. L. Kiers, R. Bro, *J. Chemom.* **2006**, *17*, 274-286.
 [13] W. J. Mack, R. M. Krauss, H. N. Hodis, *Am. Heart Assoc.* **1996**, *16*, 697-704.
 [14] B. Lamarche, A. Tcherno, S. Moorjani, B. Cantin, G. R. Dagenais, P. J. Lupien, J. P. Despres, *Am. Heart Assoc.* **1997**, *95*, 69-75.
 [15] J. D. Otvos, E. J. Jeyarajah, D. W. Bennett, *Clin. Chem.* **1991**, *37*, 377-386.
 [16] M. Dyrby, M. Petersen, A. K. Whittaker, L. Lambert, L. Nørgaard, R. Bro, S. B. Engelsen, *Anal. Chim. Acta* **2005**, *531*, 209-216.

Entry for the Table of Contents

**Mathematical chromatography –
new possibilities**

Rasmus Bro*, Nanna Viereck, Marianne Toft, Henrik Toft, Peter I. Hansen, Søren B. Engelsen _____ **Page 1–
Page 4**

Solving the nuclear magnetic cocktail party effect in systems biology



The *cocktail party effect* describes the difficulty in separating co-occurring overlapping signals. This problem is well known in nuclear magnetic resonance analysis. Data from complex samples are mostly assessed visually which makes it impossible to identify the underlying constituents. A new approach coined *mathematical chromatography* can identify and quantify individual components. Lipoprotein profiling of human plasma samples is used as an example.

Paper VI

R. Bro, P.I. Hansen, N. Viereck, M. Dyrby, H.T. Pedersen, and S.B. Engelsen, *A new principle for unique spectral decomposition of 2D NMR data*, 195-203 The RSC proceeding for the 7th International Conference on Application of Magnetic Resonance in Food Science. 2004

A NEW PRINCIPLE FOR UNIQUE SPECTRAL DECOMPOSITION OF 2D NMR DATA

Rasmus Bro¹, Peter Ibsen Hansen¹, Nanna Viereck¹, Marianne Dyrby², Henrik Toft Pedersen³ and Søren B. Engelsen¹

Dept. of Food Science, Quality & Technology, The Royal Veterinary and Agricultural University, 1958 Frederiksberg C, Denmark

² Umetrics AB, 211 34 Malmö, Sweden

³ Virology & Molecular Toxicology, Novo Nordisk A/S, 2760 Måløv, Denmark

1 INTRODUCTION

Resolving pure component spectra from complex bilinear 2D NMR spectra of mixtures is highly desirable, but not possible at present. Extracting individual-analyte information from data such as diffusion-weighted spectra (DOSY - diffusion edited spectroscopy) would be important in many applications. The mathematical principle behind a new method for reaching this goal is outlined and its successful use is demonstrated on a series of mixtures in a severely reduced design. The potential of the new method applied to multi-parametric NMR spectroscopy appears to be inexhaustible and includes aiding tools in structural assignments and the recovery of pure spectra of impurities in low-rank mixtures.

1.1 Multi-way data

*Multi-way analysis*²⁰ is an emerging data analysis technique that has been successfully applied in several chemical fields^{1, 2, 9, 10, 16, 18, 21}. The key issue in *multi-way* analysis is to have access to boxes of data rather than tables of data. Usually, a spectrum is measured for each sample. Data for several samples are then gathered in a matrix/table. If, for example, a set of spectra is obtained from a sample, then the data from just one sample are contained in a matrix. For several samples a box of data is obtained (Figure 1). Such *multi-way* data can be modelled with specialized tools that take particular advantage of the data format. Most notably, the so-called PARAFAC model is an interesting alternative to traditional data analysis tools, because it allows resolving complex mixture measurements into the underlying single-component spectra.

