

Computational analysis of oligosaccharide conformations

- methodological development, applied
studies, and design of glycomimetics

Francesco Strino

Department of Medical Biochemistry and Cell Biology
Institute of Biomedicine at Sahlgrenska Academy
University of Gothenburg



UNIVERSITY OF GOTHENBURG

Göteborg 2010

Cover illustration: Docking of the A/B histo-blood group antigen-specific domain of the β -galactosidase GH98CBM51 with a seleno-derivative of the histo-blood group A antigen. The solvent-accessible surface of the protein is colored according to lipophilicity (blue: hydrophilic, brown: lipophilic). The selenium atom is shown in green.

Computational analysis of oligosaccharide conformations

© Francesco Strino 2010

francesco.strino@biognos.se

ISBN 978-91-628-8127-6

URL: <http://hdl.handle.net/2077/22231>

Printed in Göteborg, Sweden 2010

Geson Sverige AB

It is not the possession of truth, but the success which attends the seeking after it, that enriches the seeker and brings happiness to him.

[Max Planck]

Computational analysis of oligosaccharide conformations

- methodological development, applied studies, and design of glycomimetics

Francesco Strino

Department of Medical Biochemistry and Cell Biology
Institute of Biomedicine at Sahlgrenska Academy
University of Gothenburg
Göteborg, Sweden

ABSTRACT

Carbohydrates are the most abundant class of biomolecules. Besides their roles as structural elements and energy storage, they are involved in signaling and recognition processes. Their functions and activities depend on their preferred conformations. The software GLYGAL was developed to perform conformational studies of oligosaccharides using a genetic algorithm tailored for carbohydrates. The new method was applied to the highly branched exopolysaccharide of *Burkholderia cepacia*. The results show that its heptasaccharide repeating units assume a well defined conformation, stabilized by steric interactions between consecutive units. Furthermore, GLYGAL was used to calculate favorable conformations of histo-blood group antigens. The compounds were then fitted in the binding site of the surface protein of the norovirus VA387 strain and their binding affinity was estimated by molecular dynamics and Glide scoring, giving insights into the interaction patterns involved in norovirus infection. Finally, the mimetic properties of thioglycosidic and selenoglycosidic derivatives of the ABH antigens were studied by conformational and dockings studies, indicating potentially bioactive derivatives with increased resistance to hydrolysis.

In conclusion, the computational methodologies developed during this study were successfully used, together with existing methods, for the investigation of natural carbohydrates and the rational design of glycomimetics.

Keywords: *Burkholderia cepacia*, *histo-blood group antigens*, *saccharide conformations*, *genetic algorithms*, *norovirus*, *thioglycoside*, *selenoglycoside*.

ISBN: 978-91-628-8127-6

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals:

- I. A. Nahmany, F. Strino, J. Rosén, G. J. L. Kemp and P.-G. Nyholm. The use of a genetic algorithm search for molecular mechanics (MM3)-based conformational analysis of oligosaccharides. *Carbohydr Res*, 2005. 340: 1059-64. Doi:10.1016/j.carres.2004.12.037.
- II. F. Strino, A. Nahmany, J. Rosén, G. J. L. Kemp, I. Sá-correia and P.-G. Nyholm. Conformation of the exopolysaccharide of *Burkholderia cepacia* predicted with molecular mechanics (MM3) using genetic algorithm search. *Carbohydr Res*, 2005. 340: 1019-24. Doi:10.1016/j.carres.2004.12.031.
- III. C.A.K. Koppisetty, W. Nasir, F. Strino, G.E. Rydell, G. Larson, P.-G. Nyholm. Computational studies on the interaction of ABO-active saccharides with the norovirus VA387 capsid protein can explain experimental binding data. *J Comput Aided Mol Des*, 2010. Doi:10.1007/s10822-010-9353-5.
- IV. F. Strino, J.-H. Lii, H.-J. Gabius and P.-G. Nyholm. Conformational analysis of thioglycoside derivatives of histo-blood group ABH antigens using an ab initio-derived reparameterization of MM4: implications for design of non-hydrolysable mimetics. *J Comput Aided Mol Des*, 2009. 23: 845-852. Doi:10.1007/s10822-009-9301-4.
- V. F. Strino, J.-H. Lii, C.A.K. Koppisetty, P.-G. Nyholm and H.-J. Gabius. Selenoglycosides *in silico*: ab initio-derived reparameterization of MM4, conformational analysis using histo-blood group ABH antigens and lectin docking as indication for bioactivity. *J Comput Aided Mol Des*, *under revision*.

CONTENTS

ABBREVIATIONS	iv
1 INTRODUCTION.....	1
1.1 Carbohydrate structures	1
1.1.1 Monosaccharides	1
1.1.2 Conformational properties of pyranoses	3
1.1.3 Glycosidic linkages	4
1.1.4 Methods for the study of the 3D structures of sugars.....	6
1.2 Biological and medical importance of sugars	7
1.2.1 Human histo-blood group antigens	7
1.2.2 Sugar derivatives	9
1.3 Genetic algorithms	10
1.3.1 Encoding.....	12
1.3.2 Genetic operators.....	12
1.3.3 Fitness function and selection	13
1.3.4 Variants of GAs.....	13
2 AIMS	16
3 METHODS	17
3.1 Quantum mechanics.....	17
3.1.1 Quantum mechanics theories.....	18
3.1.2 Basis sets	19
3.1.3 Common <i>ab initio</i> notations.....	20
3.2 Molecular mechanics and force fields.....	20
3.2.1 Force fields for sugars	22
3.3 Computational approaches for the study of oligosaccharide conformations.....	22
3.3.1 Adiabatic disaccharide maps.....	23
3.3.2 Filtered systematic searches.....	25
3.4 GLYGAL: Genetic algorithms for sugars (Paper I).....	26

3.4.1	Encoding.....	26
3.4.2	Genetic Algorithm operators	27
3.4.3	Fitness function and selection	28
3.4.4	Weights and adaptive parameterization.....	29
3.4.5	Other features of GLYGAL.....	29
3.5	Molecular Dynamics	30
3.6	Molecular docking	30
3.6.1	Glide.....	31
3.7	Other software.....	32
4	RESULTS AND DISCUSSION.....	33
4.1	Performance of GLYGAL (Paper I).....	33
4.2	Conformational studies on an exopolysaccharide of <i>Burkholderia cepacia</i> (Paper II).....	34
4.3	Interactions between the VA387 strain of norovirus and HBGAs (Paper III).....	36
4.4	Conformational studies on thioglycosides and selenoglycosides (Papers IV and V).....	38
5	CONCLUSIONS	44
6	FUTURE PERSPECTIVES.....	45
	ACKNOWLEDGEMENTS	46
	REFERENCES.....	47

ABBREVIATIONS

3D	Three-dimensional
B3LYP	Becke, three-parameter, Lee–Yang–Parr
EPS	Exopolysaccharide
Fuc	Fucose
GA	Genetic Algorithm
Gal	Galactose
GalNAc	N-acetylgalactosamine
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	N-acetylglucosamine
GSL	Glycosphingolipid
GLYGAL	GLYcosidic bonds Genetic ALgorithm
HBGA	Histo-Blood Group Antigen
Man	Mannose
MD	Molecular Dynamics
MP2	Second-order Møller–Plesset theory
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PDB	Protein Data Bank
Rha	Rhamnose

1 INTRODUCTION

Saccharides linked to proteins and lipids cover a large fraction of the surface area of most cells. Many of these saccharides are involved in specific recognition processes. To understand their biological function it is necessary to have information about their three-dimensional structure (Varki *et al.* 1999).

This thesis starts with an introduction of the structural properties of sugars (section 1.1) followed by a short introduction of their importance (section 1.2). Genetic Algorithms are then introduced (section 1.3) to give a background to their application to oligosaccharides described later (section 3.4). In the methods chapter (3), the computational approaches used in this work for the prediction of three-dimensional structures are described. The results from the five papers are summarized in chapter 4, after which some conclusions are drawn in chapter 5 and future directions are suggested in chapter 6.

1.1 Carbohydrate structures

Carbohydrates, also known as sugars, saccharides and glycans, are biopolymers made of components called monosaccharides. The monosaccharide units can be connected through glycosidic linkages and molecules composed of two monosaccharides linked together are called disaccharides. Saccharides of 3 to around 20 units are called oligosaccharides.

1.1.1 Monosaccharides

Monosaccharides are constituted by carbon chains of 3 to 9 carbons, with CHOH internal groups, a CH₂OH group at one end and an aldo ($-C[H]=O$) or keto ($-C[=O]-$, generally $-C[=O]CH_2OH$) group at the other end. The carbon atoms are numbered incrementally from the aldo or keto group (Figure 1a) and a monosaccharide with a chain of length n is called n -ose, where n is a Greek number, e.g. treose for 3 carbons and pentose for 5 carbons. The different chiralities of the internal CHOH group specify the differences between monosaccharides. The chiral orientation of the hydroxyl of the last chiral carbon define the L (left) and D (right) configuration. The relative chirality of the other CHOH groups defines different sugar types (Table 1). For example, the D orientation of the hydroxyl of C5 and the right-left-right

(= ≠ =) orientations of the hydroxyls of C2–C3–C4 specify a D-Glc (Table 1, Figure 1a).

Monosaccharides are generally not present as open chains, as they usually undergo intermolecular condensation between the double-bonded oxygen in the aldo or keto group and one of the CHOH groups. Monosaccharides with a 5-membered ring are called furanoses (*f*) and those with a 6-membered ring pyranoses (*p*). Similarly to the other nomenclature, the chirality assumed by the carbon (called anomeric) is called α if it matches the L/D chirality or β otherwise.

Amino, acidic, and carboxy derivatives of monosaccharides are commonly found in nature. Furthermore, hydroxyls often undergo deoxygenation, methylation, acetylation, phosphorylation or sulfation.

Table 1. Nomenclature of monosaccharides, ranging from treoses to hexoses. L/D indicates the L or D orientation of the monosaccharide. = indicates that the chirality is the same as the L/D carbon, ≠ indicates that the chirality is opposite.

	aldoses					ketoses			
	C2	C3	C4	C5	name	C3	C4	C5	name
Treose	L/D				glycer- aldehyde				dihydroxy- acetone
Tetrose	≠	L/D			threose	L/D			erythrose
	=	L/D			erythrose				
Pentose	≠	≠	L/D		lyxose	≠	L/D		xylulose
	=	≠	L/D		xylose				
	≠	=	L/D		arabinose				
	=	=	L/D		ribose				
Hexose	≠	≠	≠	L/D	talose	≠	≠	L/D	tagatose
	=	≠	≠	L/D	galactose				
	≠	=	≠	L/D	idose				
	=	=	≠	L/D	gulose				
	≠	≠	=	L/D	mannose				
	=	≠	=	L/D	glucose				
	≠	=	=	L/D	altrose				
	=	=	=	L/D	allose				

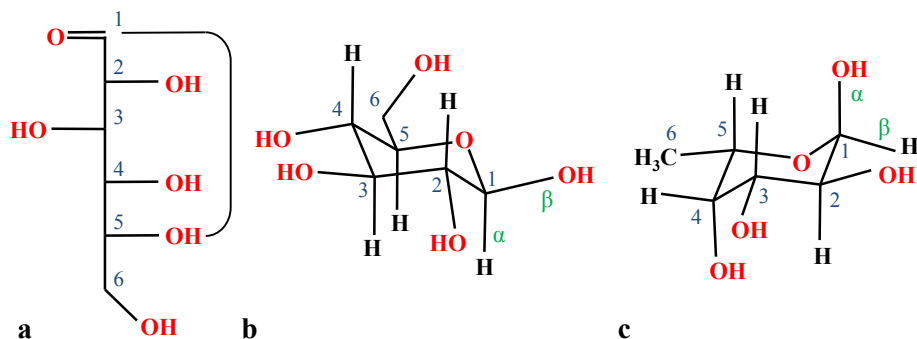


Figure 1. Monosaccharide structures: a) open chain D-Glc; b) 4C_1 conformation of β -D-Glcp; c) 1C_4 conformation of α -L-Fucp. The α and β positions of the anomeric carbon are shown in green.

1.1.2 Conformational properties of pyranoses

While the ring conformation of furanoses is rather flexible, pyranoses assume preferentially a chair conformation, where C5, O5, C2 and C3 form a plane and the relative positions of C4 and C1 define the 4C_1 and 1C_4 ring conformations. For example, β -D-Glcp assumes preferentially the 4C_1 conformation (Figure 1b), as most sugars do, while α -L-Fucp prefers the 1C_4 conformation (Figure 1c). The β position of the anomeric carbon is generally equatorial relative to the plane formed by C5, O5, C2 and C3, whereas the α position is axial (Figure 1).

The three staggered conformations about the C5-C6 bond are named after the gauche ($g \approx \pm 60^\circ$) or trans ($t \approx 180^\circ$) conformation assumed by the O5-C5-C6-O6 and C4-C5-C6-O6 torsion angles respectively, namely *gg*, *gt*, and *tg* (Figure 2b). One of these conformations is generally less favorable than the others due to unfavorable interactions of O4 and O6, for instance *tg* is unavailable for the β -D-Glcp (Hassel and Ottar 1947; Jeffrey 1990) as shown in Figure 2b.

The hydroxyl groups within a ring have a tendency to orient themselves in chains due to hydrogen bonds following a clockwise (C) or counterclockwise (R) orientation (Figure 2c). These conformations are generally energetically optimal, although a water solvent may favor other stable configurations involving hydrogen bonds with water or they may just be too ordered to be stable because of entropy reasons (Ha *et al.* 1988; Kräutler *et al.* 2007; Simons *et al.* 2009).

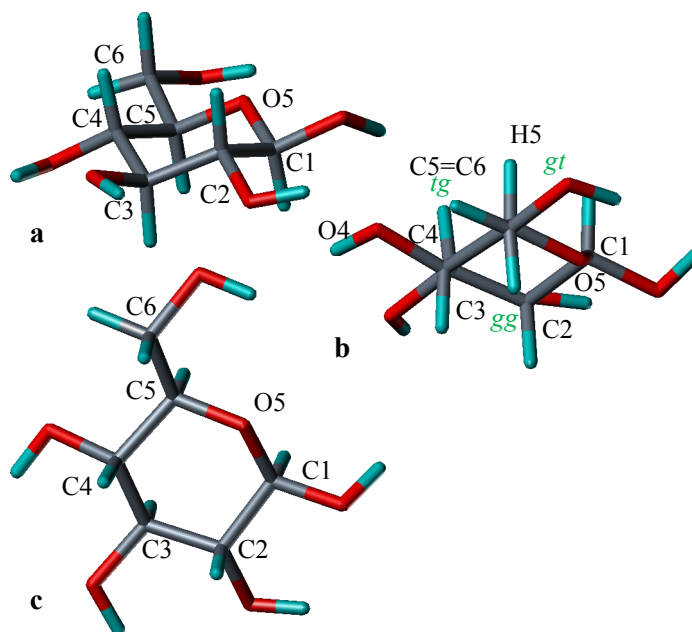


Figure 2. 3D structure of the minimum energy conformation of β -D-Glcp obtained with MM4: a) side view; b) view along the C5-C6 bond to show the staggered positions of the C5-C6 torsion (in green); c) view from above to show the counterclockwise arrangement of ring hydroxyls.

1.1.3 Glycosidic linkages

Glycosidic linkages are the result of a condensation (Figure 3) between the hydroxyl of the anomeric carbon of a monosaccharide and a hydroxyl of another saccharide. The oxygen atom of the glycosidic linkage is called glycosidic or bridge oxygen. Monosaccharides can be coupled to more than one monosaccharide and can give rise to branched structures.

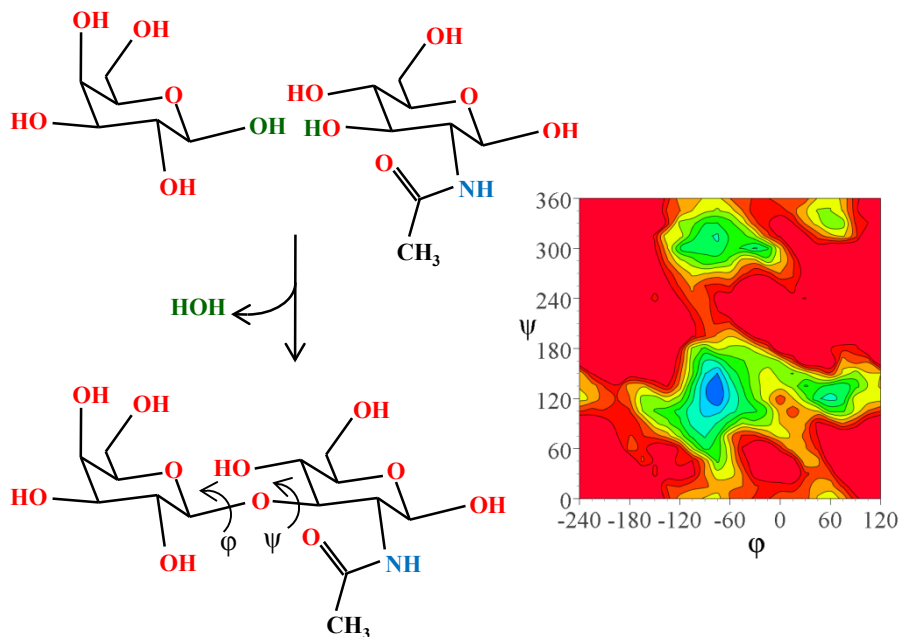


Figure 3. Left: formation of the disaccharide β -D-Galp-(1 \rightarrow 3)- β -D-GlcNAc from its monosaccharide components; Right: GLYGAL/MM4 adiabatic energy map of the glycosidic linkage. Contour levels are drawn at steps of 1 kcal/mol with high- to low-energy regions colored from red to blue gradually.

