Molecular characterization and optimization of enzymes involved in glycosaminoglycan biosynthesis

J.G.J. (John) Raedts

Thesis committee

Thesis supervisor

Prof. dr. J. van der Oost

Personal Chair at the Laboratory of Microbiology

Wageningen University

Thesis co-supervisor

Dr. S.W.M. Kengen

Assistant Professor at the Laboratory of Microbiology

Wageningen University

Other members

Prof. dr. G. Eggink, Wageningen University

Prof. dr. ir. H. Gruppen, Wageningen University

Prof. dr. W.J.H. van Berkel, Wageningen University

Dr. M.H.M. Eppink, Synthon B.V., Nijmegen

This research was conducted under the auspices of the Graduate School VLAG

Molecular characterization and optimization of enzymes involved in glycosaminoglycan biosynthesis

J.G.J. (John) Raedts

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 21 October 2011
at 1.30 p.m. in the Aula.

J.G.J. (John) Raedts

Molecular characterization and optimization of enzymes involved in glycosaminoglycan biosynthesis (160 pages).

PhD thesis, Wageningen University, The Netherlands (2011)
With propositions and summaries in English and Dutch

ISBN: 978-94-6173-019-0

Contents

Aim and outline

Aim and outline

The research described in this thesis is part of B-Basic project 3.5 entitled "Directed evolution of glycosaminoglycan synthases and glycosaminoglycan modifying enzymes". During this research project we explored alternative ways of controlled chemo-enzymatic synthesis of monodisperse and uniform glycosaminoglycans for pharmaceutical applications, with a main focus on heparin. The D-glucuronyl C5-epimerase was recognized as a key enzyme in heparin biosynthesis. Two alternative approaches were taken to get closer to an industrial applicable enzyme; improvement of the murine (mouse) C5-epimerase and identification and characterization of novel candidate C5-epimerases isolated from prokaryotes.

Chapter 1 is an introduction to the work presented in this thesis. It gives an overview on glycosaminoglycans, mainly concentrating on hyaluronan and heparin. Their biological roles and biosynthesis are discussed, with main focus on the C5-epimerization of D-glucuronic acid into L-iduronic acid in heparin biosynthesis. This reaction is facilitated by heparan sulfate D-glucuronyl C5-epimerase.

Chapter 2 is an overview of prokaryotic candidate D-glucuronyl C5-epimerases. The occurrence of L-iduronic acid in microorganisms is reviewed, and candidate C5-epimerases are identified by *in silico* screening using animal heparan sulfate D-glucuronyl C5-epimerases as a query. Potential physiological implications of these candidate C5-epimerases are discussed.

Chapter 3 describes the cloning, expression and purification of a selection of prokaryotic candidate C5-epimerases using *Escherichia coli* BL21(DE3) as a host. All candidate C5-epimerases could be expressed successfully, some have been purified and used to initiate crystallization trials. The candidate C5-epimerase have been tested for activity using an HPLC-PAD method optimized for hexuronic acid analysis. Hydrolyzed reaction mixture of de-O-sulfated heparin and the candidate C5-epimerase from the bacterium *Bermanella marisrubri* sp. Red65 shows conversion of L-iduronic acid to D-glucuronic acid. This implies the first experimental proof of D-glucuronyl C5-epimerase activity in a prokaryote.

Chapter 4 presents the attempts of the heterologous expression of the murine D-glucuronyl C5-epimerase gene in an industrially applicable expression system. Two different expression hosts have been tried; both expression in *Pichia pastoris* (yeast) as *Escherichia coli*. No protein production has been observed in *P. pastoris*. Gene expression in *E. coli* only results in the production of soluble protein when the enzyme is N-terminally fused to maltose binding protein (MBP). Truncation studies indicate increased solubility upon N-terminal truncations. Activity has been confirmed indication functional heterologous expression of the murine D-glucuronyl C5-epimerase in *E. coli*.

Chapter 5 describes a comparison study of *Pasteurella multosida* hyaluronan synthase (pmHAS) production. This enzyme has been functionally expressed in the past. However, limitations exist in expression levels, protein purification and protein stability. Fusion of this protein to a protein tag possibly helps overcome some of these problems. In this chapter we compare pmHAS expression and activity when fused to an polyhistidine-tag and fused to maltose binding protein (MBP).

Chapter 6 gives an example of protein improvement by "smart library design". We screened a mutant library of *Pyrococcus furiosus* phosphoglucose isomerase (PfPGI) for differences in PGI activity. The mutations are based on Comulator predictions, a computational method to identify correlated amino acid residues. Two conserved and correlated amino acid residues are varied in PfPGI. By enzyme kinetics we indicate a correlation between amino acid prevalence and elevated PGI activity. This higher activity is specific for manganese as cofactor, which might indicate changes in metal affinity.

Chapter 7 summarizes and discusses the results described in the previous chapters of this thesis. Three bottlenecks to overcome are discussed; ways to improve glycosaminoglycan modifying enzyme stability are discussed, suggestions are made on usable methods to assay D-glucuronyl C5-epimerase activity and potential key residues and the protein fold of the heparan sulfate D-glucuronyl C5-epimerases are discussed.

Chapter 1

Introduction

Heparin and heparin-like glycosaminoglycans

Part of this chapter has been published in:

Raedts J, Kengen SWM, van der Oost J. (2011). Occurrence of L-iduronic acid and putative D-glucuronyl C5-epimerases in prokaryotes. *Glycoconjugate Journal* **28**(2), 57-66.

Introduction

Carbohydrates are key building blocks ubiquitously found in all forms of life. The self-explanatory name already suggests the chemical nature of these molecules. Being hydrated carbon based molecules, the structure generally exists of carbon, hydrogen and oxygen atoms. Although all carbohydrates are almost exclusively composed of these elementary atoms, an enormous variety exists. Most of this diversity reflects the combination of distinct building blocks (monosaccharides), leading to long linear or branched polysaccharides.

Many different polysaccharide structures exist. Glycosaminoglycans (GAGs) are a highly abundant type of biological active polysaccharides predominantly found in animals. Glycosaminoglycans are long linear anionic hetero-polysaccharide chains, also referred to as mucopolysaccharides, that consist of repeating disaccharide units consisting of a hexuronic acid and a hexosamine. The latter is either a D-glucosamine (GlcN) or a D-galactosamine (GalN). The hexuronic acid is a D-glucuronic acid (GlcA) or its C5-epimer L-iduronic acid (IdoA). The GAG keratan sulfate is an exception, it has the hexose D-galactose (Gal) instead of a hexuronic acid. The disaccharide composition differs per type of GAG, and can vary both in sugar content and glycosidic linkage. The main classification of all described GAGs is based on these differences. The most common glycosaminoglycans are chondroitin sulfate, dermatan sulfate, heparan sulfate, heparin, hyaluronan and keratan sulfate (Table 1).

Table 1 – Overview of most common glycosaminoglycans

Glycosaminoglycan	Main disaccharide	Sulfation pattern	Localization
Chondroitin sulfate	-4)-GlcA-β(1,3)-GalNAc-β(1-	GalNAc at C4 / C6, GlcA at C2	cartilage, bone, heart valves
Dermatan sulfate	-4)-IdoA-α(1,3)-GalNAc-β(1-	GalNAc at C4 / C6, IdoA at C2	skin, blood vessels, heart valves
Heparan sulfate	-4)-GlcA-β(1,4)-GlcNAc-α(1-	GlcA(IdoA) at C2, GlcNAc(NS) at C6 / C3	cell surface membranes, ECM
Heparin	-4)-IdoA-α(1,4)-GlcNS-α(1-	like heparan sulfate but heavier sulfated	mast cells (e.g. lungs, liver, skin)
Hyaluronan	-4)-GlcA-β(1,3)-GlcNAc-β(1-	completely unsulfated chain	synovial fluid, eye, ECM, skin
Keratan sulfate	-3)-Gal-β(1,4)-GlcNAc-β(1-	Both Gal and GlcNAc at C6	cornea, bone, cartilage

Glycosaminoglycans are often sulfated at various positions, resulting in another differentiating feature apart from differences in disaccharide composition. The fact that this sulfation pattern can vary within any given type of GAG, implies that the heterogeneity of GAGs is enormous. Additional consequence of these sulfate additions, is the highly negative charge of glycosaminoglycans. Together with the carboxyl group of the hexuronic acids, this results in the GAG heparin being the most negatively charged biological macromolecule known (Lindahl, 1997).

Most glycosaminoglycans can be found in large quantities in specific cell structures called proteoglycans. These large structures consist of a linear core protein with many glycosaminoglycans covalently linked to it. Each glycosaminoglycan is linked to the core protein via a Gal-Gal-Xyl sugar linker, where the xylose (Xyl) is attached to a serine residue of the core protein and the glycosaminoglycan is linked to the galactose (Gal) residue. Also proteoglycans are highly diverse; variations occur both in the type of core protein and in type, size and number of glycosaminoglycans.

Overview common heparin-like glycosaminoglycans

A wide variety of glycosaminoglycans have been characterized to date. In this thesis the focus is on [GlcNAc – GlcA/IdoA] containing GAGs. An overview of these glycosaminoglycan structures can been found in Figure 1. A more detailed description of each individual glycosaminoglycan is given in this paragraph.

Heparin | The first report of heparin was nearly a century ago (Howell and Holt, 1918). It took another twenty years before the relevance of heparin as a pharmaceutical application became clear, when in 1935 heparin was first described as a clinical anticoagulant (Linhardt, 1991). Even though an application was found, it lasted until the 1960s before L-iduronic acid was identified as the major uronic acid residue (Cifonelli and Dorfman, 1962; Perlin *et al*., 1968). Subsequently, a general structure of heparin could be elucidated, and more insight in the molecular mechanism behind the anticoagulation effect was obtained.

Heparin is widely distributed among animals, including mammals, birds, fish. Also from invertebrates like insects, worms, lobsters and clams, heparin-like structures have been isolated (Linhardt and Toida, 1997; Luppi *et al*., 2005; Im *et al*., 2010). Also these heparin-like

compounds can have anticoagulant activities, even though not all these organisms have a blood coagulation system (Nader *et al*., 1999; Medeiros *et al*., 2000). To date (clinical) heparin is still most efficiently obtained by isolation as a polydisperse mixture from animal tissue that is rich in mast cells. Initially a frequently used source was liver (*hepatos* in Latin), resulting in the name heparin. Nowadays, heparin is isolated from highly vascularized tissue derived as side-product from livestock. Examples of such tissue include (porcine) intestine tissue and (bovine) lung tissue, although the latter is banned from US and European markets since the outbreak of bovine spongiform encephalopathy (BSE) in the 1990's. The extraction of heparin requires rather large quantities of animal tissue; one kg of tissue typically results in as little as 150 mg of isolated heparin (Alban, 1997).