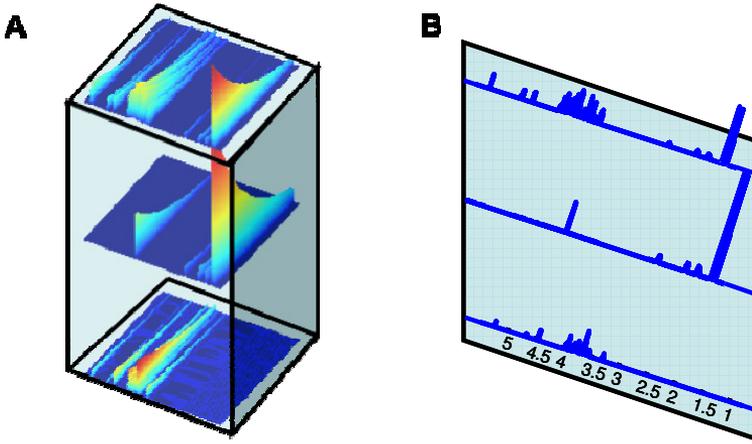


Figure 1 Structure of multi-way data (A) and of ordinary two-way data tables (B). Rather than having one spectrum for each sample, a set of spectra can lead to multi-way data.

1.1 Trilinear PARAFAC modeling

The PARAFAC model was developed in 1970 by R. Harshman¹¹. The model provides a rational way to analyse multi-way data. The box of three-way data, called $\underline{\mathbf{X}}$, is modelled as a sum of F contributions. Ideally, each of the F contributions describes one underlying chemical component. Each component consist of one score vector, \mathbf{a}_f , and two loading vectors, \mathbf{b}_f and \mathbf{c}_f . If $\underline{\mathbf{X}}$ is an $I \times J \times K$ array consisting of I samples measured at J variables at K occasions, then each score \mathbf{a}_f will be a vector of length I and similar for \mathbf{b}_f and \mathbf{c}_f . In scalar notation, the model can be written in terms of the individual elements of $\underline{\mathbf{X}}$:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf}, \quad i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K \quad (1)$$

where a_{if} is the i 'th element of \mathbf{a}_f etc. Details of the model and its derivation can be found in the literature. The one feature of this model which distinguishes it markedly from, for example, principal component analysis (PCA)^{13, 15, 17} is that no artificial constraints such as orthogonality need to be imposed to obtain a unique solution. If the fundamental laws prescribe that the data can be approximated as the above trilinear model, then PARAFAC can uniquely estimate these physical parameters. With PCA it is generally not possible to determine physically meaningful parameters, because there are infinitely many solutions to the same algebraic form as the PCA model. One "arbitrary" of these solutions is chosen in PCA on the basis of the constraints of orthogonality and maximum variance. The implication of the uniqueness properties of the PARAFAC model is that PARAFAC is able to resolve the underlying individual contributions directly from a mixture. Hence, PARAFAC performs *mathematical chromatography*.

2 EXAMPLE: HIGH-RESOLUTION NMR

In 2D diffusion-edited NMR the signal intensity is recorded as a function of chemical shift as well as of gradient strength:

$$I_{\delta g k} = \sum_{f=1}^F S_{\delta f} \cdot A_{g f} \cdot C_{k f} \quad (2)$$

where $S_{\delta f}$ denotes the spectral intensity at chemical shift δ for compound f , and where $A_{g f} = \exp[-D_f \gamma^2 g^2 \Delta' - R_f]$ denotes the attenuation due to diffusion (D_f) at gradient strength g for compound f , and $C_{k f}$ denotes the concentration in sample k of compound f . When comparing the above model with the PARAFAC model, it follows that 2D diffusion-edited data can be separated into individual constituent spectra (\mathbf{b}_f), diffusion profiles (\mathbf{c}_f) and concentrations (\mathbf{a}_f). To date, such a mathematical separation has not been possible. An example will be used to show the results of applying PARAFAC modelling on 2D diffusion-edited data.

2.1 Experimental setup

2.1.1 Samples Mixtures of glucose, lactose and iso-leucine were made according to a reduced design (Figure 2). Only seven samples were used in this work. These samples were prepared with the three compounds at 0, 10 and 20 mM. The compounds were dissolved in a pH 6.0 phosphate buffer (0.1 M) in D_2O . The samples were prepared as 500 μ l of the above described solutions with 50 μ l 1mg/ml TSP (3-trimethylsilyl-1-[2,2,3,3- $2H_4$] propionate) in D_2O added as chemical shift reference (0.0 ppm).