Glycosidic linkages are rather flexible and account for most of the flexibility of oligosaccharide structures. The conformation assumed by a glycosidic linkage can be expressed by means of the torsion angles assumed by its rotatable bonds. Most glycosidic linkages have two torsions, namely phi (ϕ) and psi (ψ), but other linkages may involve extra torsions, e.g. the (1 \rightarrow 6) linkage comprehends a third torsion angle called omega (ω). In the present work, torsions are measured using the *heavy atom* definition, which defines the angles of the (1 \rightarrow X) linkage as $\phi = \text{O5-C1-O-C'X}$ and $\psi = \text{C1-O-C'X-C'X+1}$. The three torsions of the (1 \rightarrow 6) linkage are defined as $\phi = \text{O5-C1-O-C'6}$, $\psi = \text{C1-O-C'6-C'5}$ and $\omega = \text{O-C'6-C'5-O'5}$.

The conformational properties of the glycosidic linkages can be studied through adiabatic energy maps (Figure 3), similar to Ramachandran plots used for proteins. The generation of adiabatic maps and the issues involved in conformational studies of oligosaccharide conformations are discussed in detail in section 3.3.

1.1.4 Methods for the study of the 3D structures of sugars

NMR and X-ray crystallography are the most common experimental techniques for determining the three-dimensional structures of biomolecules.

X-ray crystallography is an established technique for the determination of the 3D structure of molecules, provided that they can be obtained as well ordered crystals. Most crystallized carbohydrate structures are monosaccharides or disaccharides because of difficulties in crystallizing larger structures (Pérez and Mulloy 2005). On the other hand, many carbohydrates are crystallized in the presence of protein, providing important information about their bioactivity. X-ray structures represent one of the most important sources of sugar 3D structures, although the conformations of the crystallized structures may occasionally differ from the preferred conformations of the free sugars (Jeffrey 1990).

Nuclear magnetic resonance (NMR) is also routinely used to determine the structure of saccharides. Because of the flexibility of sugars, the quality of the data is generally not sufficient alone and Molecular Dynamics simulations are often performed to refine the predicted structures. Saturation Transfer Difference (STD) NMR can also be used to study the binding interactions of sugar ligands.

Since both X-ray and NMR are relatively expensive and time consuming, an alternative, or complementary, approach for oligosaccharide conformational analysis is to computationally search the space of possible conformations to identify favorable low energy conformations. A detailed description of computational methods for the study of carbohydrate conformations is given in chapter 3.

Online resources for carbohydrate 3D structures

The largest collection of experimentally determined 3D structures is the Cambridge Structure Database (<http://www.ccdc.cam.ac.uk/products/csd/>), which contains over 4000 carbohydrate structures. GLYCO3D (<http://www.cermav.cnrs.fr/glyco3d/>) stores many sugar structures. Several entries in the Protein Data Bank (<http://www.pdb.org/>) contain carbohydrates, either as ligands or as glycoproteins. It is important to keep in mind that sugars in the presence of proteins may undergo conformational changes (French *et al.* 2001; Johnson *et al.* 2009) and that in many cases carbohydrate structures in the PDB are simply wrong (Lütteke *et al.* 2004; Lütteke and von der Lieth 2004). Oligosaccharide conformations predicted

with SHAPE (Rosén *et al.* 2009) are stored in the 3D-BAO database (<http://www.cermav.cnrs.fr/cgi-bin/bao/3D-BAO.cgi>).

1.2 Biological and medical importance of sugars

Bacteria often synthesize and secrete large amount of saccharides also known as exopolysaccharides (EPS). The cell surface of Gram-negative bacteria is covered with saccharides for protective and mimetic purposes.

Glycans are often coupled to other biomolecules, in particular to proteins and lipids. Many proteins are glycosylated, and sugars are covalently linked to oxygen atoms of serine, threonine, tyrosine, hydroxylysine or hydroxyproline (O-linked), or to the nitrogen atoms of asparagine or arginine side chains (N-linked). Glycolipids are composed of carbohydrates O-linked to lipids, e.g. glycosphingolipids (GSL) are linked to ceramide.

Carbohydrate epitopes are recognized by bacteria (Hanada 2005) and viruses (Marsh and Helenius 2006). Furthermore, cancer cells express typical carbohydrate structures. Several carbohydrates and carbohydrate derivatives are drug candidates or marketed drugs (Ernst and Magnani 2009; Osborn and Turkson 2009).

Epitopes of bacterial saccharides can be used for vaccine development (Astronomo and Burton 2010; Hecht *et al.* 2009). In order to increase the efficiency of the vaccines they are generally conjugated to proteins, thereby eliciting a lasting immune response. There are successful vaccines for *Haemophilus influenzae* type b (Verez-Bencomo *et al.* 2004), *Neisseria meningitidis* (Giebink *et al.* 1993), and *Streptococcus pneumoniae* (Whitney *et al.* 2003), and there is ongoing development for vaccines towards HIV (Wang 2006) and some types of cancer (Guo and Wang 2009).

1.2.1 Human histo- blood group antigens

The ABH blood group system has been studied for more than a century in transfusion medicine (Landsteiner 1900), where incompatibilities led to severe complications. The carbohydrate basis of the antigens involved was discovered half a century later (Morgan 1950). These antigens are also

present in other tissues and are thus referred to as histo-blood groups antigens (HBGAs) (Clausen and Hakomori 1989).

HBGAs define two important systems: the ABH (or ABO) system and the closely related Lewis system. Both systems can be present in the same sequence as they overlap both structurally and synthetically. The structure composition and the biosynthesis paths are shown in Figure 4.

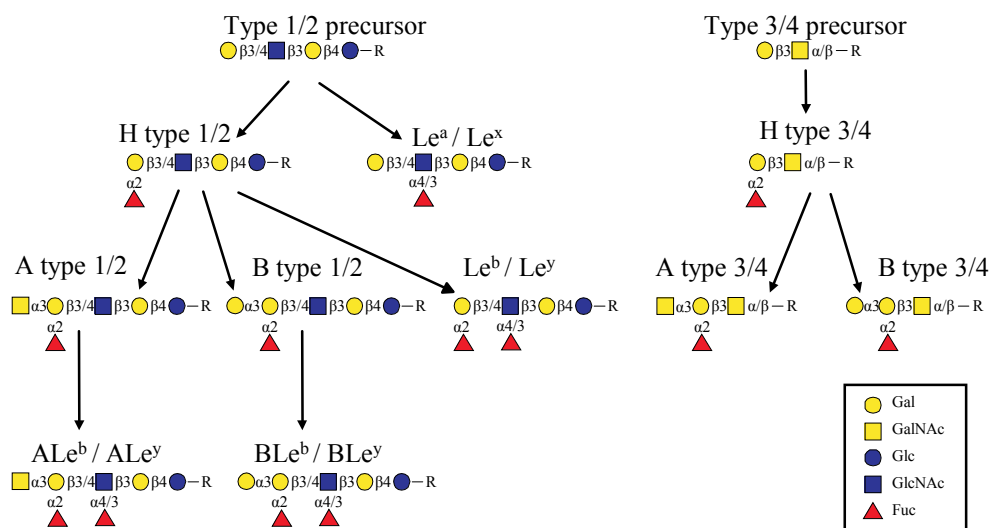


Figure 4. Biosynthetic pathways of the ABH and Lewis antigens on the chain types 1/2 (left) and 3/4 (right). Antigen names and linkages of the different chains are, when differing between the chains in each figure, separated by a slash in the antigen notation (Rydell 2009).

These antigens are found on four different carbohydrate chains: type 1 (Gal β 1-3GlcNAc β 1-R) and type 2 (Gal β 1-4GlcNAc β 1-R) chains are found in N- and O- glycoproteins and in glycosphingolipids (GSLs). Type 3 structures (Gal β 1-3GalNAc α 1-R) are found in O-glycoproteins and in GSLs. Type 4 structures (Gal β 1-3GalNAc β 1-R) are present in GSLs (Ravn and Dabelsteen 2000).

The ABH system is defined by the presence of an α 1,2-linked fucose connected to the Gal β moiety. H structures have no other monosaccharide connected to the galactose, while if a GalNAc α or a Gal α is connected in position 3, the structure becomes A or B, respectively.

Compounds of the Lewis system contain either type 1 or type 2 chains where the central GlcNAc has a Fuc α attached in position 4 (Le^a) or 3 (Le^x), respectively. If the structure is also part of the ABH system, they are denoted Le^b or Le^y instead. Lewis structures containing A- or B- trisaccharide moieties are denoted ALe^b, BLe^b, ALe^y, and BLe^y.

Apart from their role in transfusion and transplantation medicine, histo-blood group epitopes are known to be involved in other interactions of biological and medical relevance. Bacteria and viruses have been shown to bind to HBGAs (Azevedo *et al.* 2008; Le Pendu 2004). Furthermore, normal and tumor cells interact with such antigens and in lung cancer this interaction has prognostic relevance (Kayser *et al.* 1994). A complete conformational study of the ABH and Lewis systems has been performed using CICADA and the MM3 force field (Imberty *et al.* 1995). Furthermore, the 3D structures of some HBGAs have been determined by experimental methods (Azurmendi and Bush 2002; Pérez *et al.* 1996).

1.2.2 Sugar derivatives

While small differences such as the N-acetylation in HBGAs of a monosaccharide can destroy the binding properties (Galanina *et al.* 1997; Teneberg *et al.* 2003), lectins often recognize several saccharide residues. Information about the binding site of the protein can be used for a rational design of derivatives with similar or superior binding affinity. Sugar derivatives obtained by the substitution of the glycosidic oxygens also yield interest as potential drug leads because of their increased resistance against hydrolysis. The present work concentrates on substitutions of the glycosidic linkages, although a similar approach can be used for modeling other kinds of derivatives.

C-glycosides occur in nature (Haynes 1965). They have enhanced flexibility and can assume multiple conformations that would not be energetically favorable in the natural O-glycosides (Asensio *et al.* 1999; Jiménez-Barbero *et al.* 2000; Pérez-Castells *et al.* 2007; Poveda *et al.* 2000), while still retaining biological activity (Asensio *et al.* 1999; Pérez-Castells *et al.* 2007). N-glycosidic linkages are present at the attachment point of N-glycoproteins and have been thoroughly studied (Ali *et al.* 2009; Petrescu *et al.* 1999), while the more complex –N[OMe]– derivative has also showed conformational properties similar to the natural saccharide (Peri *et al.* 2004). Glycosyldisulfides have been synthesized (André *et al.* 2006; Rye and Withers 2004; Szilágyi and Varela 2006; Witczak 1999) and have

experimentally shown mimetic properties and even improved binding (André *et al.* 2006). The selenosulfo-glycosidic linkage has been shown to have similar properties to its corresponding glycosyldisulfide and to be more flexible than the natural disaccharide (Chakka *et al.* 2005).

Paper IV and V focus on the study of S-glycosides and Se-glycosides.

Thioglycosides

Thioglycosides can be synthesized either chemically (Rye and Withers 2004; Szilágyi and Varela 2006; Witczak 1999) or enzymatically with mutated glycosidases (Jahn *et al.* 2003; Kim *et al.* 2006; Müllegger *et al.* 2005). Only a few studies on conformational properties of thioglycosides (Aguilera *et al.* 1998; Mazeau and Tvaroška 1992; Saito and Okazaki 2009; Tvaroška 1984) have been reported in the literature, showing increased flexibility and accessibility of secondary conformers.

Selenoglycosides

Selenoglycosides are mostly used as intermediates for carbohydrate synthesis (Chakka *et al.* 2005; Witczak and Czernecki 1998). In crystallography applications, the selenium derivatization of oligonucleotides is used to solve the phase problem (Sheng and Huang 2008). In this context, oxygen to selenium substitutions of a carbohydrate ligand can be used instead of the selenomethionine substitution of the protein as shown in the case of the O→Se substituted GlcNAc β 1-Se-CH₃, which was co-crystallized in complex with an *Escherichia coli* adhesin (Buts *et al.* 2003). Seleno-monosaccharides are present in humans apparently as urinary metabolites for excretion of excess selenium (Gammelgaard *et al.* 2005; Kobayashi *et al.* 2002; Ogra *et al.* 2002).

1.3 Genetic algorithms

Genetic Algorithms (GAs) are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetics. These ideas were first introduced in the 1960s by Rechenberg in his work on Evolutionary Strategies (Rechenberg 1965). GAs were introduced by John Holland (1975).

The scheme of Holland (Figure 5) can characterize mathematically the evolution over time of the population within a GA. The search for an appropriate solution begins with an initial population of solutions (individuals). Individuals of the current population give rise to the next generation by means of operations such as random mutation and crossover, which are patterned after processes in biological evolution. At each step the

individuals are evaluated using a given measure of fitness, with the most fitted individuals selected probabilistically as seeds for producing the next generation (Mitchell 1997). The process iterates until the convergence criteria are met.

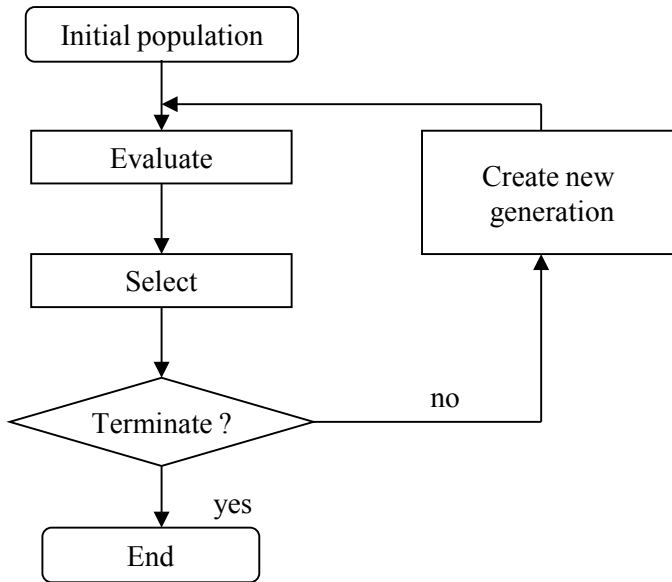


Figure 5. Flowchart of a genetic algorithm.

GAs have been applied successfully to a variety of learning tasks and other optimization problems in many areas such as bioinformatics, phylogenetics, computational science, engineering, economics, chemistry, manufacturing, mathematics, and physics. Furthermore, there are successful applications in molecular modeling with the docking programs AutoDock (Morris *et al.* 2009) and GOLD (Jones *et al.* 1997). The popularity of GAs can be motivated by the following reasons:

- a) Evolution is known to be a successful method for adaptation within biological systems.
- b) GAs can search very complex search spaces.
- c) GAs do not require deep knowledge of the system being examined.
- d) GAs are relatively easy to code and parallelize.
- e) GA parameters and techniques can be ported easily across different systems.

The idea behind GA search methods is to mimic the biological processes known from genetics and evolution. The chromosomes existing in all living cells define the genotype of the organism. Each chromosome consists of genes that encode proteins, which in turn affect the appearance and the behavior of the cells. The ability to survive in the environment defines the fitness of an organism, and more fitted individuals have higher probability of reproducing and generating new offspring.

During sexual reproduction, the recombination (crossover) process allows chromosomes from the parents to merge and create a new chromosome of the offspring. In both sexual and asexual reproduction, changes (mutations) in one or more genes can occur and they will be transmitted to the offspring.

In order to apply GAs to solve a given problem, it is only necessary to have a mapping function that can encode a candidate solution into its genomic representation and a fitness function to evaluate its quality.

1.3.1 Encoding

In the original GA and in many applications the favorable encoding is a binary encoding where a chromosome is represented as a bit string. This representation allows a straightforward application of operators such as mutation and crossover on the chromosomes. Another encoding strategy which can be seen as more natural for some applications uses a so-called value encoding. In this case the chromosomes consist of an array of values. These values can be as simple as integers, real numbers, or characters, but they might even be electronic circuit components (Rudnick *et al.* 1994) or notes in jazz solos (Biles 1994). Specialized encoding methods have the advantage of modeling specific properties of a system, but they might require special crossover and mutation operators to be designed.

1.3.2 Genetic operators

The generation of offspring in a GA is determined by a set of operators that manipulate selected members of the current population. The most common ones are mutation and crossover.