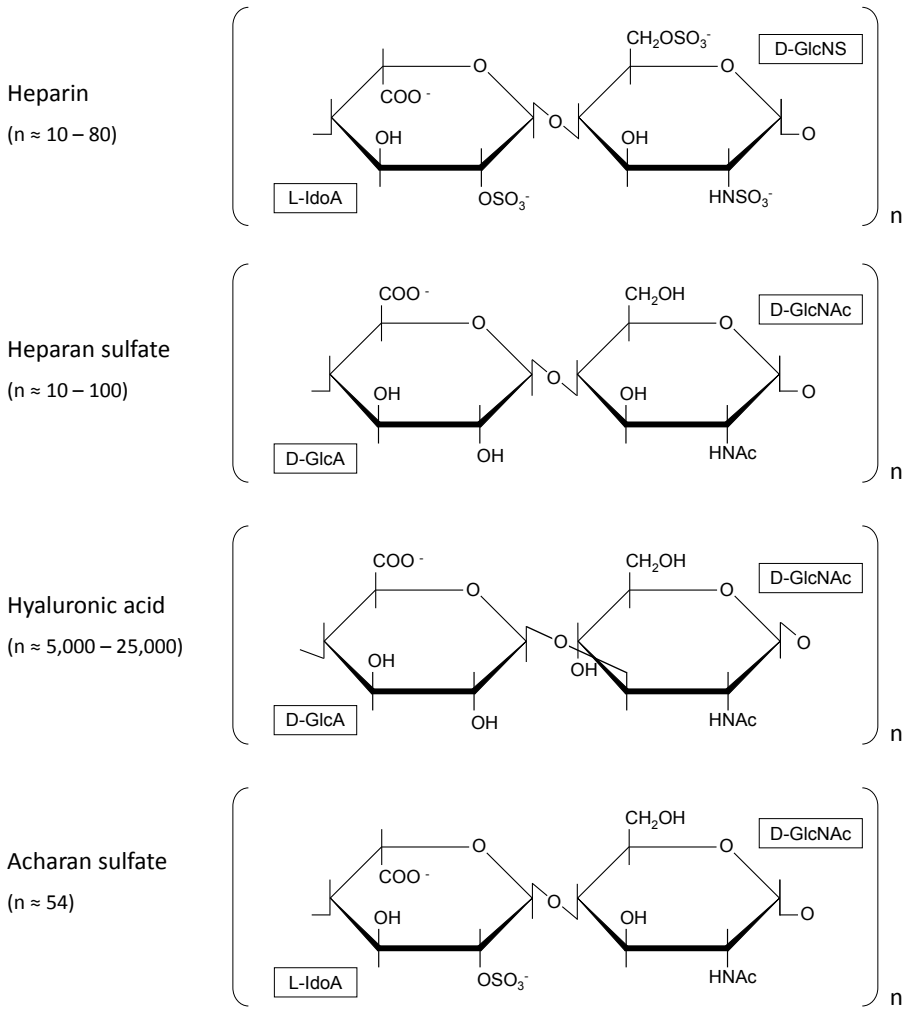


Figure 1 – Chemical structure of the repetitive disaccharide unit of the described glycosaminoglycans.

Many proteins have been described that can bind heparin. Binding is mainly due to the large number of negatively charged residues present in these GAGs, which highly contribute to electrostatic interactions based binding. Two well-studied examples of heparin protein interactions are fibroblast growth factor 2 (FGF-2) and antithrombin. For both proteins the minimal binding structure is a pentasaccharide, although both pentasaccharides have a dissimilar chain modification motif (Fig.2). Binding of heparin by FGF-2 requires an N-sulfated-D-glucosamine and an L-iduronic acid that is 2-O-sulfated (Lindahl, 1997). Structural data of heparin bound to basic fibroblast growth factor is available and amino acid residues involved in heparin binding have been characterized. However these residues are found not to be conserved throughout the FGF family (Faham *et al*., 1996).

Structural data also is available of antithrombin-heparin interactions (Lindahl *et al*., 1979). The pentasaccharide sequence required for binding (Thunberg *et al*., 1982) is depicted in Figure 2. Binding of heparin by the serine protease inhibitor antithrombin leads to inactivation of several members of the coagulation cascade including thrombin (Factor II) and Factor Xa (Bourin and Lindahl, 1993). The coagulation activity of thrombin, which is a serine protease, is based on the conversion of soluble fibrinogen into insoluble strands of fibrin. Thrombin can be released from prothrombin by prothrombinase (Factor Xa), the activated form of the serine endopeptidase Factor X.

Heparan sulfate | A glycosaminoglycan highly similar to heparin is heparan sulfate. For some time heparan sulfate was mainly an unwanted by-product of heparin purification. However, as more insight in its biological distribution and roles became known, heparan sulfate is now recognized as a large group within the family of glycosaminoglycans. Consequently, heparin now is regarded as just one of the members within the family of heparan sulfate (Casu and Lindahl, 2001).

Heparan sulfate is structurally closely related to heparin. Similar to heparin, the polysaccharide backbone consists of repetitive dimers of D-glucosamine 1-4 linked to D-glucuronic acid (Fig.1). The number of modifications on the backbone, however, is less than for heparin, especially the degree of sulfation and C5-epimerization. Heparan sulfate generally has less than one sulfate group per disaccharide, while heparin on average has three (Linhardt *et al*., 1992). The same holds true for the hexuronic acid content; while heparan sulfate predominantly consists of D-glucuronic acid, hexuronic acid residues in

heparin are up to 90% L-iduronic acid. Also the chain length of heparan sulfate is typically longer and more polydisperse than that of heparin (Griffin *et al*., 1995).

Heparan sulfate is believed to be ubiquitously present in the animal kingdom, with an exception of sponges (Nader *et al*., 1999; Medeiros *et al*., 2000). Although structurally related to heparin, it differs in topology and biological roles (Conrad, 1998). Heparan sulfate is located at the external surface of cell membranes and the extracellular matrix (ECM) as part of a proteoglycan. It has a role in multiple specific cell-cell and cell-protein interactions. Heparan sulfate is involved in many processes, mostly related to regulation of cell growth and development, proliferation and cell adhesion (Wight *et al*., 1992; Perrimon and Bernfield, 2000). Numerous compounds have been found to interact with heparan sulfate proteoglycans, e.g. signaling molecules like growth factors, cytokines, morphogens, enzymes and extracellular matrix proteins. These binding properties result in involvement of heparan sulfate in diseases like viral invasions e.g. herpes simplex virus 1 (Shukla *et al*., 1999). Tumor metastasis involves the deregulated release of heparan sulfate due to enhanced heparanase activity (Hulett *et al*., 1999; Vlodavsky *et al*., 1999). Another disorder based on deregulation of glycosaminoglycan-acting enzymes is mucopolysaccharidosis (MPS), which is the inability to degrade glycosaminoglycans due to an inherited mutation, resulting in accumulation of glycosaminoglycans in lysosomes. As much as ten different forms are described, two of which are a result of deficiencies in α-L-iduronic acid acting enzymes; a mutation in the α-L-iduronidase gene results in MPS I (Hurler syndrome) while MPS II (Hunter syndrome) is the result of a mutation in the iduronate sulfatase gene.

Hyaluronan | Hyaluronan, also referred to as hyaluronic acid, is a glycosaminoglycan with a rather simple structure. Although the polymer consists of the same building blocks as heparin and heparin sulfate, it differs in that the disaccharides are connected via different glycosidic bonds. In addition hyaluronic acid is the only GAG that is completely unsulfated. The polymer generally is a very long molecule, that is built from as many as 2,000-25,000 disaccharide repeats, resulting in molecules of up to 10^7 Da (Chong *et al*., 2005). The polymer consists of D-glucuronic acid β(1-3) linked to N-acetyl-D-glucosamine β(1-4) (Fig.1) (Weissmann and Meyer, 1954). Like all GAGs, hyaluronic acid has a large number of anionic residues, resulting in a large water binding capacity. Even at low concentration it has viscous hydrated gel forming properties. Even though the primary structure is rather simple, a wide

variety of secondary structures can occur, resulting in different biological functions. It is believed that high molecular weight hyaluronan molecules adopt an anti-parallel β-sheet like tertiary structure, stabilized by specific H-bonds and hydrophobic bonds (Scott and Heatley, 1999). Low molecular weight molecules do not adopt this structure.

Hyaluronic acid is involved in various biological processes, and can be found in various kinds of tissue (Laurent *et al*., 1996; George, 1998). Hyaluronan plays important roles in the extracellular matrix (ECM). For example, hyaluronic acid is found in abundance in the ECM of load-bearing joints. Hyaluronan absorbs significant amounts of water providing a turgor pressure, crucial for its structural role in connective tissue (Fraser *et al*., 1997). This water binding capacity also is crucial in the hydration balance. Additionally, hyaluronan has a regulatory effect since it can be bound by various cell surface receptors. It has a role in various stages of cell growth including embryonic development, cell migration, cell proliferation, inflammations, healing processes and tumor growth (Sherman *et al*., 1994).

The first isolation of hyaluronan was from the vitreous body of the eye, *hyaloid* in Greek, hence the name hyaluronan (Meyer and Palmer, 1934). Other tissues in vertebrates being rich in hyaluronic acid include the umbilical cord, the synovial fluid in joints and the dermis. The highest concentration of hyaluronic acid reported in animal tissue, is in rooster combs, making it a commonly used source for industrial hyaluronan isolation.

Acharan | Acharan sulfate is a non-canonical glycosaminoglycan, whose structure has high resemblance with heparin. This polysaccharide has been isolated from the giant African snail *Achtina fulica* (Kim *et al.,* 1996). The major repeating disaccharide structure is determined (Fig.1) as [→4)-α-D-GlcNAc-(1→4)-α-L-IdoA2S-(1→], although to a lesser extent also unsulfated disaccharides can be present (Kim *et al.,* 1998). The hexuronic acid in acharan sulfate is 100% L-iduronic acid (Chi *et al.,* 2006). However, the most remarkable difference compared to heparin, is the acetylation of the D-glucosamine residues. In heparin biosynthesis C5-epimerization of D-glucuronic acid is only possible when the GlcNAc is de-N-acetylated and subsequently re-N-sulfated. The acetylated amine group in acharan sulfate might indicate either a heparan sulfate D-glucuronyl C5-epimerase that is less substrate specific or an additional and alternative D-glucuronyl C5-epimerase. Unfortunately no genome data is available for this species to date, and therefore putative differences in the C5-epimerases remain unclear.