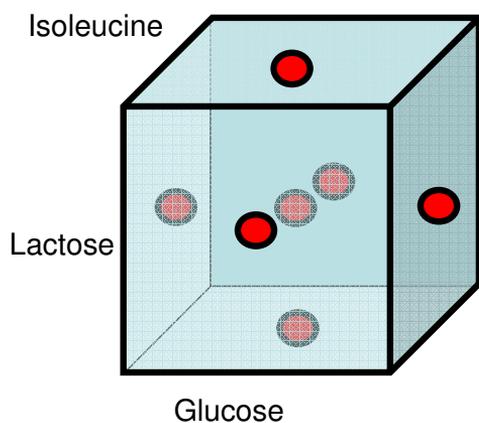


Figure 2 Design of the three compounds (glucose, lactose and iso-leucine) at three concentration levels

2.1.2 ^1H nuclear magnetic resonance. Proton spectra were measured at 298 K on a Bruker DRX600 operating at 14.1 T using a flow-NMR system and a 120 μl flow-probe. The pulse program used was a stimulated echo experiment with bipolar gradient pulses and a longitudinal eddy current delay as well as pre-saturation to suppress residual H_2O signal. The gradient was varied in 32 steps from 0.05 to 0.95 of maximum gradient power using a squared ramp.

2.2 Results

As can be seen in the subset of data shown in Figure 3, there are severe overlaps between the spectra and between the squared gradient attenuation profiles (from now on denoted diffusion profiles) of different analytes. Even though some spectral regions have limited overlap, the diffusion profiles are highly correlated throughout and hence, it is not possible to separate the mixture data into pure contributions immediately.

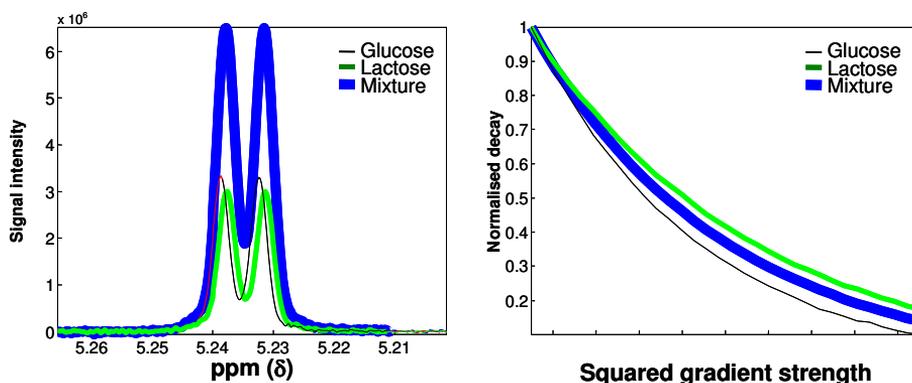


Figure 3 Spectral region (left) and diffusion profiles (right) of the α -anomer proton of lactose, glucose and a mixture of the two

When modelling the mixture data with a PARAFAC model, no assumptions on spectral band shape or distribution of concentration needs to be made. The only assumptions required for PARAFAC modelling are:

1. The data is low-rank trilinear
2. The number of components is decided by the user
3. The concentrations, spectra and profiles of different analytes are not identical

Based on the number of components and the measured data *only*, the PARAFAC solution is determined.

2.2.1 Elaboration on assumptions. It is perhaps instructive to consider the above verbal description of assumptions in somewhat more detail. The data has to be low-rank trilinear. This implies, for example, that the spectrum of an analyte does not change shape (significantly) at different gradient strengths. This is the crucial premise of the usefulness of the PARAFAC model. The model provides *one* spectrum for each analyte, so this can only be meaningful if indeed; the shape remains unaltered at different conditions. Experimental conditions have to be set so that this assumption is reasonably valid. This means proper tuning and matching as well as working in range of gradient strength where the output of the gradient amplifier is well defined for the selected diffusion time. Furthermore, the combination of diffusion time and maximum gradient power should not be set so signals becomes extinct, and it is important that the sample temperature is kept sufficiently low and constant in order to avoid convection current which will strongly influence the acquired signal. In addition, pH has to be constant from sample to sample to further avoid chemical shift differences.