Mutation

The mutation operator works on a single parent and produces a single offspring. If a binary representation is used, this operator chooses a single bit at random and changes its value. For example, given the parent represented by the bit string 1000110011 and by choosing randomly the fourth bit from the left to be changed, the offspring 1001110011 is generated.

Crossover

The crossover operator produces two offspring from two parents, by exchanging selected bits between the parents. Figure 6 illustrates the single-point crossover operator. The crossover operator can be extended in different ways, for example by choosing more than one crossover point, by choosing more than two parents, or by generating a different number of offspring.

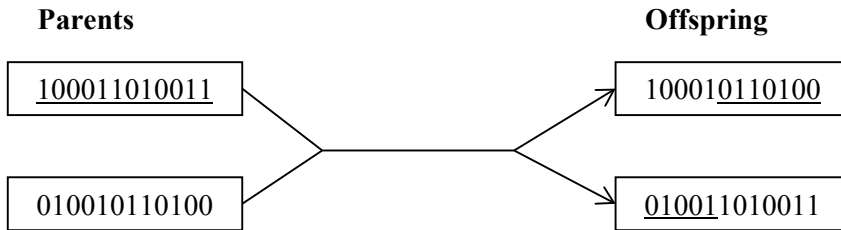


Figure 6. Single-point crossover. The parents are shown on the left and the offspring on the right. The underlined parts are the bits inherited from the first parent.

1.3.3 Fitness function and selection

The fitness function is used to rank individuals and to select them probabilistically for inclusion into the next generation.

The standard GA uses a method called “roulette wheel selection” in which the probability that an individual will be selected corresponds to a normalized function of its fitness. Other selection methods have been suggested such as “rank selection”, “tournament selection” and “steady-state selection”.

1.3.4 Variants of GAs

The standard GA described above can be extended or changed to create slightly different genetic algorithms. Only the variants that have been implemented in GLYGAL (see section 3.4) are described here.

Parallel GA

In the parallel GA several populations evolve in parallel with individuals migrating between the different populations. This simulation resembles real life evolutionary processes. Using several different populations one can avoid the so-called “crowding problem” where an individual that is highly fitted quickly reproduces and takes over a large fraction of the population. Isolation can be a requisite for evolution, as pointed out by Darwin in his report about the evolution of finches in the Galapagos Islands (Darwin 1859). However,

by allowing some migration between the different populations, successful genes can spread to other populations.

The algorithm itself changes only slightly with respect to the standard GA. For the most part, each population runs independently using the standard GA search. The only difference is that some individuals occasionally migrate to other populations. This migration can be done in different ways, and the most common way is that a fixed number of best individuals from each population moves to another population every fixed number of generations. Another advantage of parallel GA is the possibility of distributing the execution by running different populations on different machines in a cluster, thereby reducing the need for shared memory communication between the nodes.

Lamarckian GA

Most of the GA methods simulate Darwinian evolution and Mendelian genetics where an individual is evolving from one generation to the next. Another model of evolution, Lamarckian evolution, is based on the theories proposed by Lamarck (1809). His theory suggests that experiences of a single individual could directly affect the genetic material of its offspring. One example is the giraffe neck. According to Lamarck, since the giraffe had to stretch its neck to reach the higher branches for the leaves, its neck got longer and this feature was passed on to its offspring.

Although the Lamarckian theory is not accepted as a valid model of biological evolution, it has been shown to improve the effectiveness of computerized GAs in some cases (Morris *et al.* 1998). Figure 7 illustrates the difference between the Lamarckian and the standard Darwinian approach. The Lamarckian GA has only one basic difference from the standard GA: the characteristics of an individual after local minimization and evaluation are directly inherited by the offspring.

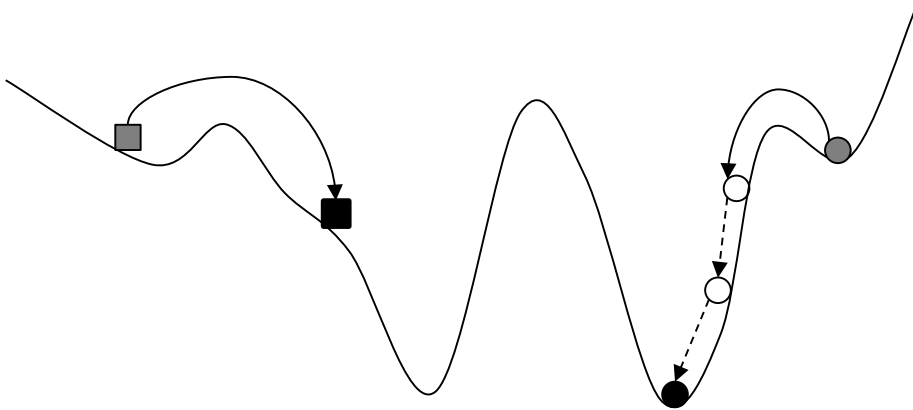


Figure 7. The difference between Darwinian (squares) and Lamarckian (circles) GA approaches. Parents are shown in gray, offspring in black. The generation of the new offspring is represented by continuous arrows. In the Lamarckian approach, optimization steps (dashed arrows) are performed until a local optimum is reached.

2 AIMS

The aim of the present thesis work was to develop and use computational methods for the study of medically relevant carbohydrate structures, as well as predicting properties of sugar derivatives.

The main goals were:

- a) Development of a genetic algorithm based computational tool to predict oligosaccharide three-dimensional structures efficiently and accurately.
- b) Conformational investigation of a highly branched exopolysaccharide of *Burkholderia cepacia* in order to identify epitopes for potential glycoconjugate vaccines.
- c) Conformational studies of the binding of histo-blood group antigens with the VA387 strain of norovirus to elucidate the binding patterns and explain binding data.
- d) Rational design of glycomimetics resistant to hydrolysis using O→S and O→Se substitutions of the glycosidic linkages.

3 METHODS

In this chapter, the methods used for conformational studies of oligosaccharide structures are discussed. After introducing *ab initio* calculations (section 3.1) which can be used to investigate in detail the properties of molecular fragments, the force fields to evaluate the energy of saccharides are discussed (section 3.2). Furthermore, methods to sample the vast conformational space are explained: systematic search (section 3.3), genetic algorithms (section 3.4) and molecular dynamics (section 3.5). Finally, the docking programs to study protein-carbohydrate interactions are described in section 3.6.

3.1 Quantum mechanics

In computational chemistry, *ab initio* methods comprehend the methods based solely on the principles of quantum chemistry. A complete description of quantum mechanics is beyond of the scope of this work, so only a brief introduction to the methods used in the rest of the thesis is given here. More comprehensive introductions have been given, e.g. by Davidson and Feller (1986), Henre (2003), and Pilar (2001). Several software packages for *ab initio* calculations exist, the most common are Gaussian (Frisch *et al.* 2009), GAMESS (Gordon and Schmidt 2005), Spartan (Hehre and Huang 1995) and Jaguar (Schrödinger Inc., Portland, OR, USA).

The quantum mechanics approach defines the system as a collection of nuclei and electrons and the interactions are defined only by the Schrödinger equation, which in its simplest form is

$$\hat{H}\psi = E\psi$$

where \hat{H} is the Hamiltonian operator (which can be seen as the sum of potential and kinetic energy), E is the electron energy and ψ is a function of the electronic coordinates describing their movements.

The Schrödinger equation can be solved exactly only for the one-electron hydrogen atom. The resulting wave functions are the known s, p, d, f, and g atomic orbitals (Figure 8). The integral of the square of the amplitude of the wave function over a certain volume defines the probability of the electron of being in that volume and is referred to as electron density.

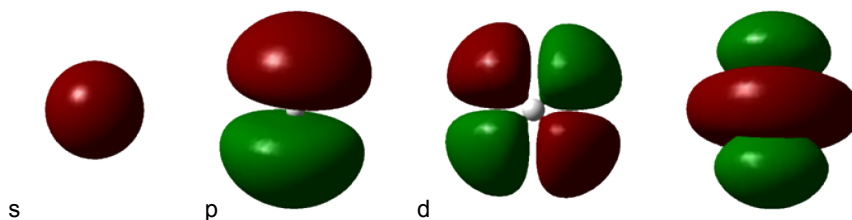


Figure 8. Examples of *s*, *p*, and *d* orbitals (red and green indicate opposite polarity).

Since all chemical systems more complex than the one-electron hydrogen atom cannot be solved exactly, many simplifications and models have been developed to make the problem tractable.

3.1.1 Quantum mechanics theories

The first simplification is assuming that nuclei do not move (Born–Oppenheimer approximation), which is well justified by the fact that electrons have a smaller mass and move close to the speed of light. The addition of other assumptions generates different theoretical models.

A commonly used simplification is the Hartree–Fock (HF) theory (Fock 1930), in which all electrons are assumed to move independently of each other, thus dividing the problem into a group of independent one-electron sub-problems. The Møller–Plesset (MP) perturbation theory (Møller and Plesset 1934) is commonly used to correct the errors introduced by the HF simplification. Perturbation theory is a mathematical way to solve a complex mathematical problem (Schrödinger equation) by first solving a simple related problem (HF approximation) and then iteratively (the number of steps is the order of the model) adding correction terms (electron–electron interaction terms) to solve the original problem. The second-order model (MP2) is widely used, but third- (MP3) and fourth-order (MP4) models can also be used when higher precision is desired.

Density functional theory (DFT) is an alternative theory where the energy is estimated using functionals (i.e. functions of functions) of the electron density instead of the wave functions. This makes the DFT approach computationally fast. Because of difficulties in modeling some intermolecular interactions accurately, hybrid approaches that borrow some terms from HF or MP theories have been developed. The B3LYP (Becke, three-parameter,

Lee–Yang–Parr) theory (Becke 1993; Kim and Jordan 1994; Lee *et al.* 1988) is one of the most commonly used DFT approaches.

3.1.2 Basis sets

It is generally assumed that the one-electron solutions for the hydrogen atom will resemble those for multi-electron atoms. Such solutions are commonly represented as a linear combination of atomic orbitals (the LCAO approximation):

$$\psi_i = \sum_{\mu} (c_{\mu i} \varphi_{\mu})$$

where ψ_i is the solution for electron i , the φ_{μ} are the functions contained in the basis set and $c_{\mu i}$ are the molecular orbital coefficients that describe the solution. This simplification turns differential equations into algebraic equations, which are much simpler to solve.

There are several types of basis sets, which correspond to different requirements in the tradeoff between accuracy and computational complexity. For computational reasons, linear combinations of Gaussian functions are generally used instead of the Slater-type orbitals (STOs) that arise as solutions to the hydrogen atom; e.g. the common STO-nG family uses n Gaussian functions to represent each STO.

Split-valence basis sets

In order to model the atoms that do not have fully spherical properties, split-valence basis sets include different models for core and valence orbitals and the valence orbitals are split into parts. They are generally named using the formula $N_C-N_{O1}\dots N_{Ok}G$, where N_C is the number of Gaussian functions used to model the core orbitals, k is the number of parts in which the valence orbitals are split and each part is modeled by N_{O_i} Gaussian functions. The most common examples are 3-21G, 6-31G and 6-311G.

The basis sets described so far are centered on the nuclei and thus cannot describe the displacements of the electron distribution (polarization). These effects can be modeled by the addition of d-type orbitals functions on main-group elements (i.e. with valence orbitals of s- and p-type) and with the addition of p-type orbitals on hydrogen (which only have s-type orbitals). Basis sets modeling only main-group atoms are labeled *, while basis sets that also model hydrogens are labeled **. Polarization functions can also be labeled more explicitly by indicating the number and type of orbitals used

for main-group and hydrogen atoms, respectively. For example, ** is often implemented as two d orbitals for main-group elements and two p orbitals for hydrogens, i.e. (2d, 2p).

Diffuse functions are often included in basis sets to model electrons that are loosely associated with specific atoms, e.g. in anions. Similarly to polarization, + indicates the presence of diffuse functions for main-group elements and ++ indicates also the presence of diffuse functions for hydrogen atoms.

3.1.3 Common *ab initio* notations

When describing *ab initio* calculations, the notation “Theory/Basis set” is commonly used: for example, MP2/6-311+G** indicates computation using the second order Møller–Plesset theory and a split-valence basis set 6-311G with polarization for all atoms and diffusion for atoms except hydrogen. If the calculation is preceded by geometry optimization with a different theory or basis set, the notation “Theory for energy/Basis set for energy//[theory for geometry]/[basis set for geometry]” is used instead.

3.2 Molecular mechanics and force fields

Molecular mechanics describes molecules as a set of bonded atoms whose interactions can be modeled using standard Newton mechanics instead of quantum mechanics. The interactions between atoms are modeled with simple parameterized functions based on experimental values or *ab initio* calculations. These functions define the so-called force field, which is used to easily calculate the potential energy of the system.

Rather than a physical model, molecular mechanics is an elaborate interpolation strategy. It is based on the assumption that the geometrical properties of atoms can be transferred from one molecule to similar ones, in particular from small fragments to larger molecules.

The most important bonded interactions concern bond lengths, bond angles and torsion angles. Bond lengths are modeled with a parabola parameterized by the equilibrium distance and a strain constant that penalizes stretched and compressed values. This model is physically equivalent to a spring. The strain of bond angles is defined similarly. The energy profile of torsion angles is more complex and is typically modeled with a Fourier series, generally a series of cosines. For example, the three-fold symmetry of the staggered conformations of C₂H₆ can be easily modeled by a third-order term

$k_3 \cos(3\omega)$ Other bonded terms such as out-of-plane bending, anomeric corrections, as well as second order terms (e.g. bend-bend, stretch-bend, torsion-bend) are often defined. Some force fields also use higher order Taylor or Fourier series to model some interaction terms more accurately.

The most important non-bonded interactions are the electrostatic and the van der Waals interactions. Electrostatic interactions are usually modeled by the Coulomb law, while van der Waals interactions are generally modeled as a Lennard–Jones potential consisting of a sum of a repulsive and an attractive term. Other common non-bonded interaction terms model hydrogen-bond and solvation effects.

The energy function of a simple force field can be expressed as:

$$\begin{aligned}
 E &= \sum_B^{\text{bonds}} k_B (r - r_B)^2 + \sum_A^{\text{angles}} \alpha_A (\alpha - \alpha_A)^2 \\
 &+ \sum_T^{\text{torsions}} [k_{1T} \cos(\omega) + k_{2T} \cos(2\omega) + k_{3T} \cos(3\omega)] \\
 &+ \sum_A^{\text{atoms}} \sum_B^{\text{atoms}} \frac{q_A q_B}{\epsilon r} + \sum_A^{\text{atoms}} \sum_B^{\text{atoms}} \left(\frac{k_{1AB}}{r^{12}} - \frac{k_{2AB}}{r^6} \right)
 \end{aligned}$$

where constants for specific atom types have been previously parameterized from experimental methods or *ab initio* calculations.

The simple definition of the force fields makes them very useful for a rapid estimation of the molecular energy of molecules with over one hundred atoms. If some distance threshold is applied, energy evaluations scale linearly with the number of atoms of the studied molecule. The computational speed is several orders of magnitude higher than *ab initio* methods and the accuracy is similar whenever parameters with good quality are available. Furthermore, the simple formulation of the force field makes the calculation of gradient and Hessian relatively easy, which is very useful for optimization applications.

Hybrid QM/MM (quantum mechanics / molecular mechanics) approaches can also be used to combine QM accuracy with MM speed, especially in cases where force field parameters are missing or where some interactions are difficult to model with force fields, e.g. interactions with metal ions.

3.2.1 Force fields for sugars

Computational analysis on saccharide conformations became established in the late 1970s. Early methods such as the HSEA, Hard-Sphere Exo-Anomeric (Thøgersen *et al.* 1982) considered monosaccharide as rigid residues and could yield satisfactory results in most cases by providing account for the exo-anomeric effect (i.e. the predilection of the glycosidic ϕ torsion for values around $\pm 60^\circ$).

Nowadays the most common popular general-purpose force fields offer parameterizations for carbohydrates which are accurate and show comparable performances (Hemmingsen *et al.* 2004; Pérez *et al.* 1998; Stortz *et al.* 2009). The most commonly used are the CSFF parameter set (Michelle *et al.* 2002) for CHARMM, the GLYCAM06 parameter set (Kirschner *et al.* 2008) for AMBER, GROMOS (Lins and Hünenberger 2005), MM3 (Allinger *et al.* 1989; Lii and Allinger 1989a; Lii and Allinger 1989b), its successor MM4 (Allinger *et al.* 2003; Langley and Allinger 2002; Lii *et al.* 2003a; Lii *et al.* 2004; Lii *et al.* 2003b; Lii *et al.* 2003c; Lii *et al.* 2005), and OPLS 2005 (Kaminski *et al.* 2001).

As an alternative to standard force fields, the hybrid quantum mechanics approach PM3CARB-1 (McNamara *et al.* 2004) has also shown promising results (Stortz *et al.* 2009). Furthermore, coarse-grained force fields have been developed or parameterized for carbohydrates (Bathe *et al.* 2005; López *et al.* 2009; Molinero and Goddard 2004). Coarse-grain force fields consider monosaccharides as single units and parameterize the interactions between them, which results in a considerable gain in speed and allows the investigation of large polysaccharides.