Glycosaminoglycan biosynthesis

The variation in glycosaminoglycan chain complexity is a result of essential differences in their biosynthesis. While the biosynthesis of hyaluronan is restricted to only a few steps, the biosynthesis of heparin / heparan sulfate is a rather complex multistep process. An overview of the biosynthesis of both types of glycosaminoglycans is given below.

Biosynthesis of hyaluronic acid

Hyaluronan is synthesized in the plasma membrane by one specific glycosyltransferase called hyaluronan synthase (HAS). This enzyme is an integral membrane protein containing multiple membrane domains present throughout the protein sequence. One single glycosyltransferase domain (GT2) is responsible for the transfer of the two different UDP-sugars; UDP-GlcNAc and UDP-GlcA (Weigel *et al*., 1997). These UDP-sugars are placed either at the reducing end (e.g. in mouse; MmHAS and human; HsHAS) or at the non-reducing end (e.g. frog; XlHAS) of the growing hyaluronan molecule (Wiegel and DeAngelis, 2007). The hyaluronan chain polymerizes on the cytoplasmic side of the plasma membrane. In vertebrates the chain growth is in a processive way. This means that upon release from the hyaluronan synthase, the growing hyaluronan chain cannot be further elongated. In vertebrates as much as three isoforms (HAS1, HAS2, HAS3) of the hyaluronan synthase exist; in both human and mouse these have been identified and characterized (Spicer *et al*., 1997). All three isoforms have HAS activity, but differ in the chain length of the hyaluronan molecules that they produce and the ease with which produced hyaluronan molecules are released from the cell surface (Itano and Kimata, 2002).

Biosynthesis of heparin and heparan sulphate

Unlike hyaluronan biosynthesis, the biosynthesis of heparin and heparan sulfate is a multistep process. The wide variation in the different heparan sulfate sequences is caused by the many enzymes involved in the biosynthesis of these glycosaminoglycans (Lindahl *et al*., 1998). Over thirty different enzymes are involved, including multiple glycosyltransferases, multiple sulfotransferases and a single C5-epimerase (Esko and Selleck, 2002). Most of these enzymes are so-called type II transmembrane proteins and are believed to be localized in the Golgi apparatus. The whole biosynthesis can be divided into three parts: (1) chain initiation,

(2) chain polymerization and (3) chain modification. A schematic representation is provided in figure 2 (Sasisekharan and Venkataraman, 2000; Esko and Lindahl, 2001; Habuchi *et al*., 2004; Mizumoto *et al*., 2005; Lindahl and Li; 2009).

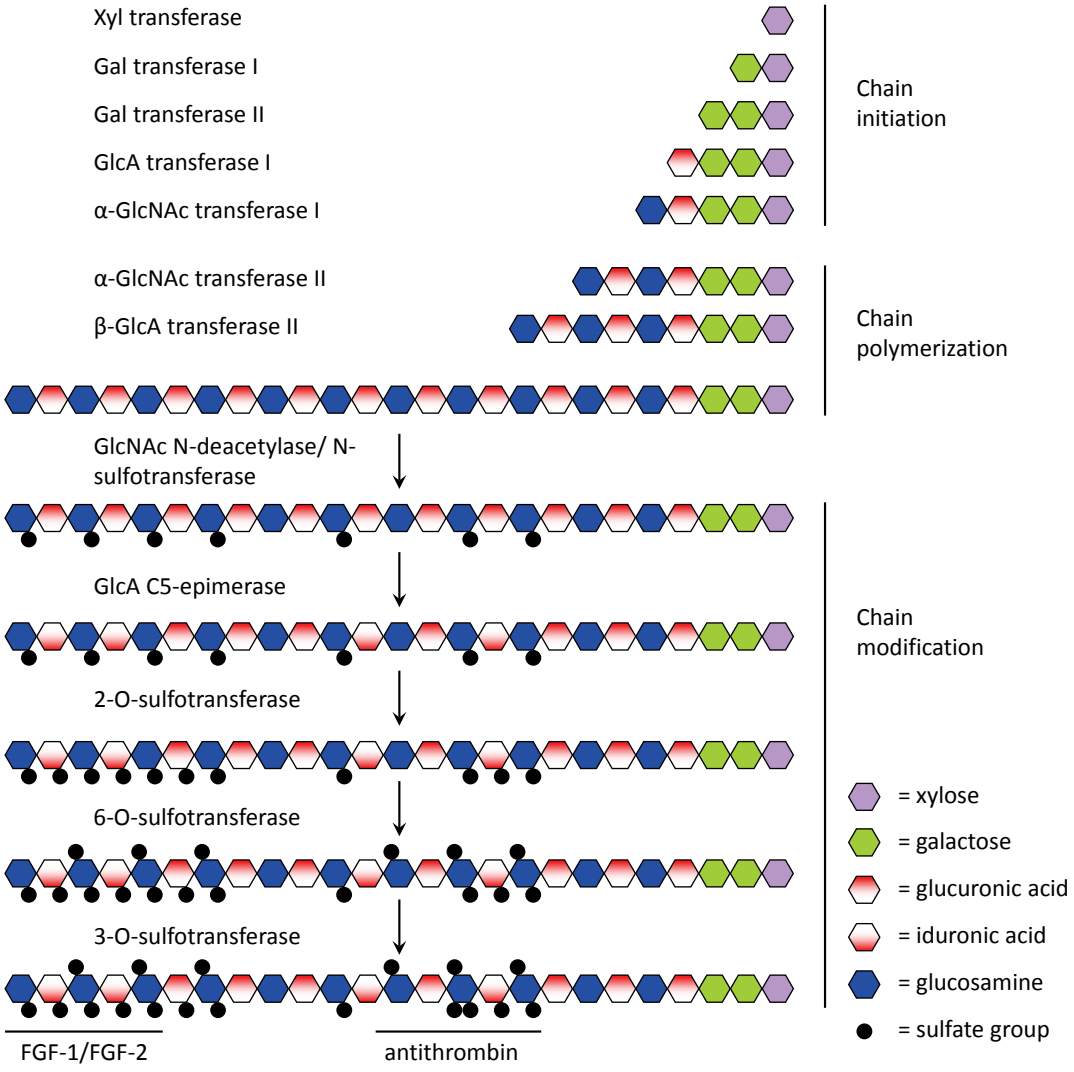


Figure 2 – Schematic representation of heparin / heparan sulfate biosynthesis. The epimerization step is the second step in the chain modification, and only possible upon the N-deacetylation / N-sulfation step of the N-acetyl-D-glucosamine residue attached at the C4 of that specific D-glucuronic acid residue (Jacobsson *et al*., 1984). Subsequent O-sulfatation steps of the hexuronic acid or flanking D-glucosamine residues results in the protection of the hexuronic acid residue for further epimerization. Picture based on (Esko and Lindahl, 2001).

Chain initiation | Most glycosaminoglycans contain the common GAG-protein linkage tetrasaccharide D-GlcA-β(1→3) – D-Gal-β(1→3) – D-Gal-β(1→4) – D-Xyl- β(1→*O*) – Ser. The chain initiation of the glycosaminoglycan starts with the xylosylation of a Serine (Ser) residue

of the core protein, catalyzed by a Xylosyltransferase (XylT). Subsequently, two galactosyltransferases (Gal-I and Gal-II) are responsible for the attachment of two D-galactose-residues (Gal) to the xylose-residue (Xyl). The linkage tetrasaccharide is completed after addition of a D-glucuronic acid (GlcA) residue to the Gal-residue, catalyzed by a glucuronyltransferase (GlcAT-I). All sugars are added at the non-reducing end of the chain, and originate from UDP-sugar, with UDP being released upon sugar-transfer.

Chain polymerization | The polymerization of heparin / heparan sulfate is mediated by a hetero-oligomeric complex of two glycosyltransferases; EXT1 and EXT2 (McCormick *et al*., 2000). *In vivo* this complex is localized in the Golgi apparatus (McCormick *et al*., 1998; Lind *et al*., 1998). It is responsible for the polymerization by alternating addition of D-glucuronic acid (GlcA) and N-acetyl-D-glucosamine (GlcNAc) linked by a 1→4 glycosidic bond. The residues are added to the non-reducing termini of the growing chain, leading to a homogeneous non-sulfated glycosaminoglycan called heparosan.

Chain modification | The polymerization reactions result in the heparin backbone (heparosan), a uniformly repetitively polysaccharide of GlcNAc and GlcA. To obtain heparin and heparan sulfate, up to six different modifications of the heparosan molecule are required.

The first two modifications are catalyzed by a single enzyme. The bifunctional enzyme N-deacetylase / N-sulfotransferase (NDST) is responsible for replacement of most N-acetyl groups of the GlcNAc residues by N-sulfated groups (Kjellén, 2003). As much as four different isoforms have been described in human and mouse (Aikawa *et al*., 2001). Each isoform has its unique preference in the acetylation : sulfation ratio of the GlcNAc residue. This difference contributes to the large variation in N-acetylation / N-sulfation patterns of D-glucosamine in heparin and heparin sulfate. N-sulfation is essential determining the overall extent of modification of the polymer chain, as some of the subsequent modification steps recognize only N-sulfated glucosamine.

The next modification step is the conversion of D-glucuronic acid (GlcA) residues into L-iduronic acid (IdoA) catalyzed by a D-glucuronyl C5-epimerase. In mammals only one single isoform of this enzyme is present. It has been demonstrated that GlcA is recognized as a substrate, when it is linked at C-1 to an N-acetyl-D-glucosamine and at C-4 to an N-sulfated-

D-glucosamine (Lindahl *et al*., 1984). The reverse structure is not recognized as a substrate, but a GlcA residue between two N-sulfated residues is recognized as substrate as well. Different than other described heparin modifying enzymes, the C5-epimerization of D-glucuronic acid in heparin biosynthesis is the only step that cannot be done chemically to date and therefore completely depends on enzymatic catalysis (Li, 2010).