Even though the low-rank trilinear requirement is not fulfilled, it is possible to perform meaningful multi-way analysis on this kind of data, but pure spectra etc. will not be obtained. Examples of such data analysis have been investigated by Dyrbjerg *et al.* when exploring lipoprotein signals in DOSY spectra from human blood serum^{5,6}.

The number of components has to be decided by the user. For the given data, it is known that the number of components should be three: one for each analyte. In general, however, such results are not known and the number of components therefore has to be determined from the data. This is similar to the situation in principal component analysis⁷ and there are numerous tools for deciding on the appropriate number of components in PARAFAC^{3, 4, 20}. Many of these are even simpler than for principal component analysis, because the intrinsic uniqueness of the PARAFAC model can help in guiding^{4,12}.

The final assumption mentioned above is that the profiles are not identical. This unique condition can be stated more mathematically and stringently^{14, 19}, but conceptually it suffices to know that the main issue is that if, for example, the two self-diffusion coefficients are identical within the signal-to-noise ratio, then complete uniqueness cannot be guaranteed. One remarkable property that follows directly from the more stringent uniqueness results is that, in principle, only two samples are needed, even when many more than two underlying analytes are modelled. This is used in so-called second-order calibration^{8, 22} where one sample contains analyte and unknown interferents and the other sample is standard with known non-zero concentration of the analyte. From two such samples the concentration of the analyte can be determined in the unknown sample.

2.3 The PARAFAC model

The so-called scores of the PARAFAC model contain estimates of the concentrations of the three analytes. The scores are shown in Figure 4.

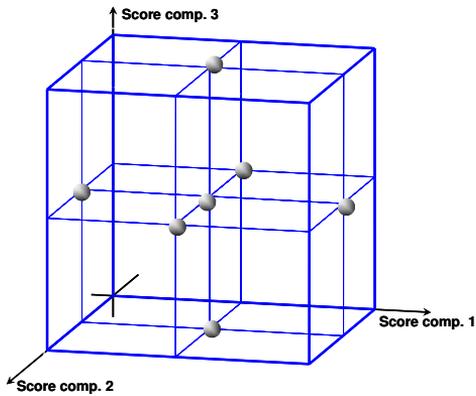


Figure 4 Estimated relative concentrations from PARAFAC model of the mixture data

When comparing with Figure 2 it is remarkable to see how well PARAFAC resolves the concentrations. However, since there is no way to determine the concentration in specific molar units from a spectrum, the PARAFAC model cannot provide concentration estimates on an absolute level. Hence, the scores shown in Figure 4 are relative concentrations that need to be scaled to the concentration of at least one sample or alternatively to a unit-concentration spectrum. This lack of absolute level is not specifically related to PARAFAC, but holds for *all* indirect measurements.

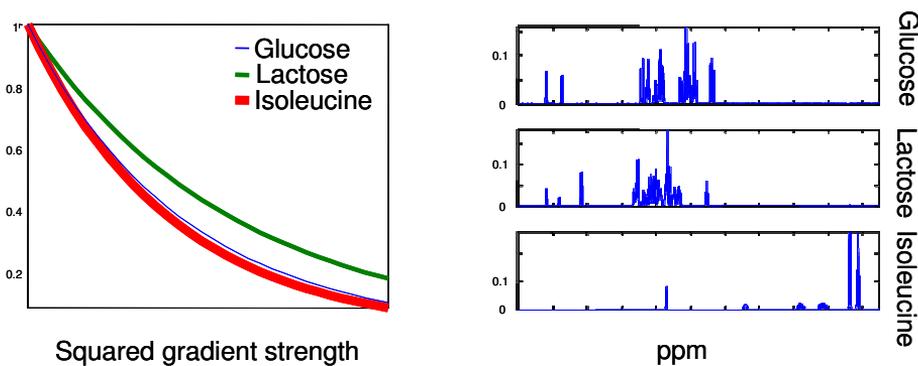


Figure 5 The PARAFAC resolved diffusion profiles (left) and spectra (right)

In Figure 5, the estimated gradient profiles and spectra are shown. In order to verify the quality of these estimates, the residuals are shown in Figure 6 comparing the estimated spectra with measured pure spectra. The similarity between estimated and measured real spectra is evident.