3.3 Computational approaches for the study of oligosaccharide conformations

Force fields give the possibility of evaluating the energy of a given conformation. However, molecules have flexibility and assume several conformations with a probability that follows the Boltzmann distribution. In the discrete formulation, the probability that a molecule assumes a conformation i (out of n possible conformations) is given by:

$$P_i = \frac{e^{-\frac{E_i}{RT}}}{\sum_{j=1}^n e^{-\frac{E_j}{RT}}}$$

where R is the molar ideal gas constant (i.e. 8.314472 J/(mol K)) and T is the temperature in degree Kelvin. It is thus very important to explore the conformational space of a molecule in order to know the conformations assumed by the compound. The exponential nature of the Boltzmann distribution ensures that only the conformations within a few kcal/mol from the global minimum are populated in relevant amounts. For example, at room temperature the probability ratios of two conformations differing 1, 2, and 3 kcal/mol quickly drop to around 1/5, 1/30 and 1/160, respectively.

It is often useful to cluster similar conformations. In this case, the probability for each conformer C_i can be formulated as:

$$P_{C_i} = \frac{\int_{C_i} e^{-\frac{E_i}{RT}} dC_i}{\sum_{j=1}^n \int_{C_j} e^{-\frac{E_j}{RT}} dC_j}$$

In most molecules, including saccharides, the conformational differences are mainly due to the different states assumed by the torsion angles. For example, the conformational space of monosaccharides (see 1.1.2) can be reduced to a few combinations of the staggered configurations assumed by the C5-C6 torsion and possibly the ring conformations of the hydroxyls (Stortz 1999).

3.3.1 Adiabatic disaccharide maps

The conformational variation of a saccharide depends mostly on the conformation assumed by its glycosidic linkages, as even small variations of their torsion angles can produce large effects in the geometrical properties.

The conformational properties of a glycosidic linkage are often summarized by adiabatic 2D ϕ/ψ maps (Figure 9), although plotting ψ can often be enough (Stortz and Cerezo 2002). In this context, adiabatic means that for each value of ϕ/ψ the energy values of the combinations of monosaccharide conformations are considered and that during each evaluation the structures are relaxed by local energy minimizations. The common way to generate such maps is to sample a grid of ϕ/ψ pairs with a step size of 5° to 30°. For each pair, several monosaccharide conformations are considered and energy minimizations are performed at each step while constraining the ϕ and ψ torsions. The lowest energy values for each ϕ/ψ pair is then saved and used to draw the maps (French *et al.* 1990; Stortz 1999).

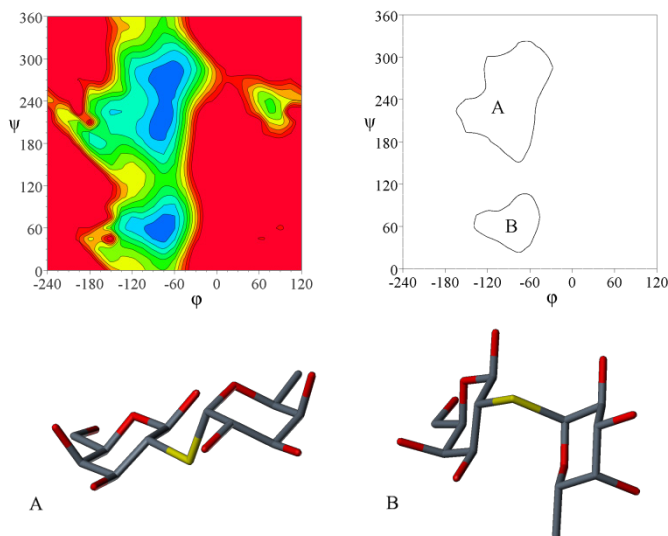


Figure 9. Conformational analysis of the thio-disaccharide α -L-Fucp-(1-S \rightarrow 2)- β -D-Glcp using GLYGAL/MM4R (see Paper IV). Adiabatic energy map, population contours and the minimum energy conformations from the A and B populations are shown. In the adiabatic map the contour levels are drawn at steps of 1 kcal/mol with high- to low-energy regions colored from red to blue gradually.

Table 2. Description of the energetically favorable conformations of the thio-disaccharide α -L-Fucp-(1-S \rightarrow 2)- β -D-Glcp (Paper IV).

Conformer	ϕ/ψ	Energy (kcal/mol)	Population
A	-90°/-90°		62.4%
B	90°/45°	+0.03	35.1%

The data can be analyzed in several ways. GLYGAL generates 2D depictions of the energy landscape (Figure 9) by drawing the energy isocontours after identifying some points on grid lines using the marching square algorithm and interpolating these points using cardinal splines (i.e. a segmented curve made of cubic Bezier segments which have the same direction in the connection points).

Another common analysis is clustering populations. In Papers IV and V, neighboring conformations within 3 kcal/mol from the global minimum were clustered together using the union-find algorithm, which also identifies the optimal conformation in each cluster. The relative population for each

conformer was then calculated by summing the Boltzmann probabilities of the conformations within the population. Finally, conformers representing less than 1% of the population or with energy at least 2 kcal/mol higher than the global minimum were excluded.

The conformational analysis for the thioglycosidic derivative of the histo-blood group H antigen is illustrated in Figure 9 and Table 2.

3.3.2 Filtered systematic searches

Because systematic searches scale exponentially, a systematic study of oligosaccharides conformations quickly becomes infeasible. The combinatorial explosion can be limited by filtering out conformations of the glycosidic linkages that are unfavorable in the respective disaccharide moieties. Even a very conservative filtering of 12 kcal/mol can reduce the amount of needed calculations by 50–80%.

Oligosaccharides often have conformational characteristics similar to those of their disaccharide moiety. This property can be used to rapidly build oligosaccharides from the preferred conformations assumed by the glycosidic linkages of their disaccharide moieties (Bohne *et al.* 1999; Engelsen *et al.* 1996). This approach may fail for structures where steric hindrances and interactions between neighboring monosaccharides may require the systematic study of trisaccharide or larger moieties. For trisaccharide moieties, systematic searches are still feasible (e.g. with a 15° step size, 331 776 conformations need to be sampled) and the time can be significantly reduced by pre-filtering conformations that are very unfavorable in the disaccharide moiety. This approach can be used to study oligosaccharides with branches (Rosén *et al.* 2002; Rosén *et al.* 2004) or kinks (Nyholm *et al.* 2001). Systematic searches are also used in conjunction with NOE data and MD simulations in FSPS, the Fast Sugar Structure Prediction Software (Xia *et al.* 2007a; Xia *et al.* 2007b; Xia and Margulis 2009).

The exponential complexity of such systematic searches can be avoided using stochastic methods. Such methods are considerably faster and, although not guaranteed to sample the conformational space sufficiently, they are able to identify the global energy minima efficiently in most cases. Several stochastic methods have been implemented for investigating the conformational space of oligosaccharides (Frank 2009). CICADA, Channels in Conformational space Analyzed by Driver Approach (Koča 1994), implements a guided search approach that directs the searches in the potential valleys of the conformational space. Monte Carlo methods (Peters *et al.*

1993) and genetic algorithms (Rosén *et al.* 2009; Strino *et al.* 2005) have also been applied to oligosaccharide conformations and are discussed in the next section.

3.4 GLYGAL: Genetic algorithms for sugars (Paper I)

As introduced in section 1.2.1, genetic algorithms can be applied to a variety of different problems. The software GLYGAL (GLYcosidic bonds Genetic ALgorithm) is a Java implementation of GAs for the conformational analysis of oligosaccharides. GLYGAL was initially developed during a master project (Nahmany and Strino 2004) and has been significantly improved during the course of the present work. Several features described in this chapter are not included in Paper I, since they have been part of successive development done between 2006 and 2009.

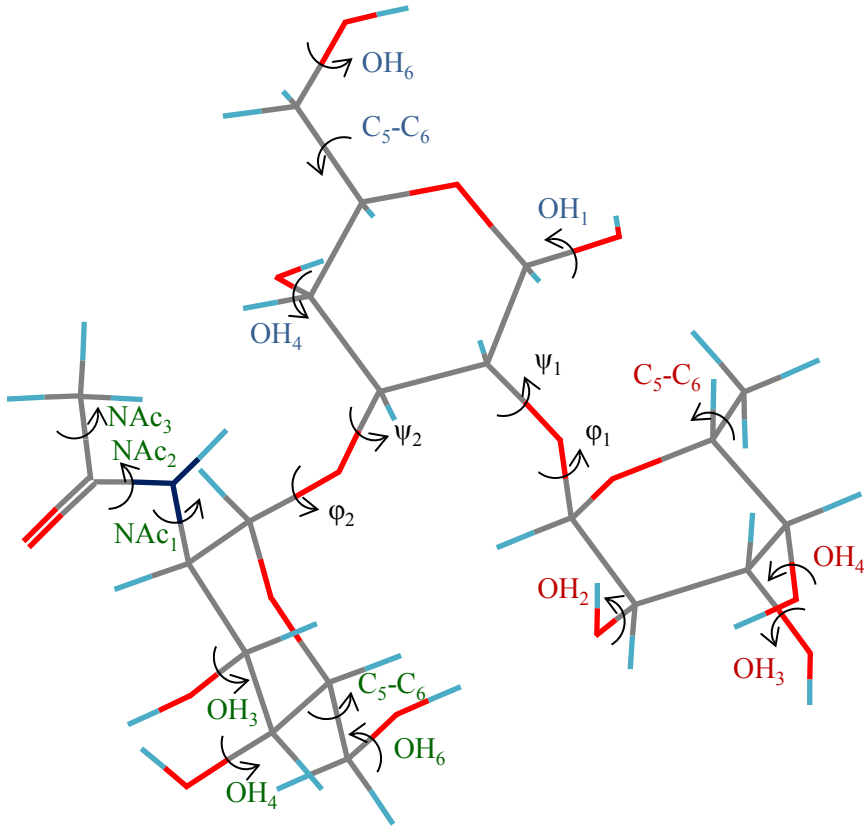
The basic ideas used to model oligosaccharide 3D structures using genetic algorithms are:

- a) Initial population of randomly generated conformations of the oligosaccharide.
- b) Evaluation using a fitness function based on molecular mechanics calculations using established force fields, in particular MM3 and MM4.
- c) A roulette wheel selection method.
- d) Standard genetic operators like mutation and crossover to generate offspring were adapted to oligosaccharide structures. *Ad hoc* operators were also created to improve the performance.
- e) Termination criteria satisfied either after a fixed number of generations or when no improvement has occurred during several generations.

3.4.1 Encoding

In order to avoid the steric clashes and the disruptions of the stereochemistry that come with jumps of Cartesian coordinates, it was chosen to represent the 3D structure of an oligosaccharide through its torsion angles. In first approximation, the conformation of the structure can be described by the values assumed by the torsion angles of each glycosidic linkage. The model can be improved later by also including the torsion angles of all the other rotatable bonds.

The torsion angles are then grouped together into units, which contain the conformational information about a glycosidic linkage or a monosaccharide. Using this approach, it is also possible to treat several dihedrals as a single item which enables some of the optimizations described later. The encoding of the torsion angles of a trisaccharide is shown in Figure 10.



$[(OH_2, OH_3, OH_4, C_5-C_6)(\phi_1, \psi_1)(OH_1, OH_4, C_5-C_6, OH_6)(\phi_2, \psi_2)(NAC_1, NAC_2, NAC_3, OH_3, OH_4, C_5-C_6, OH_6)]$

$[(-61^\circ, 60^\circ, 55^\circ-59^\circ)(-68^\circ-90^\circ)(57^\circ-59^\circ, 58^\circ-61^\circ)(83^\circ, 73^\circ)(-89^\circ, 1^\circ, 124^\circ-61^\circ-56^\circ, 55^\circ, 62^\circ)]$

Figure 10. Encoding of the three-dimensional structure of a trisaccharide using a vector of all its rotatable bonds.

3.4.2 Genetic Algorithm operators

The above representation using a torsion angle vector makes it an easy task to apply the standard GA operators described in section 1.3.2.

Mutation

The mutation operator simply changes the value of one or several torsion angles in a unit randomly. GLYGAL allows selecting the number of units to be mutated and the number of torsions to be mutated in each unit.

For glycosidic linkage units, GLYGAL offers the possibility to filter out the conformations that would be highly energetically unfavorable based on the adiabatic map of the disaccharide moiety around the linkage in question. In addition to standard random mutations, GLYGAL implements coordinated mutations for monosaccharide units so that the hydroxyls assume C/R chain conformations (see section 1.1.2) with higher probability.

Crossover

The crossover operator merges the conformations assumed by some substructures in the different parents. In order to do this, two structures are selected to be the parent structures, then a crossover point is randomly chosen and the application of the crossover operation results in two offspring. For example, let the first parent be the vector $[(38^\circ, 19^\circ), (-25^\circ, -25^\circ), (-41^\circ, 29^\circ)]$, the second parent be the vector $[(39^\circ, 25^\circ), (-19^\circ, -26^\circ), (-47^\circ, -11^\circ)]$ and the crossover point be between the first and the second glycosidic bond. The result will be the two offspring $[(38^\circ, 19^\circ), (-19^\circ, -26^\circ), (-47^\circ, -11^\circ)]$ and $[(39^\circ, 25^\circ), (-25^\circ, -25^\circ), (-41^\circ, 29^\circ)]$.

3.4.3 Fitness function and selection

As described in section 1.3.3, two important features of genetic algorithm methods are fitness evaluation and parent selection.

The natural choice for fitness evaluation is the energy of a conformation, since the Boltzmann probability of a molecule being in a certain state depends on its energy. Force field methods are widely used for predicting favorable conformations and GLYGAL can interface with MM3, MM4 and the free TINKER/MM3 version (Ponder 2010). These programs implement some of the best force fields available today for saccharides (see 3.2.1) and the minimization capabilities provided by these programs constitutes the Lamarckian part of the GLYGAL algorithm.

The construction of the next generation in a GA is straightforward after the structures have been evaluated using the fitness function and the GA operators have been applied to the present generation according to the selection criteria. The parents are chosen according to a roulette wheel selection in which the probability of being chosen is greater for energetically favorable conformers. The user can specify the probabilities of applying a

specific genetic operator to the parent structure(s), i.e. mutation, crossover, or just passing over an identical copy of the best structures (elitism).

3.4.4 Weights and adaptive parameterization

Since glycosidic linkages are more important to the overall conformation, GLYGAL provides the option to assign them a higher weight during GA searches. Such weights are used to bias the mutation and the crossover in the glycosidic linkages, avoiding less useful sampling of monosaccharide conformers, which can generally be adjusted satisfactorily by simple MM3/MM4 minimizations.

When the glycosidic torsions of the best structure remain substantially unchanged for a predefined number of generations, GLYGAL reduces the weights of the glycosidic linkages linearly until it enters a refinement stage during which the glycosidic torsions are not mutated and the conformation of the secondary torsions are optimized. If a new low-energy structure with significantly different glycosidic linkage conformations is found during this process, the initial weights are restored and the adaptive process is restarted.

3.4.5 Other features of GLYGAL

Besides the more scientifically relevant algorithm aspects described above, GLYGAL takes care of many other aspects of oligosaccharide 3D structure in over 50 000 lines of Java code. The most important are

- Input/output of MM3, MM4, TINKER xyz format files for communication with the force fields; PDB format for interfacing with the most common software; and its own compressed binary format to store millions of conformations efficiently.
- Parallelization framework to run GLYGAL jobs on a cluster of computers. Any computer can be added dynamically to the cluster, provided it has an internet connection and the possibility to run Java and the force field executables.
- Parsing of sugar molecules to identify the most common monosaccharide and linkage types. Such information is used to locate adiabatic energy maps of the disaccharide moieties relative to the glycosidic linkages specified or to create new ones before starting GA searches.
- Generation of adiabatic energy maps for disaccharides using systematic search and for trisaccharides using filtered systematic search. Creation of 2D images of such maps.

- Checking to filter out minimized conformation with artifacts such as serious steric clashes and flips of pyranose rings. This is used to filter out obviously unnatural structures that might not be treated correctly by the force fields.
- A viewer created by Ezzeddin K.B.M. Hashim during his master thesis (Hashim 2007) to visualize and cluster sugar conformers.

3.5 Molecular Dynamics

Molecular dynamics (MD) is a methodology to simulate the evolution over time for a molecular system in which the forces acting on the atoms are generally defined through the approximation of a force field (see section 3.2). The evolution is based on solving Newton's second equation $\vec{F} = m\vec{a}$ with numeric methods. In particular, time steps in the femtosecond order are considered and at each step the forces in the system are calculated and the velocities are updated.

Since snapshots of the MD simulation taken at sufficiently distant times behave like random samples of the Boltzmann energy distributions, MD simulations are routinely performed to estimate molecular flexibility, population densities of conformers and to estimate molecular properties dependent on the population distribution.

Molecular dynamics simulations have also been successfully applied to the study of oligosaccharides. They can be set up remotely with an easy web interface in Dynamic molecules (Frank *et al.* 2003), or can be set up with general software that have special parameterization for sugars like AMBER (Case *et al.* 2005; Case *et al.* 2008; Kirschner *et al.* 2008), CHARMM (MacKerell *et al.* 1998) and GROMOS (Lins and Hünenberger 2005; Scott *et al.* 1999).