The last three polymer modifying steps are all O-sulfation steps, catalyzed by three different O-sulfotransferases. All these sulfotransferases use adenosine 3'-phosphate 5'-phosphosulfate (PAPS) as a sulfate donor. The first O-sulfation step involves the sulfation of the C2 of the hexuronic acid. The responsible enzyme uronosyl-2-O-sulfotransferase (2-OST) both reacts on GlcA and IdoA, although it prefers the latter (Rong *et al*., 2001). GlcA is only sulfated occasionally while IdoA is sulfated ranging from 50 to 90% (Safaiyan *et al*., 2000). The 2-OST is likely to be physically associated to the C5-epimerase *in vivo*. Both enzymes co-localize in the Golgi-apparatus (Pinhal *et al*., 2001). Enzyme stability and activity of both, is enhanced by this physical association. However, the subsequent 2-O-sulfation is not a prerequisite for the formation of IdoA residues, as the C5-epimerase is active without 2-OST.

The other two sulfation steps both act on the D-glucosamine-residues. A D-glucosamine-6-O-sulfotransferase is responsible for sulfation at C-6 of a GlcN, while the C-3 of GlcN can be sulfated by a D-glucosamine-3-O-sulfotransferase. In contrast to 2-OST, there are multiple isoforms identified of both 6-OST (three isoforms) and 3-OST (at least five isoforms). These isoforms are expressed in a tissue and developmental stage dependent way (Habuchi *et al*., 1999; Shworak *et al*., 1999). In parallel to the C5-epimerization step, also the 2-O-sulfation and 6-O-sulfation steps occur frequently, although sulfation-levels are tissue specific (Safaiyan *et al*., 2000). The number of 3-O-sulfated residues are relatively low in heparin sulfate / heparin.

Hexuronic acid C5-epimerization

The previous paragraph describes all reaction steps involved in heparin / heparan sulfate biosynthesis. The C5-epimerization of hexuronic acids, being part of a glycosaminoglycan, cannot be performed chemically and therefore is considered as a bottleneck in controlled *in vitro* synthesis of heparin. The next paragraph focuses on this specific reaction and the involved enzyme. Three distinct types of C5-epimerases have been described to be able to

catalyze the C5-epimerization of hexuronic acids (Valla *et al*., 2001); one polymannuronic acid C5-epimerase and two distinct D-glucuronyl-C5-epimerases (Fig.3). Each of these enzymes has a very limited substrate acceptance and only acts on specific polysaccharides. There is no apparent sequence homology between these enzymes.

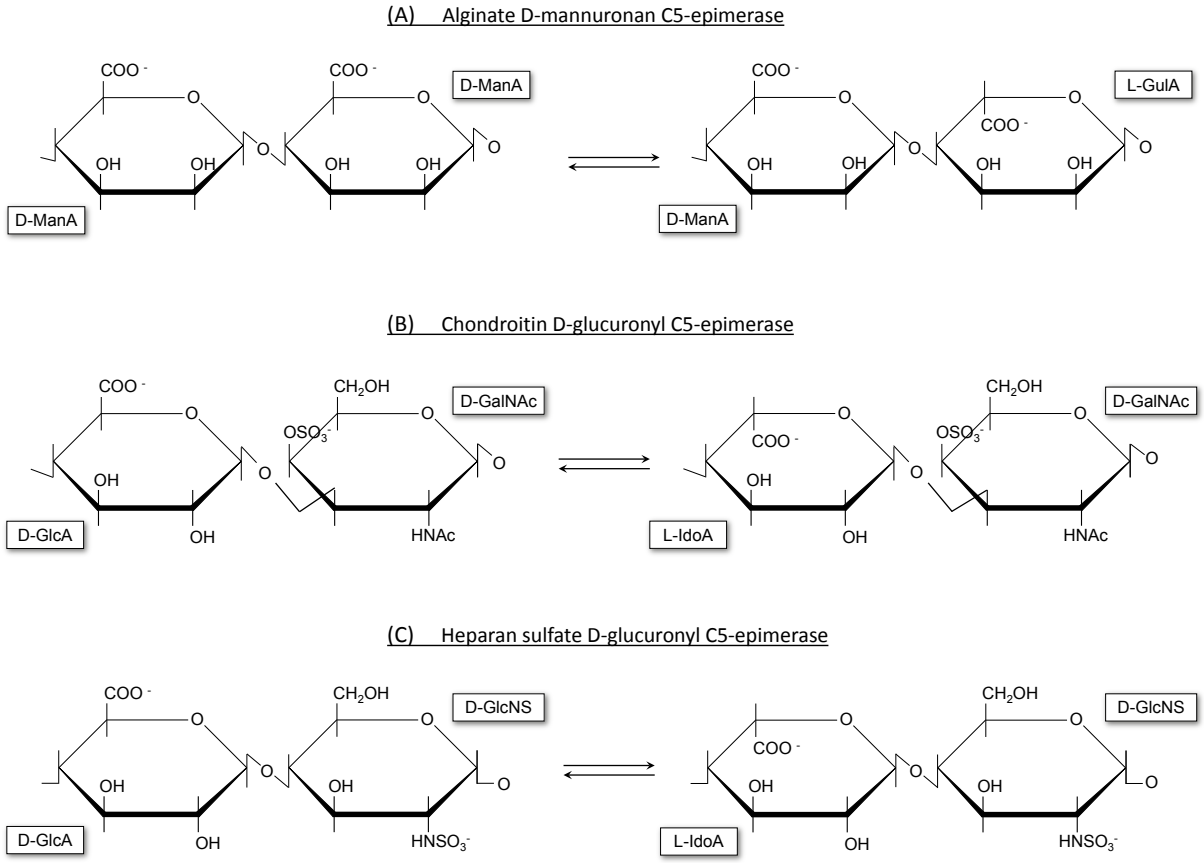


Figure 3 – Schematic representation of reactions catalyzed by the three described hexuronic acid C5-epimerases. (A) Alginate D-mannuronan C5-epimerases catalyze the conversion of D-mannuronic acid into L-guluronic acid. (B) Chondroitin D-glucuronyl C5-epimerase converts D-glucuronic acid into its C5-epimer L-iduronic acid. (C) Heparan sulfate D-glucuronyl C5-epimerase mediated catalysis also results in L-iduronic acid formation. The epimerase reaction is expected to occur via carbanion intermediate by abstraction and re-addition of the C5 proton. Water is used as the proton acceptor / donor.

Polymannuronic acid C5-epimerases

Polymannuronic acid C5-epimerases do not act on D-glucuronic acid in a glycosaminoglycan but on D-mannuronic acid in alginate. These C5-epimerases are the best described hexuronic acid C5-epimerases, involved in the biosynthesis of alginate. Alginate is an unbranched linear polysaccharide consisting of (1-4) linked β-D-mannuronic acid residues, some of these D-

mannuronic acid residues are converted to their C-5 epimer L-guluronic acid by a polymannuronic acid C5-epimerase (Fig.3A). The ratio between these two hexuronic acids are variable and depend on the source of isolation. Discovered in the late 19th century alginate was originally extracted from brown seaweed. Nowadays it is also identified in the two bacterial genera *Pseudomonas* (Franklin *et al*., 1994) and *Azotobacter* (Haug and Larsen, 1970). Alginate has a high water binding capacity, which makes it a widely used product with multiple industrial applications.

Much research has been done on the bacterial polymannuronic acid C5-epimerases (Remminghorst and Rehm, 2006). All genes involved in the alginate biosynthesis are clustered in an operon, both in *P. aeruginosa* and *A. vinelandii*. The corresponding enzymes reside in the periplasm including the polymannuronan C5-epimerase (AlgG). However *A. vinelandii* has a set of seven additional polymannuronan C5-epimerases (AlgE1-7) that are secreted by the cell and whose genes are not located in the alginate biosynthesis gene cluster (Ertesvåg *et al*., 1999). Alg1-7 have a Ca^{2+} dependency and do not share similarity at amino acid sequence level with AlgG. Most likely these additional epimerases are crucial for the high variability in the epimerization pattern.

The ease of working with these extracellular alginate C5-epimerases has led to extensive research of these enzymes. It has been shown that all seven extracellular alginate C5-epimerases are composed of varying amounts of two types of structural modules. Each type of C5-epimerase has one or two copies of the catalytically active A-module, which is about 385 amino acids long. In addition up to seven copies of the much smaller R-module (155 amino acids) are present per enzyme. A crystal structure is available for the smallest epimerase AlgE4, which has one single A-module and one single R-module. The A-module alone is both catalytically active, and can bind calcium (Ertesvåg and Valla, 1999). X-ray crystallography data (2.1-Å resolution) have revealed a right-handed β-helix with 4 parallel sheets, where the active site is positioned in a highly basic groove (Rozeboom *et al*., 2008). The R-module has a right-handed parallel β-roll subunit (NMR data) that is likely to be involved in additional calcium binding. Linkage of a R-module to an A-module enhances overall activity substantially (Aachmann *et al*., 2005).

D-glucuronyl C5-epimerases

The other two hexuronic acid C5-epimerases both act on β-D-glucuronic acid linked to a hexosamine in a polymer. In either case α-L-iduronic acid is the reaction product. However although the result of the reaction of both enzymes is identical, no significant sequence homology exists between them at amino acid level. Also biochemically both epimerases are dissimilar, since they differ in pH optimum and the requirement for cations (Malmström and Åberg, 1982). Both enzymes can only accept their natural substrate.

Dermatan sulfate D-glucuronyl C5-epimerase | The dermatan sulfate D-glucuronyl C5-epimerase (EC 5.1.3.19) converts D-glucuronic acid to L-iduronic acid (Fig.3B) in chondroitin biosynthesis (Malmström and Fransson, 1975). The gene responsible for the expression of this C5-epimerase has been identified only a few years ago, and was renamed from its original name SART2 (squamous cell carcinoma antigen recognized by T-cell 2) (Maccarana *et al*., 2006). Different than the heparosan D-glucuronyl C5-epimerase, the chondroitin D-glucuronyl C5-epimerase requires a divalent metal ion, with manganese as the most effective one. Based on several bacterial lyases a structural model of the epimerase has been proposed. The catalytic site is expected to comprehend three putative catalytic residues, two Histidines (His-205 and His-450) and one Tyrosine (Tyr-261). Site directed mutagenesis of any of these conserved residues results in complete loss of epimerase activity (Pacheco *et al*., 2008).

Heparan sulfate D-glucuronyl C5-epimerases | Also the heparan sulfate C5-epimerase (EC 5.1.3.12) catalyzes the conversion of D-glucuronic acid into L-iduronic acid, however it only accepts N-sulfated heparosan as substrate (Fig.3C). Unlike the other two hexuronic acid C5-epimerases, no structural data is available of heparosan C5-epimerases, nor has the catalytic site been identified. However, a reaction mechanism is proposed based on activity measurements. The C5-epimerase reaction occurs via a carbanion intermediate by abstraction and re-addition of the C5-proton derived from water (Hagner-McWhirter, 2000). The re-addition can occur at both sides of the pyranose ring, in such a way that it can give an inversion of configuration; the carboxy group shifts across the plane of the pyranose ring. *In vitro* the reaction is freely reversible, with a slight preference for D-glucuronic acid over L-iduronic acid. However, *in vivo* the C5-epimerization step is irreversible; possibly because the

L-iduronic acid is blocked for the C5-epimerase by the subsequent 2-O-sulfation step (Hagner-McWhirter, 2004).