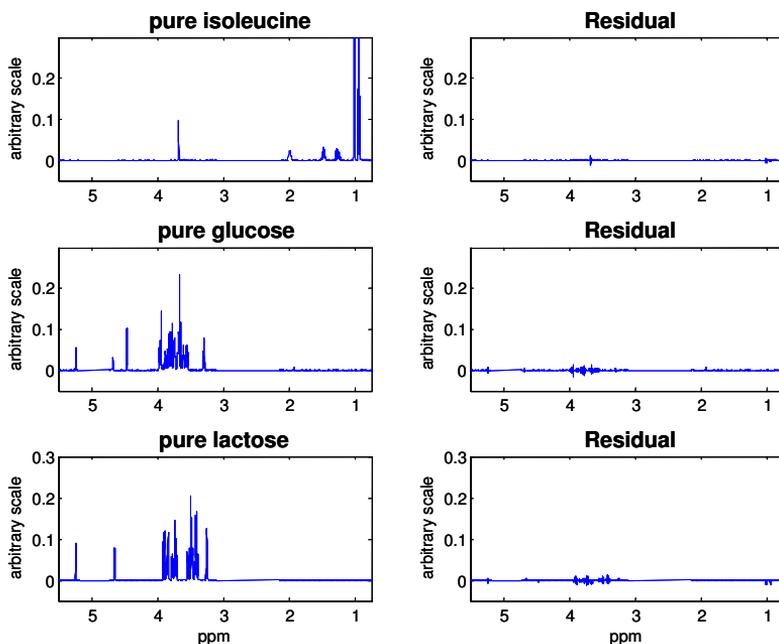


Figure 6 Pure spectra and residuals between pure measured spectra and PARAFAC estimates of spectra

3 CONCLUSION

The results presented show that diffusion-ordered NMR spectroscopy data fulfil the low-rank trilinearity requirement, since it is possible to extract the pure component spectra by application of a PARAFAC model. This is accomplished, despite the fact that the selected chemical analytes show a relatively high degree of overlapping NMR spectra and have relatively similar self-diffusion coefficients. Furthermore, it is accomplished with a limited number of samples, all including mixtures of the three chemical components. It is important to stress that the latter is not a requirement, in the sense that if one or more samples did not contain all three components, the analysis would still be handled equally well. These samples would simply show zero concentration for the missing component(s).

By measuring one pure component spectrum where the concentration of the component is known, it is possible to quantify this component in all the samples included in the PARAFAC model. This is accomplished without having to perform any kind of peak integration, which would be difficult in a case where no selective peak could be identified as originating from the chemical component in question. This is considered a tremendous improvement compared to the traditional way of working with NMR data of complex mixtures.

By the example presented here we hope to illustrate the enormous future possibilities for the *multi-way* analysis of complex 2D-NMR data with applications in the fields of metabonomics and quality control of food (bromatonomics).

Acknowledgments

The authors wish to thank The Danish Dairy Foundation, The Ministry of Food, Agriculture and Fisheries and the Danish Research Councils for generous support to our NMR project.