3.6 Molecular docking

Molecular docking predicts the preferred orientation of two molecules with respect to each other when they are in a bound state. Once the preferred interaction pose is known, the binding affinity can be estimated by evaluating the intermolecular interactions and possibly other environmental factors. The docking between a protein and a small molecule (ligand) is the most common form of docking. The first docking programs considered both protein and ligand as rigid bodies to speed up the computation, whereas today the

flexibility of the ligand is generally taken into account. Some conformational changes on the protein can also be considered, generally limited to the conformation of the side chains in the binding site.

The most common docking programs today are AutoDock (Morris *et al.* 1998; Morris *et al.* 2009), DOCK (Ewing and Kuntz 1997), GOLD (Jones *et al.* 1997), FlexX (Rarey *et al.* 1996), Glide (Friesner *et al.* 2004; Halgren *et al.* 2004), LigandFit (Venkatachalam *et al.* 2003), and Surflex (Jain 2007). A longer list and introduction to the methodologies used by these programs are given by Souza *et al.* (2006). The performances are difficult to compare (Jason *et al.* 2005) and many studies with different results exist, as shown in Table 1 of Xun *et al.* (2009).

3.6.1 Glide

In Paper III Glide (Friesner *et al.* 2004; Halgren *et al.* 2004) was used because it has shown to be the current state of the art in many cases (Xun *et al.* 2009) and has given good results for protein-carbohydrate dockings (Agostino *et al.* 2009; Blanchard *et al.* 2008).

Glide (Grid-based Ligand Docking with Energetics) represents the shape and interaction affinities of the receptor on a grid defined by several components of interactions. It treats the ligands, but not the receptors, flexibly.

Before a docking job can be started, Glide creates several grids representing geometries and properties of the binding site on the given receptor. During docking, an exhaustive sampling of the ligand torsional space is performed to generate ligand binding poses, while several hierarchical filters reduce the search space and avoid the evaluation of computationally expensive interaction terms for non-promising conformations.

The two initial steps examine the steric complementarities between the ligand and the binding site and evaluate protein-ligand interactions in various poses with GlideScore (Friesner *et al.* 2004). During the third stage, the poses selected by the initial screening are minimized *in situ* with the OPLS-AA force field (Jorgensen *et al.* 1996). Finally, the resulting ligand binding poses are ranked according to a composite score which considers GlideScore, non-bonded interaction energies, and internal steric energies.

The extra precision (XP) mode of Glide implements a more thorough sampling procedure based on the results of the standard precision mode, as well as an improved scoring function (GlideScore XP) that includes

additional terms and a better treatment of some interaction terms with particular regard to lipophilicity and solvation effects.

3.7 Other software

Molecular visualization and editing was done using Sybyl 8.0 and Sybyl X1.1 (Tripos Inc., St. Louis, MO, USA). Solvent-accessible surfaces were generated using the MOLCAD (Heiden *et al.* 1993) module of Sybyl.

Ab initio calculations were performed using Gaussian 09A (Frisch *et al.* 2009). Molecular orbitals depictions were created with GaussView 5.0.8 (Gaussian Inc, Wallingford).

The initial three-dimensional conformations of several oligosaccharides were generated through the SWEET Web service (Bohne *et al.* 1999).

Protein-ligand interactions were analyzed using LIGPLOT (Wallace *et al.* 1995) diagrams obtained from the PDBsum Web site (Laskowski *et al.* 1997).

4 RESULTS AND DISCUSSION

In this chapter, the results from the papers are summarized. First, a comparison of the GLYGAL program with other software is given in section 4.1. Furthermore, conformational studies on the exopolysaccharide of *Burkholderia cepacia* (section 4.2) and docking studies between the capsid protein of VA387 and several histo-blood group antigens (section 4.3) are described. Finally, conformational and mimetic properties of thioglycosides and selenoglycosides are reported (section 4.4).

4.1 Performance of GLYGAL (Paper I)

The use of a stochastic method reduces drastically the exponential complexity of systematic searches. A comparison is given in Paper I where the prediction of the conformation of the O-specific oligosaccharide of *Shigella dysenteriae* type 2 required only 3000 structures to identify the global minimum and 8000 to reach the termination criteria, in contrast with 350 000 required by a systematic approach (Rosén *et al.* 2002). While GAs are not guaranteed to find all energy minima, they generally do, as shown in the case of *Shigella dysenteriae* type 4 which had also been investigated by a systematic search (Rosén *et al.* 2004).

The program SHAPE (Rosén *et al.* 2009) also implements a GA approach for the study of oligosaccharides. The main difference is in the speed/accuracy tradeoff: SHAPE introduces a taboo search approach (i.e. the energy of a conformation is not evaluated if a similar conformation has previously been considered) to speed up the computation. This increase in speed can come at the cost of ending up somewhat outside the optimal energy conformations. For example, SHAPE computational studies on the exopolysaccharide of *Burkholderia cepacia* (see Paper II) completed in a fraction of the time of our study, although only a conformation 1.2 kcal/mol higher in energy was found (Rosén *et al.* 2009). Furthermore, SHAPE does not interface with MM4 yet, although the high similarity between MM3 and MM4 should make this an easy task.

The major drawback of GLYGAL is that it does not model carbohydrate-solvent interactions accurately. This can be done using MD studies, although they have problems in jumping energy barriers efficiently and thus may not sample the conformational space thoroughly. The approaches can be combined by running short MD simulations starting from the optimized poses found using other methods. This approach is implemented automatically in

the FSPS software (Xia *et al.* 2007a; Xia *et al.* 2007b; Xia and Margulis 2009). A similar approach was used in Paper III to study the behavior of oligosaccharide ligands in the presence of protein.

4.2 Conformational studies on an exopolysaccharide of *Burkholderia cepacia* (Paper II)

Burkholderia cepacia (also known as the *Burkholderia cepacia* complex) is a family of very similar opportunistic bacteria that can cause serious lung infections and are the major cause of premature death in cystic fibrosis patients (Coenye and LiPuma 2003; Speert 2002). During the infection, they secrete large amounts of an exopolysaccharide, whose heptasaccharide repeating unit (Cescutti *et al.* 2000) is shown in Figure 11.

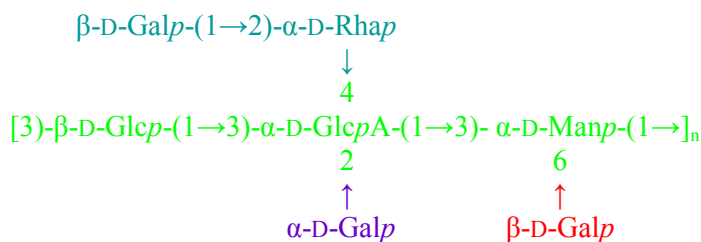


Figure 11. Sequence of the repeating unit of the exopolysaccharide of *Burkholderia cepacia*.

Paper II describes GA searches on several structures of this EPS. For the repeating unit, a global minimum and two secondary minima were found. The first minimum has a different ω angle in the $\beta\text{-D-Galp-(1}\rightarrow\text{6)-}\alpha\text{-D-Manp}$ linkage and has similar energy (only +0.1 kcal/mol), while the second has a different conformation of the $\alpha\text{-D-Rhap-(1}\rightarrow\text{4)-}\alpha\text{-D-GlcpA}$ linkage and is about 1 kcal/mol higher in energy. The 3D structures of the best conformers are shown in Figure 12a, with secondary minima shown in thin lines. Because of steric clashes, larger structures assume only the global minimum conformation of the repeating unit. A small epitope with this property is the upstream frame-shifted octasaccharide (Figure 12b).

Because of the high branching of the GlcA moiety, a systematic search study would have been unfeasible. Moreover, the branching gives this oligosaccharide structure a certain rigidity, which had also been experimentally observed (Sist *et al.* 2003). In such cases, genetic algorithms

have a clear advantage, because the minima are relatively few and well defined, in contrast to linear polysaccharides which may have too many conformations to be exhaustively sampled with a stochastic search method.

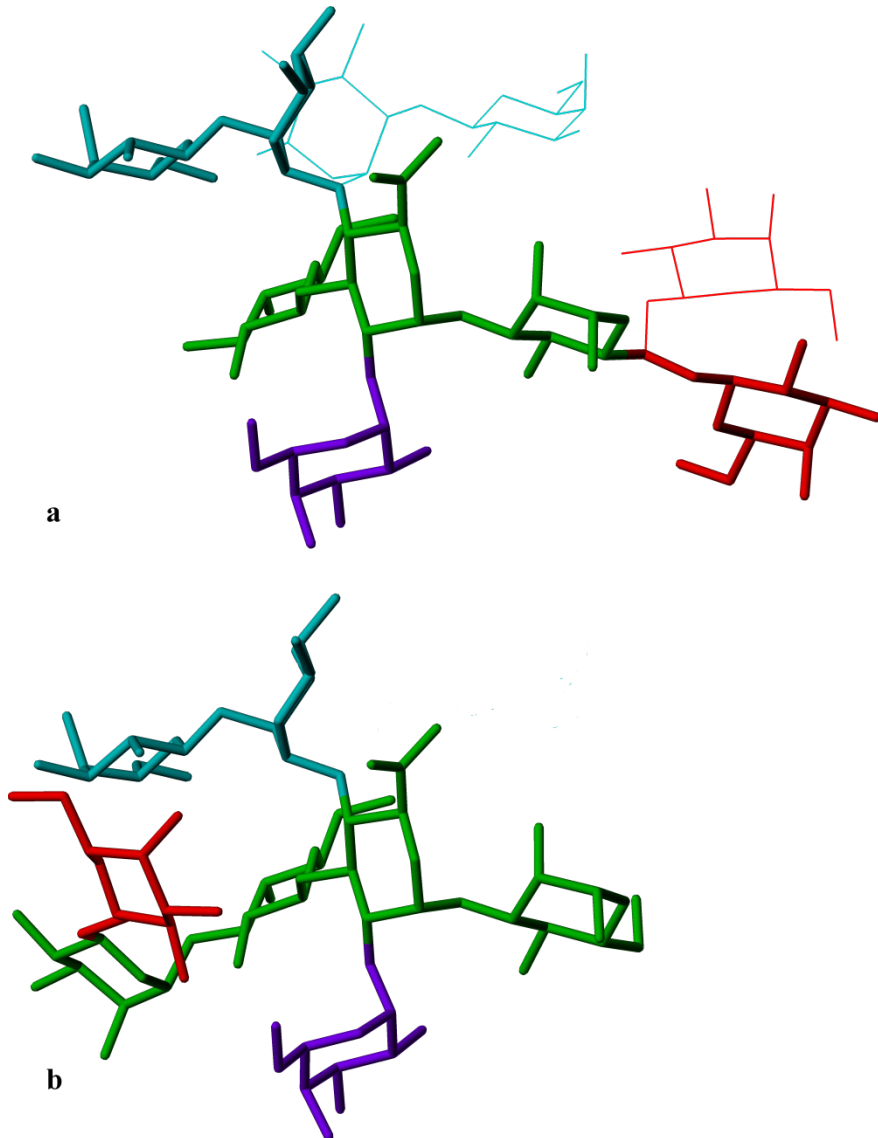


Figure 12. Favorable conformers of the exopolysaccharide of Burkholderia cepacia: (a) heptasaccharide; (b) frame-shifted octasaccharide. Secondary conformers are shown in thin lines.

Contemporaneously with our GA study, the EPS of *Burkholderia cepacia* was studied through MD (Sampaio Nogueira *et al.* 2005) in water, Me₂SO, and a 30% water-Me₂SO solution using GROMOS96 (Scott *et al.* 1999) with a specialized force field for carbohydrates (Pereira *et al.* 2004). The simulation in presence of explicit solvent gave similar results. Although the MD did not sample the less favorable secondary conformer of the Rhaα1-4GlcA moiety, the two most energetically favorable conformers were found. Nonetheless, there are minor discrepancies in the torsion angles, most likely due to differences in force field, similar to what can be observed on the superposition (Sampaio Nogueira *et al.* 2005) of the MD trajectories over the energy maps obtained with the Brant force field (Goebel *et al.* 1970).

Successive NMR studies (Herasimenka *et al.* 2008) identified several intermolecular proton-proton interactions and the estimated distances are in good agreement with the predicted conformations. These NMR results, together with atomic force microscopy (Herasimenka *et al.* 2008), also show that the larger structures of the exopolysaccharide tend to form double stranded segments in water (but not in Me₂SO). Such interactions are compatible with the topology of the solvent accessible surface (Figure 13) of our model of several repeating units, although the conformational aspects in large scale have not been thoroughly investigated yet.

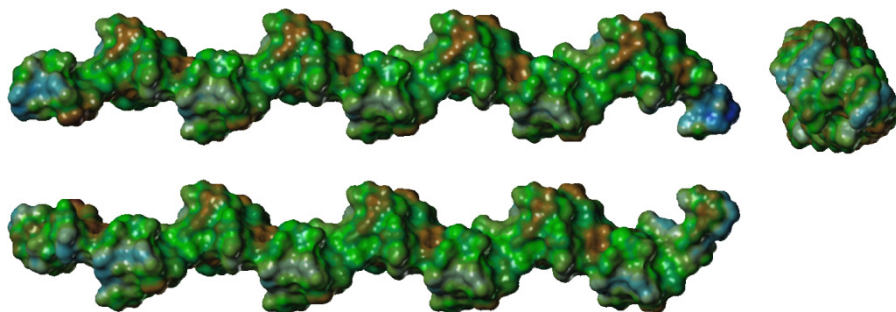


Figure 13. Orthogonal projections of the solvent-accessible surface of eight repeating units of the exopolysaccharide of *Burkholderia cepacia* colored according to lipophilicity (blue: hydrophilic, brown: lipophilic).

4.3 Interactions between the VA387 strain of norovirus and HBGAs (Paper III)

Noroviruses are the most common cause of winter gastroenteritis outbreaks in humans and are known to bind histo-blood group antigens. In Paper III the

interactions between eleven HBGA structures and the capsid protein of the VA387 strain of norovirus were studied.

Conformational studies on the free saccharides were initially performed by generating adiabatic energy maps for the disaccharide moieties and finding the optimal conformations using GLYGAL with the MM4 force field.

The initial structures were then fitted in the binding site of the VA387 capsid protein by superimposing the α 1,2-linked fucose on the respective position assumed in the crystal structure with PDB id 2OBT (Cao *et al.* 2007). Later, molecular dynamics studies with implicit solvent were performed using AMBER 8 (Case *et al.* 2005) and the GLYCAM06 (Kirschner *et al.* 2008) force field. During the MD runs, the position of the fucose moiety was constrained. The binding energies of six snapshots were finally calculated for each saccharide using Glide XP (Friesner *et al.* 2004; Halgren *et al.* 2004) and the results are graphed in Figure 14. The constraint is well justified by the strong interaction between the α 1,2-linked Fuc and the protein, which alone is responsible for more than half of the binding energy of tri- to pentasaccharide ligands.

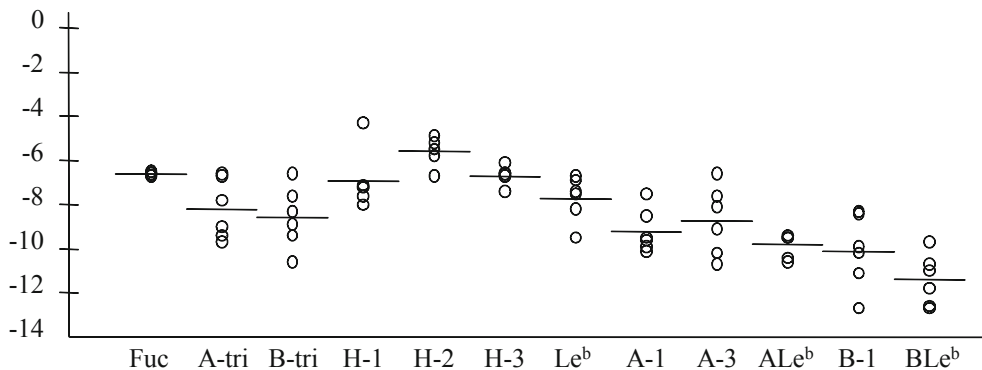


Figure 14. Estimated binding energies (in kcal/mol) between the capsid protein of VA387 and 11 HBGA saccharides. For definitions see section 1.2.1.

The difference between type 1 and the related Le^b structures also indicates that the addition of an α 1,4-linked fucose increases the binding affinity of the saccharide ligands.

Furthermore, the MD simulations show that A structures, but not B structures, interact with the residues I389 and Q331, in agreement with mutational data (Tan *et al.* 2008).

4.4 Conformational studies on thioglycosides and selenoglycosides (Papers IV and V)

Sugar derivatives in which glycosidic oxygens are substituted are of pharmaceutical interest because of their enhanced resistance to hydrolysis. Sulfur and selenium substitution are a natural choice because they are in the same column of the periodic table as oxygen.