Although the exact catalytic mechanism still is unclear, the biological importance of D-glucuronyl C5-epimerase activity in early vertebrate development has been shown in previous research. Animals lacking D-glucuronyl C5-epimerase activity either die perinatally or have very early embryonic patterning defects. In zebra fish it has been shown that dorso-ventral axis formation and patterning is highly affected in the absence of D-glucuronyl C5-epimerase activity (Ghiselli and Farber, 2005). Similar findings have been reported for axonal development of neurons in a null allele mutant of the nematode *Caenorhabditis elegans* (Bülow and Hobert, 2004). Also the development of some organs requires D-glucuronyl C5-epimerase activity. Mouse mutants having a disrupted D-glucuronyl C5-epimerase have also been demonstrated to lack L-iduronic acid, resulting in a neonatal lethality with malformations in kidney, lung and skeletal development (Li *et al*., 2003; Jia *et al*., 2009).

The catalysis of D-glucuronic acid into L-iduronic acid is typically mediated by a single epimerase. As mentioned, an exception is reported in zebra fish, where two genes have been identified (*glce-A* and *glce-B*). Both genes are expressed, and the gene products both are homologous to the D-glucuronyl C5-epimerase gene present in mammals, like human (Ghiselli and Farber, 2005). Examination of the genome data of *Branchiostoma floridae* (Florida lancelet fish) reveals another example of two distinct heparan sulfate D-glucuronyl C5-epimerase genes present in one single organism. Examination of other (fish) genomes did not reveal other examples, indicating that such recent gene duplication is not a general feature in fish.

L-iduronic acid

As described in the previous paragraph the lack of a heparan sulfate D-glucuronyl C5-epimerase results in severe disturbance in early embryonic patterning and morphogenesis, stressing the importance of L-iduronic acid as a substituent in heparin. Much research has been done to study this rather unusual hexuronic acid.

The first report of an L-iduronic acid containing glycosaminoglycan, was by Hoffman and co-workers. They showed its presence in dermatan sulphate (Hoffman *et al*., 1956). Initially the L-iduronic acid was considered to be incorporated in heparin biosynthesis as a

UDP-sugar. Jacobson & Davidson reported catalytic activity of rabbit skin extracts, in which L-iduronic acid seemed to be formed from UDP-D-glucuronic acid (Jacobson and Davidson, 1963). However, later research demonstrated that L-iduronic acid is formed via C5-epimerization of D-glucuronic acid, only after UDP-D-glucuronic acid is incorporated in the polymer. This process is identical for the biosynthesis of both dermatan sulfate (Malmström *et al*., 1975) and heparin (Lindahl *et al*., 1972; Höök *et al*., 1974). UDP-L-iduronic acid has never been isolated from any natural source and is considered as an unnatural sugar.

As a C5-epimer of D-glucuronic acid, L-iduronic acid only differs with respect to the orientation of the carboxyl group at the C5 position. Even though this might seem only a small change, glycosaminoglycans that contain L-iduronic acid generally do have stronger binding properties resulting in elevated biological activities, compared to structures that only contain D-glucuronic acid. The reason for this can be found in the nature of both hexuronic acids. While D-glucuronic acid is restricted to only a $^4C_1$ conformation, L-iduronic acid has three stable conformations and can occur additionally also in the other boat conformation $^1C_4$ and in a skew-boat conformation $^2S_0$ (Fig.4).

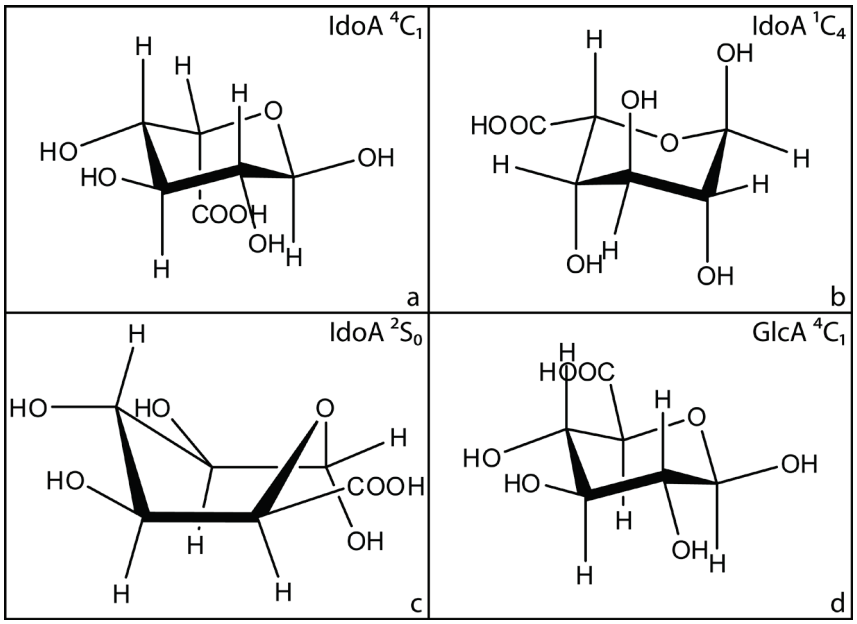


Figure 4 – Structural conformation of D-glucuronic acid (GlcA), and its C5-epimer L-iduronic acid (IdoA). Three stable conformations are possible for L-iduronic acid; the most stable form is $^4C_1$ with a calculated energy of 67.66 kcal/mole (a). The other boat conformation (b) ($^1C_4$; 73.80 kcal/mole) and a skew boat (c) ($^2S_0$; 76.66 kcal/mole). The theoretical energy value of $^1C_4$ D-glucuronic acid (d) is 63.97 kcal/mol. Theoretical energy values are calculated by Chembio 3D.

Lower calculated energy values correspond to less tension on the ring, and thus result in a more stable conformation. Theoretically, the most stable conformer is the most abundant one. All three conformers of L-iduronic acid are rather stable and therefore do occur. However, the occurrence highly depends on the position and modifications of L-iduronic acid in the polymer (Ferro *et al*., 1990). Still, L-iduronic acid has a greater conformational flexibility which results in a more flexible structure, and therefore in less (electrostatic) stress in a polymer chain (Casu *et al*., 1988). It has been suggested that this increased chain flexibility is crucial for the binding capacity of L-iduronic acid containing glycosaminoglycans like heparin (Jia *et al*., 2009).

In vitro synthesis of heparin – Needed: improved D-glucuronyl C5-epimerase

Heparin is extracted rather inexpensively from animal derived tissue. However, several (potential) limitations of this methods exist. Firstly, like all animal-derived materials, there is a potential risk of contamination of the product with either allergens or pathogens. Secondly, there is a potential shortage of crude material since for the European and U.S. market only heparin isolated from porcine material is allowed since the outbreak of BSE in the 1990's. A third and major drawback of these extraction methods is that isolated heparin is always a heterogeneous mixture of polymers. For pharmaceutical purposes polydispersity should be reduced to a minimum, resulting in defined product size and product modifications. Also limiting is the restricted natural variety of heparin. Non-natural heparin analogs that have potentially better therapeutic characteristics, cannot be obtained using the existing isolation methods.

Much effort has been taken to reach chemical synthesis of heparin-like oligomers, leading to successful chemical synthesis methods. Several synthetic heparin derivatives have been described (De Kort *et al*., 2005). The first synthetic antithrombotic pentasaccharide (Fondaparinux) is available on the world-market as Arixtra® (Petitou and van Boeckel, 2004). These completely chemical ways of synthesis require, mainly due to the many protection and de-protection events, over 60 steps. On the contrary, the absolute strength of this process is the degree of control of the process; the end-product is very well defined.

Despite the success of these chemical synthesis approaches on relatively small scale, these methods are not applicable to bulk-scale production processes, as typical yields are

less than 0.5% (Kuberan *et al*., 2003). Alternatively, also a more feasible enzymatic biosynthesis approach has been described. This shows the modification of a relatively cheap and largely available starter material, generating a heparin-like product (Kuberan *et al*., 2003; Lindahl *et al*., 2005). Modification of the capsular polysaccharide of a bacterium (*E. coli* K5), results in a heparin-like polymer (neoheparin) that has similar anticoagulant properties as heparin. Compared to the chemical synthesis, higher yields can be achieved, in largely reduced synthesis time.

Most heparin modifications can be done in a chemical way (Casu *et al*., 1994), however the epimerization of D-glucuronic acid towards L-iduronic acid is an exception. An enzyme for C5-epimerization therefore is crucial when aiming for controlled synthesis of heparin. Two alternative ways to reach controlled enzymo-chemical heparin synthesis could be (1) controlled heparin chain building via incorporation of UDP-IdoA or (2) obtaining a D-glucuronyl C5-epimerase that is easily produced in large quantities and robust enough to use in an industrial process.

(1) The controlled chain elongation using UDP-sugars would give the highest degree of control, and therefore the most defined end-product. However, UDP-L-iduronic acid is considered as an unnatural sugar that only is proven to exist when produced chemically, and its limited availability implies high costs. Additionally, there is a bottleneck in the UDP-IdoA incorporation. Previous attempts have been made to incorporate chemically synthesized UDP-L-iduronic acid by glucuronyl-transferases (Weïver *et al*., 2008). Until now this was not successful, suggesting that specific glycosyltransferases will be required for the incorporation of L-iduronic acid. The lack of natural UDP-IdoA and the disability of L-iduronic acid incorporation via an UPD-sugar means the requirement of two novel types of enzymes; a glycosyltransferase that accepts UDP-IdoA and a C5-epimerase that can convert UDP-GlcA to UDP-IdoA.

(2) Alternatively the potential of the mouse D-glucuronyl C5-epimerase (MmC5) as an industrial tool can be improved by increasing its robustness and production levels. The D-glucuronyl C5-epimerase as described in the heparin / heparan sulfate biosynthesis has been cloned and heterologously expressed (Li *et al*., 2001). Unfortunately the enzyme has serious limitation for use in (large scale industrial) *in vitro* processes in terms of protein stability and protein production. Protein engineering of the MmC5 would be an option to improve the enzyme for industrial purposed. Another approach would be the identification and

production of robust D-glucuronyl C5-epimerases orthologs that can efficiently catalyze the same reaction.