References

- 1 Baunsgaard D, Nørgaard L, Godshall MA, Specific screening for color precursors and colorants in beet and cane sugar liquors in relation to model colorants using spectrofluorometry evaluated by hplc and multiway data analysis, *J Agric Food Chem*, 2001, **49**, 1687-1694.
- 2 Booksh KS, Jiji RD, Andersson GG, Multi-way modeling of non-linear data: Practical application to excitation-emission matrix fluorometry of aqueous samples, *Abstracts of Papers of the American Chemical Society*, 1999, **217**, 065-ANYL
- 3 Bro R, Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications. Ph.D. thesis, University of Amsterdam (NL), <http://www.mli.kvl.dk/staff/foodtech/brothesis.pdf>, 1998.
- 4 Bro R, Kiers HAL, A new efficient method for determining the number of components in PARAFAC models, *J Chemom*, 2003, **17**, 274-286.
- 5 Dyrby M, Baunsgaard D, Bro R, Engelsen SB, Multiway chemometric analysis of the metabolic response to toxins monitored by NMR, *Chemom Intell Lab Syst*, 2004, in press.
- 6 Dyrby M, Peteresen M, Whittaker AD, Lambert L, Nørgaard L, Bro R, Engelsen SB, Analysis of lipoproteins using 2D diffusion-edited NMR spectroscopy and multi-way chemometrics, *Anal Chim Acta*, 2004, in press.
- 7 Eastman HT, Krzanowski WJ, Cross-validators Choice of the Number of Components From a Principal Component Analysis, *Technometrics*, 1982, **24**, 73-77.
- 8 Faber NM, Ferre J, Boque R, Kalivas JH, Second-order bilinear calibration: the effects of vectorising the data matrices of the calibration set, *Chemom Intell Lab Syst*, 2002, **63**, 107-116.
- 9 Gemperline P, Puxty G, Maeder M, Walker D, Tarczynski F, Bosserman M, Calibration-free estimates of batch process yields and detection of process upsets using in situ spectroscopic measurements and nonisothermal kinetic models: 4-(dimethylamino)pyridine-catalyzed esterification of butanol, *Anal Chem*, 2004, **76**, 2575-2582.
- 10 Gurden SP, Westerhuis JA, Bijlsma S, Smilde AK, Modelling of spectroscopic batch process data using grey models to incorporate external information, *J Chemom*, 2001, **15**, 101-121.

- 11 Harshman RA, Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis, *UCLA working papers in phonetics*, 1970, **16**, 1-84.
- 12 Harshman RA, De Sarbo WS, An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques, *Research Methods for Multimode Data Analysis*, (Eds. Law, HG, Snyder, CW, Hattie, JA, and McDonald, RP), Praeger Special Studies, New York, 1984, 602-642.
- 13 Hotelling H, Analysis of a complex of statistical variables into principal components, *Journal Of Educational Psychology*, 1933, **24**, 417-441.
- 14 Kruskal JB, Three-way arrays: rank and uniqueness of trilinear decompositions with applications to arithmetic complexity and statistics, *Linear Algebra and its Applications*, 1977, **18**, 95-138.
- 15 Laster L, Statistical background of methods of principal component analysis, *J Periodontol*, 1967, **38**, Suppl-66.
- 16 Nikolajsen RPH, Booksh KS, Hansen AM, Bro R, Quantifying catecholamines using multi-way kinetic modelling, *Anal Chim Acta*, 2003, **475**, 137-150.
- 17 Pearson K, On lines and planes of closest fit to points in space, *Philosophical Magazine*, 1901, **2**, 559-572.
- 18 Selli E, Zaccaria C, Sena F, Tomasi G, Bidoglio G, Application of multi-way models to the time-resolved fluorescence of polycyclic aromatic hydrocarbons mixtures in water, *Water Research*, 2004, **38**, 2269-2276.
- 19 Sidiropoulos ND, Bro R, On the Uniqueness of Multilinear Decomposition of N-way Arrays, *J Chemom*, 2000, **14**, 229-239.
- 20 Smilde AK, Bro R, Geladi P, *Multi-way analysis. Applications in the chemical sciences*. John Wiley & Sons Ltd., UK, 2004.
- 21 Vega-Montoto L, Wentzell PD, Maximum likelihood parallel factor analysis (MLPARAFAC), *J Chemom*, 2003, **17**, 237-253.
- 22 Wang YD, Borgen OS, Kowalski BR, Gu M, Turecek F, Advances in 2nd-order calibration, *J Chemom*, 1993, **7**, 117-130.