Conformational properties of hetero- substituted linkages

The first step of the investigation was to study the differences between the natural and the substituted glycosidic linkages. It can be seen that O→S, O→Se and, for comparison, O→C substitutions of the interglycosidic linkage substantially affect the geometrical properties, as shown in Table 3.

Table 3. Geometric properties of the natural glycosidic linkage and its C-, S- and Se- derivatives. For each linkage type van der Waals radius of the central atom, length of the C-x bond, value of the C-x-C angle and distance between the two ends of the glycosidic linkage are listed.

Linkage	vdW Radius	C-x Bond	C-x-C Angle	C-x-C Distance
O	1.5 Å	1.4 Å	115°	2.4 Å
C	1.7 Å	1.5 Å	110°	2.5 Å
S	1.8 Å	1.8 Å	95°	2.9 Å
Se	1.9 Å	1.9 Å	95°	3.0 Å

The substitution of the glycosidic oxygen can also affect the conformational properties of the glycosidic linkage, in particular the preferences of the φ torsion. The torsional profiles for φ (i.e. O-C-x-C) calculated at the MP2/6-311++G** level of theory (paper V and unpublished data) locking the C-O-C-x torsion at 60° and 180° are shown in Figure 15. In pyranoses where the anomeric carbon assumes the α conformation, the C-O-C-x ring torsion is constrained at around $\pm 60^\circ$ (the profile at -60° is the mirror image of that at 60°), while in β -pyranoses the C-O-C-x torsion is constrained at around 180°.

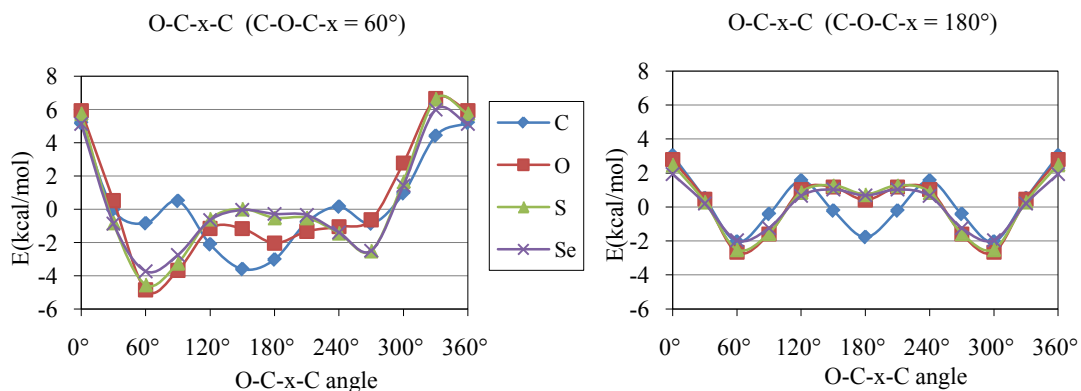


Figure 15. Torsional profiles of the ϕ torsion of the glycosidic linkages containing C, O, S and Se.

A comparison with the profile of the C-glycosidic linkage illustrates that O-, S- and Se- derivatives are subject to the exo-anomeric effect and prefer gauche conformations around $\pm 60^\circ$. The S- and Se- profiles are very similar and their main difference from the natural glycosidic linkage is in the profile with C-O-C-x fixed at 60° , where the secondary conformation at -60° has an energy difference from the global minimum of around 4 kcal/mol in the O-linkage and of 2 kcal/mol in the S- and Se- linkages.

The *ab initio* data for the thioglycosidic and the selenoglycosidic linkages were used to derive new MM4 parameters (called MM4R), which have better performance than those obtained with its parameter estimator (Allinger *et al.* 1994). The MM4R parameter set is attached as supplementary material in paper V and will be integrated in the next release of MM4. The results for the thioglycosides are in good agreement with similar parameters developed for use with AMBER (Saito and Okazaki 2009).

Conformational properties of ABH derivatives

After generating a proper force field parameterization for S- and Se- derivatives, the mimetic properties of the ABH blood group antigen derivatives were considered next.

The conformational properties and population densities were studied by the generation of adiabatic energy maps through systematic and filtered systematic searches. As previously observed for a thio-disaccharide (Aguilera *et al.* 1998), the investigations on the disaccharide moieties indicated that S-

and Se- linkages have enhanced flexibility and can assume two different conformers, although they generally prefer the conformation assumed by the O-glycosidic linkage (Table 4).

Table 4. Characterization of the energetically favorable conformations of the disaccharide moieties of histo-blood group ABH antigens and their derivatives.

Compound	ϕ/ψ	Energy (kcal/mol)	Population
Fuc-(1→2)-Gal	-90°/-120°		99.0%
Fuc-(1-S→2)-Gal	-90°/-90°		62.4%
	-90°/45°	+ 0.03	35.1%
Fuc-(1-Se→2)-Gal	-75°/60°		45.3%
	-75°/-90°	+0.37	51.9%
GalNAc-(1→3)-Gal	75°/90°		93.4%
GalNAc-(1-S→3)-Gal	75°/60°		77.3%
	75°/-75°	+ 0.40	14.8%
GalNAc-(1-Se→3)-Gal	75°/-75°		35.5%
	120°/60°	+0.24	58.0%
Gal-(1→3)-Gal	75°/90°		93.4%
Gal-(1-S→3)-Gal	90°/75°		75.4%
	75°/-60°	+ 0.43	19.2%
Gal-(1-Se→3)-Gal	75°/-75°		49.5%
	60°/75°	+1.01	42.2%

The studies on the trisaccharide derivatives predict a high variability in mimetic properties between the different derivatives. It can be noted that substitutions in the α -L-Fucp-(1→2)- β -D-Galp linkage do not seem to affect the conformational properties significantly, while fully substituted derivatives can generally assume several different conformers (Table 5 and Table 6).

Table 5. Characterization of the energetically favorable conformations of the histo-blood group A antigen and its derivatives. The first element of the A - x/x notation refers to the α -L-Fucp-(1-X \rightarrow 2)- β -D-Galp linkage, while the second refers to the α -D-GalpNAc-(1-X \rightarrow 3)- β -D-Galp linkage.

Compound	Fuc-(1-X \rightarrow 2)-Gal	GalNAc-(1-X \rightarrow 3)-Gal	Energy (kcal/mol)	Population
A - O/O	-75°/-105°	90°/75°		90.9%
A - O/S	-90°/-120°	75°/-60°		89.0%
	-90°/-120°	105°/75°	+2.38	3.7%
A - O/Se	-105°/-165°	120°/150°		29.5%
	-75°/-120°	75°/75°	+0.12	56.6%
	-90°/-150°	105°/-60°	+1.48	4.3%
A - S/O	-75°/-135°	90°/90°		94.0%
A - S/S	-75°/-135°	90°/105°		42.3%
	-75°/-120°	75°/-60°	+0.31	19.6%
	-120°/60°	60°/75°	+0.86	12.8%
	-90°/60°	75°/-60°	+0.91	13.0%
A - S/Se	-75°/75°	120°/75°		57.9%
	-120°/-165°	75°/-90°	+1.07	4.7%
	-75°/-120°	75°/90°	+1.30	20.5%
A - Se/O	-75°/-120°	90°/90°		90.9%
	-60°/75°	105°/75°	+1.66	2.5%
A - Se/S	-60°/-120°	75°/-60°		30.5%
	-60°/75°	105°/75°	+0.37	38.9%
	-75°/-105°	75°/90°	+1.88	11.1%
	-75°/60°	90°/-75°	+1.88	2.7%
A - Se/Se	-75°/-135°	75°/-60°		40.0%
	-60°/75°	120°/75°	+0.95	7.9%
	-75°/-135°	75°/90°	+1.07	31.9%
	-105°/-60°	45°/60°	+1.90	2.6%

Table 6. Characterization of the energetically favorable conformations of the histo-blood group B antigen and its derivatives. The first element of the B - x/x notation refers to the α -L-Fucp-(1-X \rightarrow 2)- β -D-Galp linkage, while the second refers to the α -D-Galp-(1-X \rightarrow 3)- β -D-Galp linkage.

Compound	Fuc-(1-X \rightarrow 2)-Gal	Gal-(1-X \rightarrow 3)-Gal	Energy (kcal/mol)	Population
B - O/O	-75°/-105°	90°/75°		91.4%
	-135°/-150°	150°/135°	+1.24	1.6%
B - O/Se	-75°/-120°	60°/75°		65.3%
	-105°/-165°	120°/150°	+0.03	27.4%
B - Se/O	-60°/-105°	90°/90°		94.1%
	-75°/75°	105°/90°	+1.78	1.6%
B - Se/Se	-75°/-135°	60°/75°		60.3%
	-75°/-135°	75°/-60°	+0.21	19.4%
	-105°/60°	45°/60°	+0.76	7.4%
	-75°/60°	75°/-60°	+1.31	2.9%
B - S/Se	-120°/60°	60°/75°		31.9%
	-75°/-120°	60°/90°	+0.52	53.7%
	-90°/60°	75°/-75°	+1.87	1.6%
B - Se/S	-60°/75°	105°/75°		31.3%
	-75°/-105°	75°/90°	+0.57	41.6%
	-75°/-120°	90°/-60°	+1.06	6.9%
	-75°/60°	90°/-75°	+1.25	5.3%
	-60°/-105°	120°/150°	+1.29	1.9%
B - O/S	-75°/-135°	75°/-60°		60.4%
	-75°/-120°	90°/75°	+0.82	33.3%
B - S/O	-75°/-135°	90°/90°		90.4%
B - S/S	-75°/-135°	90°/105°		53.4%
	-90°/-105°	75°/-60°	+0.91	21.9%
	-120°/60°	60°/75°	+1.16	6.9%
	-90°/60°	75°/-60°	+1.55	4.4%

The potential bioactivity of the derivatives was finally investigated through dockings with eight known carbohydrate-binding proteins for which crystal structures and in some cases experimental binding data were available, namely FcsSBP (Higgins *et al.* 2009), CGL2 (Walser *et al.* 2004), GH98CBM51 (Gregg *et al.* 2008), VAA (Jiménez *et al.* 2008), MOA (Grahn *et al.* 2009), PA-IL (Blanchard *et al.* 2008) and AAA (Bianchet *et al.* 2002). Although the results of such studies should be considered as semi-quantitative and in-depth studies of the structure-activity relationship may be needed for each protein, the results suggested that the increased flexibility of

some derivatives may in fact improve the interactions with the receptor, as shown in Figure 16 for CGL2.

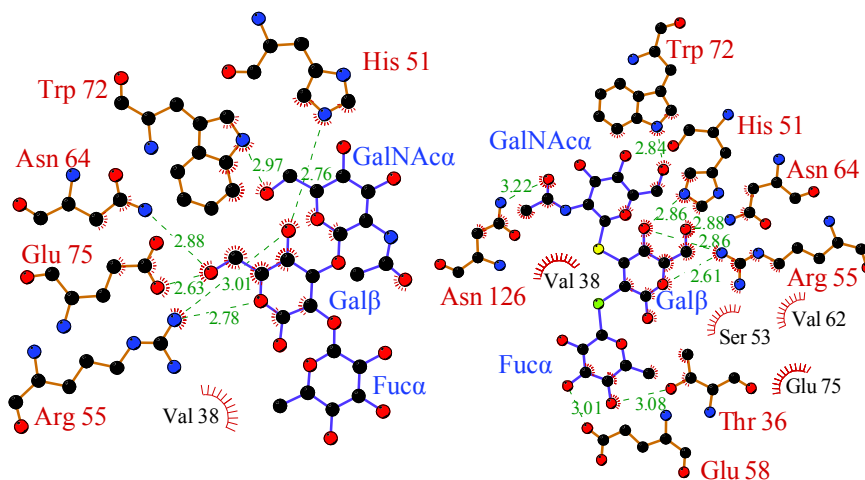
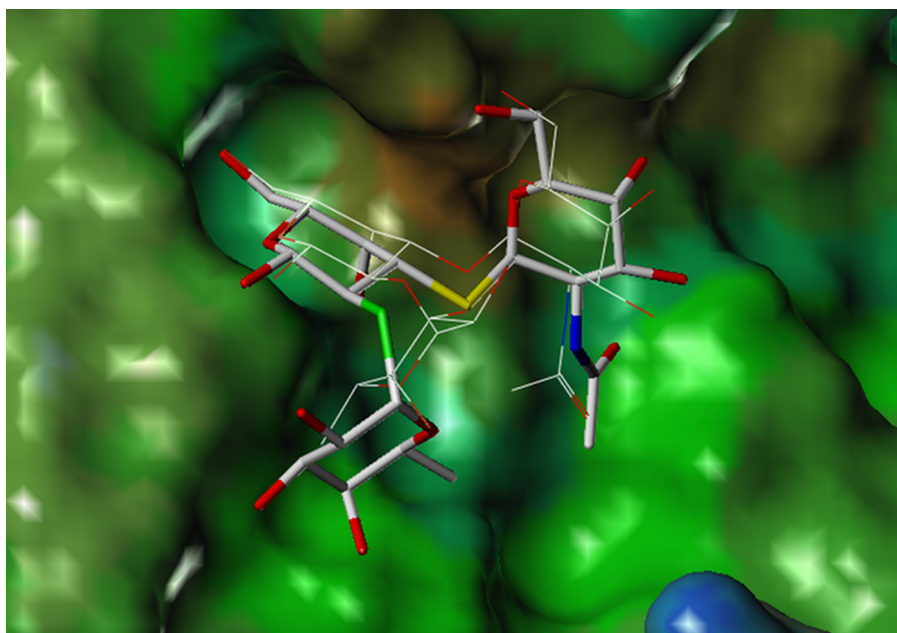


Figure 16. Binding interactions of CGL2 with the blood group antigen A and its Se/S-derivative. The docking pose of A-Se/S and the co-crystallized A structure (thin lines) are shown above, with the solvent-accessible surface of the protein colored according to hydrophilicity (blue: hydrophilic, brown: hydrophobic). 2D protein-ligand interaction plots for A (left) and A-Se/S (right) are depicted below.

5 CONCLUSIONS

During the present work, a methodology for the study of the conformations of oligosaccharides and some derivatives has been developed. GLYGAL has been shown to be able to predict the conformational properties of oligosaccharides and, after some *ab initio* studies, of some sugar derivatives.

The conformational studies on the exopolysaccharide of *Burkholderia cepacia* suggests a rigid structure stabilized by several intramolecular interactions.

The docking studies with the norovirus capsid protein give insights into the binding specificity of a series of histo-blood group type 1 and 3 antigens. Furthermore, the results can explain some mutational data on the binding site and provide a basis for structure-based design of new inhibitors.

The work on thioglycosides and selenoglycosides suggests a novel class of glycomimetics that should have increased resistance to hydrolysis and may have enhanced binding activity.

6 FUTURE PERSPECTIVES

After several years of development the GLYGAL methodology is now mature, although some improvements could be done. For instance, a method to estimate population densities from GA results (and possibly some reduced systematic searches) would be an interesting improvement.

The work on norovirus could be extended to other histo-blood group antigens, to some sugar mimetics, and to other virus strains.

The work on thioglycosides and selenoglycosides indicates a pharmaceutical potential for this class of compounds, but the quality of the structure-activity relationship models for specific proteins could be improved by the hybrid MD-docking approach used for the norovirus. The investigation could be expanded to other medically relevant carbohydrate epitopes.

Finally, experimental testing of the *in silico* results could give validation or input for improvements to the methodology used and the predictions produced.

ACKNOWLEDGEMENTS

The present work could not have been possible without the ideas, the help, the support, and the patience of several persons.

First of all, I want to thank my supervisor and employer Per-Georg Nyholm for supporting this work, for coming up with many new and interesting ideas and for always listening to mine.

Another special thank goes to Graham Kemp for countless ideas, precious suggestions and for spotting many grammatical mistakes.

Avi, Ezzeddin and Jimmy for their great help and consistent contributions to the development of GLYGAL.

Irmin Pascher for sharing his enormous knowledge of glycolipid chemistry, for several corrections and for invaluable feedbacks.

Göran Larson and Gustaf Rydell for a very interesting and stimulating collaboration on norovirus.

Hans-Joachim Gabius for interesting ideas on carbohydrate mimetics and many insights on sugar biology and biochemistry.

Jenn-Huei Lii for a great help with *ab initio* calculations and support for the use and parameterization of MM4.

Thank you Chaitanya for help with the dockings, coffee breaks, a good friendship and for making the office life more interesting. The current master students Waqas and Niklas, as well as all former colleagues at Biognos, for creating a very enjoyable working environment, interesting conversations and challenging pool games. Furthermore, I'm very grateful for the financial support from Biognos AB, which made this work possible in the first place.

Thank you Edmée for the most pleasurable form of constant annoyance. Thank you Fabio for a myriad of moderately crazy ideas and memorable trips. Thank you Giulia, Paola, Gaia, Marco, Alessandro, Giulio, and a list of names too long to enumerate for your friendship and for the good times.

My final thanks go to my mother Laura, my father Carlo and my brother Alessandro for support, understanding, and a lot of patience.