Conclusions

Glycosaminoglycans are a specific class of biological active polysaccharides. As described above, a small number of glycosaminoglycans is involved in numerous biological processes. Most of these glycosaminoglycans can vary in their degree of polymerization, D-glucuronic acid epimerization and acetylation / sulfation pattern, each can have major impact on biochemical / biophysical features and as such on their biological functionality. Many different enzymes are involved in glycosaminoglycan biosynthesis, and many of them exist in multiple isoforms, allowing the variations as seen for glycosaminoglycans. Altogether this can lead to the large variation in chain modification as seen in glycosaminoglycans like heparin / heparan sulfate.

Traditionally glycosaminoglycans are isolated from animal tissue, however this brings some undesired side-effects. Not only there is a risk of immunogenic reactions upon co-isolation of animal derived antigens, also the isolated fractions are typically highly polydisperse and show a large variability in degree of sulfation and epimerization. To obtain monodisperse and uniform glycosaminoglycans is a major challenge when starting from such sources, especially when aiming for pharmaceutical applications. Chemical synthesis of these structures is rather complex and thus economically not feasible for bulk-scale production.

The use of heparosan as a starting material for heparin production is an alternative approach. Such chemo-enzymatic approaches have been tested previously on small scale, however there is ample room for improvement at the level of enzyme optimization. A key enzyme required to move the chemo-enzymatic production process to a cost effective level concerns a D-glucuronyl C5-epimerase. Most steps in heparin chain modification can be done chemically. Although the C5-epimerization step can be done chemically at disaccharide-level (Chiba *et al*., 1988), it is not possible to C5-epimerize GlcA at a polymer level. This implies the requirement for a industrial applicable D-glucuronyl C5-epimerase when aiming for chemo-enzymatic synthesis of heparin.

References

Aachmann FL, Svanem BIG, Güntert P, Petersen SB, Valla S, Wimmer R. (2006). NMR structure of the R-module. A parallel β-roll subunit from an *Azotobacter vinelandii* mannuronan C-5 epimerase. *The Journal of Biological Chemistry* **281**, 7350-7356.

Aikawa J, Grobe K, Tsujimoto M, Esko JD. (2001). Multiple isozymes of heparan sulfate/heparin GlcNAc N-Deacetylase/GlcN N-Sulfotransferase: Structure and activity of the fourth member, NDST4. *The Journal of Biological Chemistry* **276**, 5876-5882.

Alban S. (1997). Carbohydrates with anticoagulant and antithrombotic properties. *In Carbohydrates in Drug Design (1997)*, 209-276.

Bourin MC, Lindahl U. (1993). Glycosaminoglycans and the regulation of blood coagulation. *Biochemical Journal* **289**, 313-330.

Bülow H, Hobert O. (2004). Differential sulfations and epimerization define heparan sulfate specificity in nervous system development. *Neuron* **41**, 723-736.

Casu B, Grazioli G, Razi N, Guerrini M, Naggi A, Torri G, Oreste P, Tursi F, Zoppetti G, Lindahl U. (1994). Heparin-like compounds prepared by chemical modification of capsular polysaccharide from *E. coli* K5. *Carbohydrate Research* **263**(2), 271-284.

Casu B, Lindahl U. (2001). Structure and biological interactions of heparin and heparan sulfate. *Advances in Carbohydrate Chemistry and Biochemistry* **57**, 159-206.

Casu B, Petitou M, Provasoli M, Sinaÿ P. (1988). Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans. *Trends in Biochemical Sciences* **13**(6), 221-225.

Chi L, Munoz EM, Choi HS, Ha YW, Kim YS, Toida T, Linhardt RJ. (2006). Preparation and structural determination of large oligosaccharides derived from acharan sulfate. *Carbohydrate Research* **341**, 864-869.

Chiba T, Jacquinet JC, Sinaÿ P. (1988). Chemical synthesis of L-iduronic acid-containing disaccharidic fragments of heparin. *Carbohydrate Research* **174**, 253-264.

Chong BF, Blank LM, Mclaughlin R, Nielsen LK. (2005). Microbial hyaluronic acid production. *Applied microbiology and biotechnology* **66**(4), 341-351.

Cifonelli JA, Dorfman A. (1962). The uronic acid of heparin. *Biochemical and biophysical research communications* **7**, 41-45.

Conrad HE. (1998). Heparin vs Heparan Sulfate **In:** Heparin-Binding Proteins*. Academic Press, San Diego.* Pages 1-5.

Ertesvåg H, Høidal HK, Schjerven H, Svanem BIG, Valla S. (1999). Mannuronan C-5-epimerases and their application for in vitro and in vivo design of new alginates useful in biotechnology. *Metabolic engineering.* **1**(3), 262-269.

Ertesvåg H, Valla S. (1999). The A modules of the *Azotobacter vinelandii* mannuronan-C-5-epimerase

AlgE1 are sufficient for both epimerization and binding of Ca^{2+}. *Journal of Bacteriology* **181**(10), 3033-3038.

Esko JD, Lindahl U. (2001). Molecular diversity of heparin sulfate. *The Journal of Clinical Investigation* **108**(2), 169-173.

Esko JD, Selleck SB. (2002). Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annual Review of Biochemistry* **71**, 435-471.

Faham S, Hileman RE, Fromm JR Linhardt RJ, Rees DC. (1996). Heparin structure and interactions with basic Fibroblast Growth Factor. *Science* **271**, 1116-1120.

Ferro DR, Provasoli A, Ragazzi M, Casu B, Torri G, Bossennec V, Perley B, Sinaÿ P, Petitou M, Choay J. (1990). Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydrate Research* **195**, 157-167.

Franklin MJ, Chitnis CE, Gacesa P, Sonesson A, White DC, Ohman DE. (1994). *Pseudomonas aeruginosa* AlgG is a polymer level alginate C5-mannuronan epimerase. *Journal of bacteriology* **176**(7), 1821-1830.

Fraser JR, Laurent TC, Laurent UB. (1997). Hyaluronan: its nature, distribution, functions and turnover. *Journal of internal medicine* **242**(1), 27-33.

George E. (1998). Intra-articular hyaluronan treatment for osteoarthritis. *Annals of the Rheumatic Diseases* **57**, 637-640.

Giancarlo G, Farber SA. (2005). D-glucuronyl C5-epimerase acts in dorso-ventral axis formation in zebrafish. *BMC Developmental Biology* **5**:19.

Griffin CC, Linhardt RJ, Van Gorp CL, Toida T, Hileman RE, Schubert RL, Brown SE. (1995). Isolation and characterization of heparan sulfate from crude porcine intestinal mucosal peptidoglycan heparin. *Carbohydrate Research* **276**, 183-197.

Habuchi H, Tanaka M, Habuchi O, Yoshida K, Suzuki H, Ban K, Kimata K. (2000). The occurrence of three isoforms of heparan sulfate 6-O-sulfotransferase having different specificities for hexuronic acid adjacent to the targeted N-sulfoglucosamine. *The Journal of Biological Chemistry* **275**(4), 2859-2868.

Habuchi H, Habuchi O, Kimata K. (2004). Sulfation pattern in glycosaminoglycan: Does it have a code? *Glycoconjugate Journal* **21**, 47-52.

Hagner-McWhirter Å, Lindahl U, Li J. (2000). Biosynthesis of heparin/heparan sulphate: mechanism of epimerization of glucuronyl C-5. *Biochemical Journal* **347**, 69-75.

Hagner-McWhirter Å, Li J, Oscarson S, Lindahl U. (2004). Irreversible glucuronyl C5-epimerization in the biosynthesis of heparan sulfate. *The Journal of Biological Chemistry* **279**(15), 14631-14638.

Haug A, Larsen B. (1970). Biosynthesis of alginate: Part II. Polymannuronic acid C-5-epimerase from *Azotobacter vinelandii* (Lipman). *Carbohydrate Research* **17**(2), 297-308.

Hoffman P, Linker A, Meyer K. (1956). Uronic acid of chondroitin sulfate B. *Science* **124**, 1252.

Höök M, Lindahl U, Bäckström G, Malmström A, Fransson LA. (1974). Biosynthesis of heparin. III. Formation of L-iduronic acid residues. *The Journal of Biological Chemistry* **249**, 3908-3915.

Howell WH, Holt E. (1918). Two new factors in blood coagulation – Heparin and pro-antithrombin. *American journal of physiology* **47**(3), 328-341.

Hulett MD, Freeman C, Hamdorf BJ, Baker RT, Harris MJ, Parish CR. (1999). Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis. *Nature Medicine* **5**(7), 803-809.

Im AR, Park Y, Sim JS, Zhang Z, Liu Z, Linhardt RJ, Kim YS. (2010). Glycosaminoglycans from earthworms (*Eisenia andrei*). *Glycoconjugate Journal* **27**, 249-257.

Itano N, Kimata K. (2002). Mammalian Hyaluronan Synthases. *IUBMB Life.* **54**, 195–199.

Jacobson B, Davidson EA. (1963). Biosynthesis of uronic acids by skin enzymes. II. UDP d-glucuronic acid 5-epimerase. *The Journal of Biological Chemistry* **237**, 638-642.

Jacobsson I, Lindahl U, Jensen JW, Rodén L, Prihar H, Feingold DS. (1984). Biosynthesis of Heparin - Substrate specificity of heparosan N-sulfate D-Glucuronyl 5-epimerase. *The Journal of Biological Chemistry* **259**(2), 1056-1063.

Jia J, Maccarana M, Zhang X, Bespalov M, Lindahl U, Li JP. (2009). Lack of L-iduronic acid in heparan sulfate affects interaction with growth factors and cell signaling. *The Journal of Biological Chemistry* **284**(23), 15942-15950.

Kim YS, Jo YT, Chang IM, Toida TE, Park Y, Linhardt RJ. (1996). A new glycosaminoglycan from the giant African snail *Achatina fulica*. *The Journal of Biological Chemistry* **271**, 11750–11755.

Kim YS, Ahn MY, Wu SJ, Kim DH, Toida T, Teesch LM, Park Y, Yu G, Lin J, Linhardt RJ. (1998). Determination of the structure of oligosaccharides prepared from acharan sulfate. *Glycobiology* **8**(9), 869-877.

Kjellén L. (2003). Glucosaminyl N-deacetylase/N-sulphotransferases in heparan sulphate biosynthesis and biology. *Biochemical Society transactions* **31**(2), 340-342.

de Kort M, Buijsman RC, van Boeckel CA. (2005). Synthetic heparin derivatives as new anticoagulant drugs. *Drug discovery today* **10**(11), 769-779.

Kuberan B, Lech MZ, Beeler DL, Wu ZL, Rosenberg RD. (2003). Enzymatic synthesis of antithrombin III-binding heparan sulfate pentasaccharide. *Nature Biotechnology* **21**(11), 1343-1346.

Laurent TC, Laurent UB, Fraser JR. (1996). The structure and function of hyaluronan: An overview. *Immunology and cell biology* **74**(2), A1-A7.

Li J, Gong F, El Darwish K, Jalkanen M, Lindahl U. (2001). Characterization of the D-glucuronyl C5-epimerase involved in the biosynthesis of heparin and heparan sulfate. *The Journal of Biological Chemistry* **276**(23), 20069-20077.

Li JP, Gong F, Hagner-McWhirter A, Forsberg E, Abrink M, Kisilevsky R, Zhang X, Lindahl U. (2003). Targeted disruption of a murine glucuronyl C5-epimerase gene results in heparan sulfate lacking L-iduronic acid and in neonatal lethality. *The Journal of Biological Chemistry* **278**(31), 28363-28366.

Li JP. (2010). Glucuronyl C5-epimerase an enzyme converting glucuronic acid to iduronic acid in heparan sulfate/heparin biosynthesis. *Progress in molecular biology and translational science* **93**, 59-78.

Lind T, Tufaro F, McCormick C, Lindahl U, Lidholt K. (1998). The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *The Journal of Biological Chemistry* **273**(41), 26265-26268.

Lindahl U, Bäckström G, Malmström A, Fransson LA. (1972). Biosynthesis of L-iduronic acid in heparin: Epimerization of D-glucuronic acid on the polymer level. *Biochemical and Biophysical Research Communications* **46**(2), 985-991.

Lindahl U, Bäckström G, Höök M, Thunberg L, Fransson LA, Linker A. (1979). Structure of the antithrombin-binding site in heparin. *Proceedings of the National Academy of Sciences* **76**(7), 3198–3202.

Lindahl U. (1997). Heparan sulfate – a polyanion with multiple messages. *Pure and Applied Chemistry* **69**(9), 1897-1902.

Lindahl, U., Kusche-Gullberg, M., Kjellen, L. (1998). Regulated diversity of heparan sulfate. *Journal of Biological Chemistry* **273**(39), 24979-24982.

Lindahl U, Li JP. (2009). Interactions between heparan sulfate and proteins-design and functional implications. *International review of cell and molecular biology* **276**, 105-159.

Lindahl U, Li JP, Kusche-Gullberg M, Salmivirta M, Alaranta S, Veromaa T, Emeis J, Roberts I, Taylor C, Oreste P, Zoppetti G, Naggi A, Torri G, Casu B. (2005). Generation of "neoheparin" from *E. coli* K5 capsular polysaccharide. *Journal of medicinal chemistry* **48**(2), 349-352.

Linhardt RJ. (1991). Heparin: An important drug enters its seventh decade. *Chemistry and Industry* **2**, 45-50.

Linhardt RJ, Ampofo SA, Fareed J, Hoppensteadt D, Mulliken JB, Folkman J. (1992). Isolation and characterization of human heparin. *Biochemistry* **31**(49), 12441-12445.

Linhardt RJ, Toida T. (1997). Heparin oligosaccharides: new analogues-development and applications. **In**: **Witczak ZB, Nieforth KA (eds).** (1997) Carbohydrates in Drug Design *Marcel Dekker, New York*, 277-308.

Luppi E, Cesaretti M, Volpi N. (2005). Purification and characterization of heparin from the Italian clam *Callista chione. Biomacromolecules* **6**, 1672-1678.

Maccarana M, Olander B, Malmström A,Tiedemann K, Aebersold R, Lindahl U. (2006). Biosynthesis of dermatan sulfate. Chondroitin-glucuronate C5-epimerase is identical to SART2. *The Journal of Biological Chemistry* **281**(17), 11560-11568.

Malmström A, Fransson LA, Höök M, Lindahl U. (1975). Biosynthesis of dermatan sulfate. I. Formation of L-iduronic acid residues. *The Journal of Biological Chemistry* **250**, 3419-3425.

Malmström A, Åberg L. (1982). Biosynthesis of dermatan sulphate. Assay and properties of the uronosyl C-5 epimerase. *Biochemical Journal* **201**, 489-493.

McCormick C, Duncan G, Goutsos KT, Tufaro F. (2000). The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proceedings of the National Academy of Sciences of the United States of America* **97**(2), 668-673.

McCormick C, Leduc Y, Martindale D, Mattison K, Esford LE, Dyer AP, Tufaro F. (1998). The putative

tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. *Nature genetics* **19**(2), 158-161.

Medeiros GF, Mendes A, Castro RAB, Baú EC, Nader HB, Dietrich CP. (2000). Distribution of sulfated glycosaminoglycans in the animal kingdom: widespread occurrence of heparin-like compounds in invertebrates. *Biochimica et Biophysica Acta* **1475**, 287-294.

Meyer K, Palmer JW. (1934). The polysaccharide of the vitreous humor. *The Journal of Biological Chemistry* **107**, 629-634.

Mizumoto S, Kitagawa H, Sugahara K. (2005). Biosynthesis of heparin and heparan sulfate. **In: Garg HG, Linhardt RJ, Hales CA (eds).** (2005). Chemistry and biology of heparin and heparan sulfate. *Elsevier, Oxford*, 203-243.

Nader HB, Chavante SF, dos-Santos EA, Oliveira TW, de-Paiva JF, Jerônimo SM, Medeiros GF, de-Abreu LR, Leite EL, de-Sousa-Filho JF, Castro RA, Toma L, Tersariol IL, Porcionatto MA, Dietrich CP. (1999). Heparan sulfates and heparins: similar compounds performing the same functions in vertebrates and invertebrates? *Brazilian journal of medical and biological research* **32**(5), 529-538.

Pacheco B, Maccarana M, Goodlett DR, Malmström A, Malmström L. (2008). Identification of the active site of DS-epimerase 1 and requirement of N-glycosylation for enzyme function. *The Journal of Biological Chemistry* **284**(3), 1741-1747.

Perlin AS, Mazurek M, Jaques LB, Kavanaugh LW. (1968). A proton magnetic resonance spectral study of heparin. L-iduronic acid residues in commercial heparins. *Carbohydrate Research* **7**, 369-379.

Perrimon N, Bernfield M. (2000). Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725-728.

Petitou M, van Boeckel CA. (2004). A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angewandte Chemie (International ed. in English* **43**(24), 3118-3133.

Pinhal MAS, Smith B, Olson S, Aikawa J, Kimata K, Esko JD. (2001). Enzyme interactions in heparin sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact *in vivo. Proceedings of the National Academy of Sciences* **98**(23) 12984-12989.

Remminghorst U, Rehm BH. (2006). Bacterial alginates: from biosynthesis to applications. *Biotechnology letters* **28**(21), 1701-1712.

Rong J, Habuchi H, Kimata K, Lindahl U, Kusche-Gullberg M. (2001). Substrate specificity of the heparan sulfate hexuronic acid 2-O-sulfotransferase. *Biochemistry* **40**(18), 5548-5555.

Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvåg H, Holtan S, Aachmann FL, Valla S Dijkstra BW. (2008). Structural and mutational characterization of the catalytic A-module of the Mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii. The Journal of Biological Chemistry* **283**, 23819-23828.

Safaiyan F, Lindahl U, Salmivirta M. (2000). Structural diversity of N-sulfated heparan sulfate domains: distinct modes of glucuronyl C5 epimerization, iduronic acid 2-O-sulfation, and glucosamine 6-O-sulfation. *Biochemistry* **39**(35), 10823-10830.

Sasisekharan R, Venkataraman, G. (2000). Heparin and heparan sulfate: biosynthesis, structure and function. *Current Opinion in Chemical Biology* **4**, 626-631.

Scott JE, Heatley F. (1999). Hyaluronan forms specific stable tertiary structures in aqueous solution: a 13C NMR study. *Proceedings of the National Academy of Sciences of the United States of America* **96**(9), 4850-4855.

Sherman L, Sleeman J, Herrlich P, Ponta H. (1994). Hyaluronate receptors: key players in growth, differentiation, migration and tumor progression. *Current opinion in cell biology* **6**(5), 726-733.

Shukla D, Liu J, Blaiklock P, Shworak NW, Bai X, Esko JD, Cohen GH, Eisenberg RJ, Rosenberg RD, Spear PG. (1999). A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* **99**, 13-22.

Shworak NW, Liu J, Petros LM, Zhang L, Kobayashi M, Copeland NG, Jenkins NA, Rosenberg RD. (1999). Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase. Isolation, characterization, and expression of human cdnas and identification of distinct genomic loci. *Journal of biological chemistry* **274**(8), 5170-5184.

Spicer AP, Seldin MF, Olsen AS, Brown N, Wells DE, Doggett NA, Itano N, Kimata K, Inazawa J, McDonald JA. (1997). Chromosomal localization of the human and mouse hyaluronan synthase genes. *Genomics* **41**(3), 493-497.

Thunberg L, Bäckström G, Lindahl U. (1982). Further characterization of the antithrombin-binding sequence in heparin. *Carbohydrate Research* **100**, 393-410.

Valla S, Li J, Ertesvåg H, Barbeyron T, Lindahl U. (2001). Hexuronyl C5-epimerases in alginate and glycosaminoglycan biosynthesis. *Biochimie* **83**(8), 819-830.

Vlodavsky I, Friedmann Y, Elkin M, Aingorn H, Atzmon R, Ishai-Michaeli R, Bitan M, Pappo O, Peretz T, Michal I, Spector L, Pecker I. (1999). Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis. *Nature Medicine* **5**(7), 793-802.

Weigel PH, Hascall VC, Tammi M. (1997). Hyaluronan synthases. *Journal of biological chemistry* **272**(22), 13997-14000.

Weigel PH, DeAngelis PL. (2007). Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *The Journal of biological chemistry* **282**(51), 36777-36781.

Weissmann B, Meyer K. (1954). The structure of hyalobiuronic acid and of hyaluronic acid from umbilical cord. *Journal of the American Chemical Society* **76**(7), 1753-1757.

Weïwer M, Sherwood T, Green DE, Chen M, DeAngelis PL, Liu J, Linhardt RJ. (2008). Synthesis of Uridine 5′-diphosphoiduronic Acid: A Potential Substrate for the Chemoenzymatic Synthesis of Heparin. *The Journal of Organic Chemistry* **73**, 7631-7637.