REFERENCES

- Agostino, M., C. Jene, T. Boyle, P. A. Ramsland and E. Yuriev (2009). "Molecular docking of carbohydrate ligands to antibodies: structural validation against crystal structures." *J Chem Inf Model* 49(12): 2749-60.
- Aguilera, B., J. Jiménez-Barbero and A. Fernández-Mayoralas (1998). "Conformational differences between Fuc(α 1-3)GlcNAc and its thioglycoside analogue." *Carbohydr Res* 308(1-2): 19-27.
- Ali, M. M. N., U. Aich, S. Pérez, A. Imberty and D. Loganathan (2009). "Examination of the effect of structural variation on the N-glycosidic torsion (Φ_N) among N-(β -D-glycopyranosyl)acetamido and propionamido derivatives of monosaccharides based on crystallography and quantum chemical calculations." *Carbohydr Res* 344(3): 355-61.
- Allinger, N. L., K.-H. Chen, J.-H. Lii and K. A. Durkin (2003). "Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds." *J Comput Chem* 24(12): 1447-72.
- Allinger, N. L., Y. H. Yuh and J.-H. Lii (1989). "Molecular mechanics. The MM3 force field for hydrocarbons. 1." *J Am Chem Soc* 111(23): 8551-66.
- Allinger, N. L., X. Zhou and J. Bergsma (1994). "Molecular mechanics parameters." *J Mol Struct* 312(1): 69-83.
- André, S., Z. Pei, H.-C. Siebert, O. Ramström and H.-J. Gabius (2006). "Glycosyldisulfides from dynamic combinatorial libraries as O-glycoside mimetics for plant and endogenous lectins: their reactivities in solid-phase and cell assays and conformational analysis by molecular dynamics simulations." *Bioorg Med Chem* 14(18): 6314-26.
- Asensio, J., J. Espinosa, H. Dietrich, R. Schmidt, M. Martín-Lomas, S. André, H.-J. Gabius and J. Jiménez-Barbero (1999). "Bovine Heart Galectin-1 Selects a Unique (Syn) Conformation of C-Lactose, a Flexible Lactose Analogue." *J Am Chem Soc* 121(39): 8995-9000.
- Astronomo, R. D. and D. R. Burton (2010). "Carbohydrate vaccines: developing sweet solutions to sticky situations?" *Nat Rev Drug Discov* 9(4): 308-24.
- Azevedo, M., S. Eriksson, N. Mendes, J. Serpa, C. Figueiredo, L. P. Resende, N. Ruvoën-Clouet, R. Haas, T. Borén, J. Le Pendu and L. David (2008). "Infection by *Helicobacter pylori* expressing the BabA adhesin is influenced by the secretor phenotype." *J Pathol* 215(3): 308-16.
- Azurmendi, H. F. and C. A. Bush (2002). "Conformational studies of blood group A and blood group B oligosaccharides using NMR residual dipolar couplings." *Carbohydr Res* 337(10): 905-15.

- Bathe, M., G. C. Rutledge, A. J. Grodzinsky and B. Tidor (2005). "A coarse-grained molecular model for glycosaminoglycans: application to chondroitin, chondroitin sulfate, and hyaluronic acid." *Biophys J* 88(6): 3870-87.
- Becke, A. D. (1993). "Density-functional thermochemistry. III. The role of exact exchange." *Chem Phys* 98(1): 5648-52.
- Bianchet, M. A., E. W. Odom, G. R. Vasta and L. M. Amzel (2002). "A novel fucose recognition fold involved in innate immunity." *Nat Struct Biol* 9(8): 628-34.
- Biles, J. A. (1994). *GenJam: A genetic algorithm for generating jazz solos. Proceedings of the International Computer Music Conference.*
- Blanchard, B., A. Nurisso, E. Hollville, C. Tétaud, J. Wiels, M. Pokorná, M. Wimmerová, A. Varrot and A. Imberty (2008). "Structural basis of the preferential binding for globo-series glycosphingolipids displayed by *Pseudomonas aeruginosa* lectin I." *J Mol Biol* 383(4): 837-53.
- Bohne, A., E. Lang and C. W. von der Lieth (1999). "SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides." *Bioinformatics* 15(9): 767-8.
- Buts, L., R. Loris, E. De Genst, S. Oscarson, M. Lahmann, J. Messens, E. Brosens, L. Wyns, H. De Greve and J. Bouckaert (2003). "Solving the phase problem for carbohydrate-binding proteins using selenium derivatives of their ligands: a case study involving the bacterial F17-G adhesin." *Acta Crystallogr D Biol Crystallogr* 59(Pt 6): 1012-5.
- Cao, S., Z. Lou, M. Tan, Y. Chen, Y. Liu, Z. Zhang, X. C. Zhang, X. Jiang, X. Li and Z. Rao (2007). "Structural basis for the recognition of blood group trisaccharides by norovirus." *J Virol* 81(11): 5949-57.
- Case, D. A., T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods (2005). "The Amber biomolecular simulation programs." *J Comput Chem* 26(16): 1668-88.
- Case, D. A., T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman (2008). "AMBER 10." University of California, San Francisco.
- Cescutti, P., M. Bosco, F. Picotti, G. Impallomeni, J. H. Leitão, J. A. Richau and I. Sá-Correia (2000). "Structural study of the exopolysaccharide produced by a clinical isolate of *Burkholderia cepacia*." *Biochem Biophys Res Commun* 273(3): 1088-94.
- Chakka, N., B. D. Johnston and B. M. Pinto (2005). "Synthesis and conformational analysis of disaccharide analogues containing disulfide and

- selenosulfide functionalities in the interglycosidic linkages.*" *Can J Chem* 83(6-7): 929-36.
- Clausen, H. and S.-I. Hakomori (1989). "ABH and Related Histo-Blood Group Antigens; Immunochemical Differences in Carrier Isotypes and Their Distribution." *Vox Sang* 56(1): 1-20.
- Coenye, T. and J. J. LiPuma (2003). "Molecular epidemiology of *Burkholderia* species." *Front Biosci* 8: e55-67.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* D. Appleton.
- Davidson, E. R. and D. Feller (1986). "Basis set selection for molecular calculations." *Chemical Reviews* 86(4): 681-96.
- Engelsen, S. B., S. Cros, W. Mackie and S. Pérez (1996). "A molecular builder for carbohydrates: application to polysaccharides and complex carbohydrates." *Biopolymers* 39(3): 417-33.
- Ernst, B. and J. L. Magnani (2009). "From carbohydrate leads to glycomimetic drugs." *Nat Rev Drug Discov* 8(8): 661-77.
- Ewing, T. J. A. and I. D. Kuntz (1997). "Critical evaluation of search algorithms for automated molecular docking and database screening." *J Comput Chem* 18(9): 1175-89.
- Fock, V. (1930). "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems." *Z Phys A: Hadrons Nucl* 61(1): 126-48.
- Frank, M. (2009). *Predicting Carbohydrate 3D Structures Using Theoretical Methods.* In: C. W. von der Lieth, T. Lütteke and M. Frank (eds). *Bioinformatics for Glycobiology and Glycomics: an Introduction.* Wiley-Blackwell:359-88.
- Frank, M., P. Gutbrod, C. Hassayoun and C. W. von Der Lieth (2003). "Dynamic molecules: molecular dynamics for everyone. An internet-based access to molecular dynamic simulations: basic concepts." *J Mol Model* 9(5): 308-15.
- French, A. D., G. P. Johnson, A.-M. Kelterer, M. K. Dowd and C. J. Cramer (2001). "QM/MM distortion energies in di- and oligosaccharides complexed with proteins." *Int J Quantum Chem* 84(4): 416-25.
- French, A. D., V. H. Tran and S. Pérez (1990). *Conformational Analysis of a Disaccharide (Cellobiose) with the Molecular Mechanics Program (MM2).* In: A. D. French and J. W. Brady (eds). *Computer Modeling of Carbohydrate Molecules. Volume 340.* American Chemical Society:191-212.
- Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin (2004). "Glide: A New Approach for Rapid, Accurate

Docking and Scoring. I. Method and Assessment of Docking Accuracy. "J Med Chem 47(7): 1739-49.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A. Peralta, J. E. Ogliaro, M. F.; Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. C. Tomasi, M., N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox (2009). Gaussian 09, Revision A. 1. Gaussian Inc., Wallingford.

Galanina, O. E., H. Kaltner, L. S. Khraltsova, N. V. Bovin and H.-J. Gabius (1997). "Further refinement of the description of the ligand-binding characteristics for the galactoside-binding mistletoe lectin, a plant agglutinin with immunomodulatory potency." *J Mol Recognit* 10(3): 139-47.

Gammelgaard, B., L. Bendahl, N. Jacobsen and S. Stürup (2005). "Quantitative determination of selenium metabolites in human urine by LC-DRC-ICP-MS." *J Anal At Spectrom* 20(9): 889-93.

Giebink, G. S., M. Koskela, P. P. Vella, M. Harris and C. T. Le (1993). "Pneumococcal capsular polysaccharide-meningococcal outer membrane protein complex conjugate vaccines: Immunogenicity and efficacy in experimental pneumococcal otitis media." *J Infect Dis* 167(2): 347-55.

Goebel, K., D. Brant and W. Dimpfl (1970). "The Conformational Energy of Maltose and Amylose." *Macromolecules* 3(5): 644-54.

Gordon, M. S. and M. W. Schmidt (2005). *Advances in electronic structure theory: GAMESS a decade later*. In: C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria (eds). *Theory and Applications of Computational Chemistry*. Elsevier: 1167-89.

Grahn, E. M., H. C. Winter, H. Tateno, I. J. Goldstein and U. Krenzel (2009). "Structural characterization of a lectin from the mushroom *Marasmius oreades* in complex with the blood group B trisaccharide and calcium." *J Mol Biol* 390(3): 457-66.

Gregg, K. J., R. Finn, D. W. Abbott and A. B. Boraston (2008). "Divergent modes of glycan recognition by a new family of carbohydrate-binding modules." *J Biol Chem* 283(18): 12604-13.

- Guo, Z. and Q. Wang (2009). "Recent development in carbohydrate-based cancer vaccines." *Curr Opin Chem Biol* 13(5-6): 608-17.
- Ha, S. N., L. J. Madsen and J. W. Brady (1988). "Conformational analysis and molecular dynamics simulations of maltose." *Biopolymers* 27(12): 1927-52.
- Halgren, T. A., R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks (2004). "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening." *J Med Chem* 47(7): 1750-9.
- Hanada, K. (2005). "Sphingolipids in infectious diseases." *Jpn J Infect Dis* 58(3): 131-48.
- Hashim, E. K. B. M. (2007). *Visualizing 3D oligosaccharide structures predicted by the Genetic Algorithm program GLYGAL [M.Sc. thesis]*. Göteborg: Chalmers University of Technology, 2007.
- Hassel, O. and B. Ottar (1947). "The structure of molecules containing cyclohexane or pyranose rings." *Acta chem scand* 1: 929-42.
- Haynes, L. J. (1965). *Naturally Occurring C-Glycosyl Compounds*. In: L. W. Melville (ed). *Advances in Carbohydrate Chemistry*. Volume 20. Academic Press:357-69.
- Hecht, M.-L., P. Stallforth, D. V. Varón Silva, A. Adibekian and P. H. Seeberger (2009). "Recent advances in carbohydrate-based vaccines." *Curr Opin Chem Biol* 13(3): 354-9.
- Hehre, W. J. (2003). *A guide to molecular mechanics and quantum chemical calculations*. Wavefunction Inc.
- Hehre, W. J. and W. W. Huang (1995). *Chemistry with computation: an introduction to SPARTAN*. Wavefunction, Inc.
- Heiden, W., G. Moeckel and J. Brickmann (1993). "A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces." *J Comput Aided Mol Des* 7(5): 503-14.
- Hemmingsen, L., D. E. Madsen, A. L. Esbensen, L. Olsen and S. B. Engelsen (2004). "Evaluation of carbohydrate molecular mechanical force fields by quantum mechanical calculations." *Carbohydr Res* 339(5): 937-48.
- Herasimenka, Y., P. Cescutti, C. E. Sampaio Noguera, J. R. Ruggiero, R. Urbani, G. Impallomeni, F. Zanetti, S. Campidelli, M. Prato and R. Rizzo (2008). "Macromolecular properties of cepacian in water and in dimethylsulfoxide." *Carbohydr Res* 343(1): 81-9.
- Higgins, M. A., D. W. Abbott, M. J. Boulanger and A. B. Boraston (2009). "Blood group antigen recognition by a solute-binding protein from a serotype 3 strain of *Streptococcus pneumoniae*." *J Mol Biol* 388(2): 299-309.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.

Imberty, A., E. Mikros, J. Koča, R. Mollicone, R. Oriol and S. Pérez (1995). "Computer simulation of histo-blood group oligosaccharides: energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens." *Glycoconj J* 12(3): 331-49.

Jahn, M., J. Marles, R. A. J. Warren and S. G. Withers (2003). "Thioglycoligases: mutant glycosidases for thioglycoside synthesis." *Angew Chem Int Ed Engl* 42(3): 352-4.

Jain, A. N. (2007). "Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search." *J Comput Aided Mol Des* 21(5): 281-306.

Jason, C. C., W. M. Christopher, J. W. M. Nissink, D. T. Richard and T. Robin (2005). "Comparing protein-ligand docking programs is difficult." *Proteins* 60(3): 325-32.

Jeffrey, G. A. (1990). "Crystallographic studies of carbohydrates." *Acta Crystallogr B* 46(Pt 2): 89-103.

Jiménez-Barbero, J., J. F. Espinosa, J. L. Asensio, F. J. Cañada and A. Poveda (2000). "The conformation of C-glycosyl compounds." *Adv Carbohydr Chem Biochem* 56: 235-84.

Jiménez, M., S. André, C. Barillari, A. Romero, D. Rognan, H.-J. Gabius and D. Solís (2008). "Domain versatility in plant AB-toxins: evidence for a local, pH-dependent rearrangement in the 2gamma lectin site of the mistletoe lectin by applying ligand derivatives and modelling." *FEBS Lett* 582(15): 2309-12.

Johnson, G. P., L. Petersen, A. D. French and P. J. Reilly (2009). "Twisting of glycosidic bonds by hydrolases." *Carbohydr Res* 344(16): 2157-66.

Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor (1997). "Development and validation of a genetic algorithm for flexible docking." *J Mol Biol* 267(3): 727-48.

Jorgensen, W. L., D. S. Maxwell and J. Tirado-Rives (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." *J Am Chem Soc* 118(45): 11225-36.

Kaminski, G. A., R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen (2001). "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides." *J Phys Chem B* 105(28): 6474-87.

Kayser, K., N. V. Bovin, E. Y. Korchagina, C. Zeilinger, F.-Y. Zeng and H.-J. Gabius (1994). "Correlation of expression of binding sites for synthetic blood

- group A-, B- and H-trisaccharides and for sarcolectin with survival of patients with bronchial carcinoma." *Eur J Cancer* 30A(5): 653-7.
- Kim, K. and K. D. Jordan (1994). "Comparison of Density Functional and MP2 Calculations on the Water Monomer and Dimer." *J Phys Chem* 98(40): 10089-94.
- Kim, Y.-W., A. L. Lovering, H. Chen, T. Kantner, L. P. McIntosh, N. C. J. Strynadka and S. G. Withers (2006). "Expanding the thioglycoligase strategy to the synthesis of α -linked thioglycosides allows structural investigation of the parent enzyme/substrate complex." *J Am Chem Soc* 128(7): 2202-3.
- Kirschner, K. N., A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods (2008). "GLYCAM06: a generalizable biomolecular force field. Carbohydrates." *J Comput Chem* 29(4): 622-55.
- Kobayashi, Y., Y. Ogra, K. Ishiwata, H. Takayama, N. Aimi and K. T. Suzuki (2002). "Selenosugars are key and urinary metabolites for selenium excretion within the required to low-toxic range." *Proc Natl Acad Sci U S A* 99(25): 15932-6.
- Koča, J. (1994). "Computer program CICADA - travelling along conformational potential energy hypersurface." *J Mol Struct* 308: 13-24.
- Kräutler, V., M. Müller and P. H. Hünenberger (2007). "Conformation, dynamics, solvation and relative stabilities of selected β -hexopyranoses in water: a molecular dynamics study with the gromos 45A4 force field." *Carbohydr Res* 342(14): 2097-124.
- Lamarck, J. B. (1809). *Philosophie zoologique, ou Exposition des considérations relatives à l'histoire naturelle des animaux*. Dentu.
- Landsteiner, K. (1900). "Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe." *Zbl Bakt* 27: 357-62.
- Langley, C. H. and N. L. Allinger (2002). "Molecular Mechanics (MM4) Calculations on Amides." *J Phys Chem A* 106(23): 5638-52.
- Laskowski, R. A., E. G. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones and J. M. Thornton (1997). "PDBsum: a web-based database of summaries and analyses of all PDB structures." *Trends Biochem Sci* 22(12): 488-90.
- Le Pendu, J. (2004). "Histo-blood group antigen and human milk oligosaccharides: genetic polymorphism and risk of infectious diseases." *Adv Exp Med Biol* 554: 135-43.
- Lee, C., W. Yang and R. G. Parr (1988). "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density." *Phys Rev B* 37(2): 785-9.

Lii, J.-H. and N. L. Allinger (1989a). "Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics." *J Am Chem Soc* 111(23): 8566-75.

Lii, J.-H. and N. L. Allinger (1989b). "Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons." *J Am Chem Soc* 111(23): 8576-82.

Lii, J.-H., K.-H. Chen and N. L. Allinger (2003a). "Alcohols, ethers, carbohydrates, and related compounds. IV. Carbohydrates." *J Comput Chem* 24(12): 1504-13.

Lii, J.-H., K.-H. Chen and N. L. Allinger (2004). "Alcohols, Ethers, Carbohydrates, and Related Compounds Part V. The Bohlmann Torsional Effect." *J Phys Chem A* 108(15): 3006-15.

Lii, J.-H., K.-H. Chen, K. A. Durkin and N. L. Allinger (2003b). "Alcohols, ethers, carbohydrates, and related compounds. II. The anomeric effect." *J Comput Chem* 24(12): 1473-89.

Lii, J.-H., K.-H. Chen, T. B. Grindley and N. L. Allinger (2003c). "Alcohols, ethers, carbohydrates, and related compounds. III. The 1,2-dimethoxyethane system." *J Comput Chem* 24(12): 1490-503.

Lii, J.-H., K.-H. Chen, G. P. Johnson, A. D. French and N. L. Allinger (2005). "The external-anomeric torsional effect." *Carbohydr Res* 340(5): 853-62.

Lins, R. D. and P. H. Hünenberger (2005). "A new GROMOS force field for hexopyranose-based carbohydrates." *J Comput Chem* 26(13): 1400-12.

López, C. A., A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger and S. J. Marrink (2009). "Martini Coarse-Grained Force Field: Extension to Carbohydrates." *J Chem Theory Comput*: 269-93.

Lütteke, T., M. Frank and C. W. von der Lieth (2004). "Data mining the protein data bank: automatic detection and assignment of carbohydrate structures." *Carbohydr Res* 339(5): 1015-20.

Lütteke, T. and C. W. von der Lieth (2004). "pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files." *BMC Bioinformatics* 5: 69.

MacKerell, A. D., D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus (1998). "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins" *J Phys Chem B* 102(18): 3586-616.

- Marsh, M. and A. Helenius (2006). "Virus Entry: Open Sesame." *Cell* 124(4): 729-40.
- Mazeau, K. and I. Tvaroška (1992). "PCILO quantum-mechanical relaxed conformational energy map of methyl 4-thio- α -maltoside in solution." *Carbohydr Res* 225(1): 27-41.
- McNamara, J. P., A.-M. Muslim, H. Abdel-Aal, H. Wang, M. Mohr, I. H. Hillier and R. A. Bryce (2004). "Towards a quantum mechanical force field for carbohydrates: a reparametrized semi-empirical MO approach." *Chem Phys Lett* 394(4-6): 429-36.
- Michelle, K., J. W. Brady and J. N. Kevin (2002). "Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations." *J Comput Chem* 23(13): 1236-43.
- Mitchell, T. (1997). *Machine learning*. WCB McGraw Hill.
- Molinero, V. and W. A. Goddard (2004). "M3B: A Coarse Grain Force Field for Molecular Simulations of Malto-Oligosaccharides and Their Water Mixtures." *J Phys Chem B* 108(4): 1414-27.
- Møller, C. and M. S. Plesset (1934). "Note on an approximation treatment for many-electron systems." *Phys Rev* 46(7): 618-22.
- Morgan, W. T. (1950). "Nature and relationships of the specific products of the human blood group and secretor genes." *Nature* 166(4216): 300-2.
- Morris, G. M., D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." *J Comput Chem* 19(14): 1639-62.
- Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson (2009). "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility." *J Comput Chem* 30(16): 2785-91.
- Müllegger, J., M. Jahn, H.-M. Chen, R. A. J. Warren and S. G. Withers (2005). "Engineering of a thioglycoligase: randomized mutagenesis of the acid-base residue leads to the identification of improved catalysts." *Protein Eng Des Sel* 18(1): 33-40.
- Nahmany, A. and F. Strino (2004). *Prediction of Oligosaccharide 3D Structure using Genetic Algorithm Search Methods [M.Sc. thesis]*. Göteborg: Chalmers University of Technology, 2004.
- Nyholm, P.-G., L. A. Mulard, C. E. Miller, T. Lew, R. Olin and C. P. Glaudemans (2001). "Conformation of the O-specific polysaccharide of *Shigella dysenteriae* type 1: molecular modeling shows a helical structure with efficient

exposure of the antigenic determinant α -L-Rhap-(1 \rightarrow 2)- α -D-Galp." *Glycobiology* 11(11): 945-55.

Ogra, Y., K. Ishiwata, H. Takayama, N. Aimi and K. T. Suzuki (2002). "Identification of a novel selenium metabolite, Se-methyl-N-acetylselenohexosamine, in rat urine by high-performance liquid chromatography - inductively coupled plasma mass spectrometry and - electrospray ionization tandem mass spectrometry." *J Chromatogr B Analyt Technol Biomed Life Sci* 767(2): 301-12.

Osborn, H. M. I. and A. Turkson (2009). Sugars as pharmaceuticals. In: H.-J. Gabius (ed). *The sugar code: Fundamentals of glycosciences*. Wiley-VCH:469-83.

Pereira, C. S., R. D. Lins, I. Chandrasekhar, L. C. G. Freitas and P. H. Hünenberger (2004). "Interaction of the Disaccharide Trehalose with a Phospholipid Bilayer: A Molecular Dynamics Study." *Biophys J* 86(4): 2273-85.

Pérez-Castells, J., J. J. Hernández-Gay, R. W. Denton, K. A. Tony, D. R. Mootoo and J. Jiménez-Barbero (2007). "The conformational behaviour and P-selectin inhibition of fluorine-containing sialyl LeX glycomimetics." *Org Biomol Chem* 5(7): 1087-92.

Pérez, S., A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jiménez-Barbero, A. Poveda, J.-F. Espinosa, B. P. van Eyck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenuis, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten and K. Rasmussen (1998). "A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates." *Carbohydr Res* 314(3-4): 141-55.

Pérez, S., N. Mouhous-Riou, N. E. Nifant'ev, Y. E. Tsvetkov, B. Bachet and A. Imberty (1996). "Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide." *Glycobiology* 6(5): 537-42.

Pérez, S. and B. Mulloy (2005). "Prospects for glycoinformatics." *Curr Opin Struct Biol* 15(5): 517-24.

Peri, F., J. Jiménez-Barbero, V. García-Aparicio, I. Tvaroška and F. Nicotra (2004). "Synthesis and Conformational Analysis of Novel N(OCH₃)-linked Disaccharide Analogues." *Chem Eur J* 10(6): 1433-44.

Peters, T., B. Meyer, R. Stuike-Prill, R. Somorjai and J.-R. Brisson (1993). "A Monte Carlo method for conformational analysis of saccharides." *Carbohydr Res* 238: 49-73.

Petrescu, A. J., S. M. Petrescu, R. A. Dwek and M. R. Wormald (1999). "A statistical analysis of N- and O-glycan linkage conformations from crystallographic data." *Glycobiology* 9(4): 343-52.

Pilar, F. L. (2001). *Elementary quantum chemistry*. Dover publications, Inc.

- Ponder, J. W. (2010). "TINKER Version 5.0." *Tinker Software Tools for Molecular Design*.
- Poveda, A., Juan L. Asensio, T. Polat, H. Bazin, Robert J. Linhardt and J. Jiménez-Barbero (2000). "Conformational Behavior of C-Glycosyl Analogues of Sialyl- α -(2 \rightarrow 3)Galactose." *Eur J Org Chem* 2000(9): 1805-13.
- Rarey, M., B. Kramer, T. Lengauer and G. Klebe (1996). "A fast flexible docking method using an incremental construction algorithm." *J Mol Biol* 261(3): 470-89.
- Ravn, V. and E. Dabelsteen (2000). "Tissue distribution of histo-blood group antigens." *APMIS* 108(1): 1-28.
- Rechenberg, I. (1965). *Cybernetic solution path of an experimental problem*. Royal Aircraft Establishment. Library Translation:1122.
- Rosén, J., L. Miguet and S. Pérez (2009). "Shape: automatic conformation prediction of carbohydrates using a genetic algorithm." *J Cheminform* 1(1): 16.
- Rosén, J., A. Robobi and P.-G. Nyholm (2002). "Conformation of the branched O-specific polysaccharide of *Shigella dysenteriae* type 2: molecular mechanics calculations show a compact helical structure exposing an epitope which potentially mimics galabiose." *Carbohydr Res* 337(18): 1633-40.
- Rosén, J., A. Robobi and P.-G. Nyholm (2004). "The conformations of the O-specific polysaccharides of *Shigella dysenteriae* type 4 and *Escherichia coli* O159 studied with molecular mechanics (MM3) filtered systematic search." *Carbohydr Res* 339(5): 961-6.
- Rudnick, E. M., J. H. Patel, G. S. Greenstein and T. M. Niermann (1994). *Sequential circuit test generation in a genetic algorithm framework*. Proceedings of the 31st annual Design Automation Conference. San Diego, California, United States, ACM: 698-704.
- Rydell, G. E. (2009). *Norovirus, causative agent of winter vomiting disease, exploits several histo-blood group glycans for adhesion [Ph.D. thesis]*. Gothenburg: University of Gothenburg, 2009.
- Rye, C. S. and S. G. Withers (2004). "The synthesis of a novel thio-linked disaccharide of chondroitin as a potential inhibitor of polysaccharide lyases." *Carbohydr Res* 339(3): 699-703.
- Saito, M. and I. Okazaki (2009). "Force-field parameters of the and around glycosidic bonds to oxygen and sulfur atoms." *J Comput Chem* 30(16): 2656-65.
- Sampaio Nogueira, C. E., J. R. Ruggiero, P. Sist, P. Cescutti, R. Urbani and R. Rizzo (2005). "Conformational features of cepacian: the exopolysaccharide produced by clinical strains of *Burkholderia cepacia*." *Carbohydr Res* 340(5): 1025-37.

Scott, W. R. P., P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger and W. F. van Gunsteren (1999). "The GROMOS Biomolecular Simulation Program Package." *J Phys Chem A* 103(19): 3596-607.

Sheng, J. and Z. Huang (2008). "Selenium Derivatization of Nucleic Acids for Phase and Structure Determination in Nucleic Acid X-ray Crystallography." *Int J Mol Sci* 9(3): 258-71.

Simons, J. P., B. G. Davis, E. J. Cocinero, D. P. Gamblin and E. C. Stanca-Kaposta (2009). "Conformational change and selectivity in explicitly hydrated carbohydrates." *Tetrahedron: Asymmetry* 20(6-8): 718-22.

Sist, P., P. Cescutti, S. Skerlavaj, R. Urbani, J. H. Leitão, I. Sá-Correia and R. Rizzo (2003). "Macromolecular and solution properties of Cepacian: the exopolysaccharide produced by a strain of *Burkholderia cepacia* isolated from a cystic fibrosis patient." *Carbohydr Res* 338(18): 1861-7.

Sousa, S. F., P. A. Fernandes and M. J. Ramos (2006). "Protein-ligand docking: Current status and future challenges." *Proteins* 65(1): 15-26.

Speert, D. P. (2002). "Advances in *Burkholderia cepacia* complex." *Paediatr Respir Rev* 3(3): 230-5.

Stortz, C. A. (1999). "Disaccharide conformational maps: how adiabatic is an adiabatic map?" *Carbohydr Res* 322(1-2): 77-86.

Stortz, C. A. and A. S. Cerezo (2002). "Disaccharide conformational maps: 3D contours or 2D plots?" *Carbohydr Res* 337(20): 1861-71.

Stortz, C. A., G. P. Johnson, A. D. French and G. I. Csonka (2009). "Comparison of different force fields for the study of disaccharides." *Carbohydr Res* 344(16): 2217-28.

Strino, F., A. Nahmany, J. Rosén, G. J. L. Kemp, I. Sá-correia and P.-G. Nyholm (2005). "Conformation of the exopolysaccharide of *Burkholderia cepacia* predicted with molecular mechanics (MM3) using genetic algorithm search." *Carbohydr Res* 340(5): 1019-24.

Szilágyi, L. and O. Varela (2006). "Non-conventional Glycosidic Linkages: Syntheses and Structures of Thiooligosaccharides and Carbohydrates with Three-bond Glycosidic Connections." *Curr Org Chem* 10: 1745-70.

Tan, M., M. Xia, S. Cao, P. Huang, T. Farkas, J. Meller, R. S. Hegde, X. Li, Z. Rao and X. Jiang (2008). "Elucidation of strain-specific interaction of a GII-4 norovirus with HBGA receptors by site-directed mutagenesis study." *Virology* 379(2): 324-34.

Teneberg, S., B. Alsén, J. Ångström, H. C. Winter and I. J. Goldstein (2003). "Studies on Gal α 3-binding proteins: comparison of the glycosphingolipid

binding specificities of *Marasmius oreades* lectin and *Euonymus europaeus* lectin." *Glycobiology* 13(6): 479-86.

Thøgersen, H., R. U. Lemieux, K. Bock and B. Meyer (1982). "Further justification for the exo-anomeric effect. Conformational analysis based on nuclear magnetic resonance spectroscopy of oligosaccharides." *Can J Chem* 60(1): 44-57.

Tvaroška, I. (1984). "Theoretical study of stereochemistry of methoxy (methylthio) methane as a model of thioacetal segment in thiosaccharides." *Collect Czech Chem Commun* 49(2): 345-54.

Varki, A., R. Cummings, J. Esko, H. Freeze, G. Hart and J. Marth, Eds. (1999). *Essentials of glycobiology*, Cold Spring Harbor Laboratory Press.

Venkatachalam, C. M., X. Jiang, T. Oldfield and M. Waldman (2003). "LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites." *J Mol Graph Model* 21(4): 289-307.

Verez-Bencomo, V., V. Fernández-Santana, E. Hardy, M. E. Toledo, M. C. Rodríguez, L. Heynngnezz, A. Rodriguez, A. Baly, L. Herrera, M. Izquierdo, A. Villar, Y. Valdés, K. Cosme, M. L. Deler, M. Montane, E. Garcia, A. Ramos, A. Aguilar, E. Medina, G. Toraño, I. Sosa, I. Hernandez, R. Martínez, A. Muzachio, A. Carmenates, L. Costa, F. Cardoso, C. Campa, M. Diaz and R. Roy (2004). "A Synthetic Conjugate Polysaccharide Vaccine Against *Haemophilus influenzae* Type b." *Science* 305(5683): 522-5.

Wallace, A. C., R. A. Laskowski and J. M. Thornton (1995). "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions." *Protein Eng Des Sel* 8(2): 127-34.

Walser, P. J., P. W. Haebel, M. Künzler, D. Sargent, U. Kües, M. Aebi and N. Ban (2004). "Structure and functional analysis of the fungal galectin CGL2." *Structure* 12(4): 689-702.

Wang, L. X. (2006). "Toward oligosaccharide- and glycopeptide-based HIV vaccines." *Curr Opin Drug Discov Devel* 9(2): 194-206.

Whitney, C. G., M. M. Farley, J. Hadler, L. H. Harrison, N. M. Bennett, R. Lynfield, A. Reingold, P. R. Cieslak, T. Pilishvili, D. Jackson, R. R. Facklam, J. H. Jorgensen and A. Schuchat (2003). "Decline in Invasive Pneumococcal Disease after the Introduction of Protein-Polysaccharide Conjugate Vaccine." *N Engl J Med* 348(18): 1737-46.

Witczak, Z. J. (1999). "Thio sugars: biological relevance as potential new therapeutics." *Curr Med Chem* 6(2): 165-78.

Witczak, Z. J. and S. Czernecki (1998). "Synthetic applications of selenium-containing sugars." *Adv Carbohydr Chem Biochem* 53: 143-99.

Xia, J., R. P. Daly, F.-C. Chuang, L. Parker, J. H. Jensen and C. J. Margulis (2007a). "Sugar Folding: A Novel Structural Prediction Tool for Oligosaccharides and Polysaccharides 1." *J Chem Theory Comput* 3(4): 1620-8.

Xia, J., R. P. Daly, F.-C. Chuang, L. Parker, J. H. Jensen and C. J. Margulis (2007b). "Sugar Folding: A Novel Structural Prediction Tool for Oligosaccharides and Polysaccharides 2." *J Chem Theory Comput* 3(4): 1629-43.

Xia, J. and C. J. Margulis (2009). "Computational Study of the Conformational Structures of Saccharides in Solution Based on J Couplings and the "Fast Sugar Structure Prediction Software". *Biomacromolecules*: 2370-6.

Xun, L., L. Yan, C. Tiejun, L. Zhihai and W. Renxiao (2009) "Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes." *J Comput Chem*, Doi:10.1002/jcc.21498.