Wight TN, Kinsella MG, Qwarnström EE. (1992). The role of proteoglycans in cell adhesion, migration and proliferation. *Current Opinion in Cell Biology* **4**(5), 793-801.

Chapter 2

Occurrence of L-iduronic acid and putative D-glucuronyl C5-epimerases in prokaryotes

This chapter has been published in:

Raedts J, Kengen SWM, van der Oost J. (2011). Occurrence of L-iduronic acid and putative D-glucuronyl C5-epimerases in prokaryotes. *Glycoconjugate Journal* **28**(2), 57-66.

Abstract

Glycosaminoglycans (GAGs) are polysaccharides that are typically present in a wide diversity of animal tissue. Most common GAGs are well-characterized and pharmaceutical applications exist for many of these compounds, e.g. heparin and hyaluronan. In addition, also bacterial glycosaminoglycan-like structures exist. Some of these bacterial GAGs have been characterized, but until now no bacterial GAG has been found that possesses the modifications that are characteristic for many of the animal GAGs such as sulfation and C5-epimerization. Nevertheless, the latter conversion may also occur in bacterial and archaeal GAGs, as some prokaryotic polysaccharides have been demonstrated to contain L-iduronic acid. However, experimental evidence for the enzymatic synthesis of L-iduronic acid in prokaryotes is as yet lacking. We therefore performed an *in silico* screen for D-glucuronyl C5-epimerases in prokaryotes. Multiple candidate C5-epimerases were found, suggesting that many more microorganisms are likely to exist possessing an L-iduronic acid residue as constituent of their cell wall polysaccharides.

Introduction

Glycosaminoglycans (GAGs) are long linear negatively charged hetero-polysaccharides, consisting of repeating disaccharide residues usually of a hexuronic acid linked to a hexosamine. Glycosaminoglycans are ubiquitously found throughout the animal kingdom, where they are involved in a wide variety of biological processes. The most common glycosaminoglycans are chondroitin sulfate, dermatan sulfate, heparan sulfate, heparin, hyaluronan and keratan sulfate (Chapter 1).

The glycosaminoglycan hexuronic acid residue is either D-glucuronic acid or its C5 epimer L-iduronic acid. The latter is a rather unique hexuronic acid that is typically found in glycosaminoglycans, while D-glucuronic acid is a very common compound in nature. The epimerization of D-glucuronic acid towards L-iduronic acid is mediated by a D-glucuronyl C5-epimerase acting at a polysaccharide level after D-glucuronic acid incorporation (Lindahl *et al*., 1972; Höök *et al*., 1974). The C5-epimerization of D-glucuronic acid is essential for specific binding properties of versatile glycosaminoglycans like heparin and heparan sulfate. Another D-glucuronyl C5-epimerase does exist responsible for dermatan sulfate biosynthesis, as well as a C5-epimerase involved in alginate biosynthesis. These epimerases do not show sequence homology to the heparin D-glucuronyl C5-epimerase (Valla *et al*., 2001).

Glycosaminoglycans traditionally are isolated from animal tissue. A major drawback is the great polydispersity of animal-derived material, both in chain length and degree of epimerization/sulfation. Large scale chemical synthesis of heparin is not feasible, because of the C5-epimerization of D-glucuronic acid being one of the major bottlenecks. L-iduronic acid can be synthesized chemically, however C5-epimerization of D-glucuronic acid as part of a polymer only can be done enzymatically. The C5-epimerase in animal heparin biosynthesis has been identified and characterized, and has some major limitations to use in large scale biotechnological production processes.

Novel D-glucuronyl C5-epimerases that have less restrictions in substrate acceptance, better stability and easier production methods would have great potential in controlled chemo-enzymatic synthesis of L-iduronic acid containing polymers like heparin and heparin-analogs. In this review we discuss putative novel C5-epimerases that might convert D-glucuronic acid to L-iduronic acid with less restrictions**.**

Glycosaminoglycan-like structures in bacterial capsules

Traditionally, GAGs were considered as polymeric structures only to be found in (in)vertebrate animals. However, in the past decades, some polyanionic bacterial cell wall polysaccharides have been described, of which the structure has resemblance to some well-characterized animal glycosaminoglycans. Microorganisms possessing GAGs generally are pathogenic bacteria in which the surface exposed capsular polysaccharides are likely to serve as a virulence factor. The resemblance between the bacterial capsule and the animal GAG results in very limited or no response of the hosts' immune system. GAG-containing pathogens, including serotypes of both *Escherichia coli* and *Pasteurella multocida*, are discussed below.

***E. coli* GAGs |** Many different serotypes of *E. coli* have been described. Discrimination between these strains is generally based on antigenic studies. Mostly differences can be found in a specific part of the bacterial lipopolysaccharide (LPS), the so-called O-antigen. To a lesser extent differences are seen in antigenic properties of the flagella (H-antigen) and the bacterial capsule (K-antigen). The latter is a protective layer of polysaccharides that generally can be easily observed by light microscopy. Over 70 different K-antigens have been described (Ørskov *et al*., 1977). As discussed below, two of them are analogous to glycosaminoglycans as found in animals.

A chondroitin-like glycosaminoglycan has been isolated from *E. coli* O5 : K4 : H4 (Rodriguez *et al*., 1988). Identical to chondroitin this "K4 capsular polysaccharide" consists of equimolar amounts of β-D-glucuronic acid and β-N-acetyl-galactosamine. Unlike chondroitin an additional fructose is β(1,3) linked as a substituent to each GlcA residue. However, this modification can easily be removed by mild acidification, resulting in a chondroitin backbone. Upon removal of the fructose residue, the immune response against the K4 polysaccharide decreases considerably.

Another glycosaminoglycan-like structure has been described for the K5 antigen of *E. coli* O10 : K5 : H4. This capsular polysaccharide has an identical structure to heparosan, the unsulfated and non-epimerized backbone structure of heparan sulfate and heparin. The K5 capsule is a linear polysaccharide containing α-N-acetyl-glucosamine and β-glucuronic acid in equimolar amounts, linked by an (1-4) glycosidic bond (Vann *et al*., 1981). In contrast to

animals, no post-polymerization modifications occur on the heparosan molecule. This makes the K5 polysaccharide a useful substrate to study the enzymes in heparin biosynthesis (Kusche *et al*., 1991; Hagner-McWhirter *et al*., 2000), and a potential precursor for chemo-enzymatic synthesis of heparin.

***P. multocida* GAGs |** Similar glycosaminoglycans have been isolated from several serotypes of another pathogenic gamma-proteobacterium, namely *Pasteurella multocida*. The capsules of *P. multocida* type A, D and F could be removed upon treatment with different glycosaminoglycan hydrolases (Rimler, 1994). A more detailed analysis of the capsular polysaccharides of *P. multocida* type D and F has revealed similarity with the K-antigens of *E. coli* K5 and K4 respectively. The *P. multocida* type D polymer is identical to heparosan, the type F polymer is unmodified chondroitin (DeAngelis *et al*., 2002).

Additionally, another *P. multocida* capsular polysaccharide has been described to be analogous to a vertebrate glycosaminoglycan. The extracellular capsule of *P. multocida* type A is chemically identical to the animal GAG hyaluronan (Rosner *et al*., 1992). In addition, multiple species of Streptococci have been described to have such a hyaluronan capsule. All these strains are pathogenic to human or other mammals, the hyaluronan capsule having an important role in preventing an immune response. Since these molecules are identical to mammalian hyaluronan, bacterially produced hyaluronic acid has substantial commercial value. In addition to animal derived hyaluronan, it is nowadays widely commercially available for numerous existing applications.

Although multiple examples of bacterial GAGs do exist, for none of these glycosaminoglycans modifications have been observed similar to those in the GAG biosynthesis pathways in animals. A key modification step in these pathways is the C5-epimerization of D-glucuronic acid towards L-iduronic acid, catalyzed by a D-glucuronyl C5-epimerase. No bacterial counterpart of this enzyme has been experimentally characterized to date. However, the presence of L-iduronic acid in extracellular polysaccharides of several microorganisms, suggests that D-glucuronyl C5-epimerases do exist in prokaryotes.

Identification of iduronic acid in microbes

While being a well-known component of (animal) glycosaminoglycans, the presence of L-iduronic acid in prokaryotes is rather uncommon. For some time it was believed that L-iduronic acid could only be found in multicellular eukaryotes. However, as discussed below, in the last decades multiple examples of microbial L-iduronic acid have been published.

Bacteria

The first case of L-iduronic acid being present in a prokaryote was reported in a study of the gram-positive bacterium *Clostridium perfringens* NCTC 10578 (Darby *et al*., 1970). L-iduronic acid was identified in a purified "type-specific" polysaccharide from *Clostridium perfringens* strain Hobbs 10. The exact polysaccharide structure is still unknown but the L-iduronic acid level in the isolated polysaccharide is estimated to be 7%. Most likely this "type-specific" polysaccharide is part of a bacterial capsular polysaccharide (Lee and Cherniak, 1974).

Another report on the presence of L-iduronic acid in a prokaryote concerns the analysis of specific extracellular polysaccharide (EPS) of *Butyrivibrio fibrisolvens* strain X6C61 (Stack *et al*., 1988). As much as 37 strains of *B. fibrisolvens* were screened in total, but only a single strain appeared to contain L-iduronic acid. This indicates that L-iduronic acid is part of a type-specific EPS. The exact composition of the EPS remains to be characterized, although it has been proposed that L-iduronic acid is associated to a galactosamine residue.

In addition, several reports exist in which L-iduronic acid is identified as a compound of an O-specific polysaccharide. An O-antigen is the highly variable part of a lipopolysaccharide (LPS), which is present in the outer membrane of gram-negative bacteria. The first report of L-iduronic acid being present in an O-antigen was after structure elucidation of the O-antigen of the marine bacterium *Pseudoalteromonas haloplanktis* strain KMM 223 (44-1) (Hanniffy *et al*., 1998). The L-iduronic acid residue is part of a pentasaccharide (Fig.1) that additionally contains two D-glucuronic acid residues and two residues of the uncommon QuiN4N (2,4-diamino-2,4,6-trideoxyglucose). The high amount of hexuronic acids results in a highly acidic O-antigen. In addition, the uncommon QuiN4N residues and GlcA residues have been found in other serotypes of *Pseudoalteromonas* (Nazarenkoa *et al*., 2003), however strain KMM 223 remains the only example that has an L-iduronic acid containing O-antigen.

Organism	Structure	Reference
Pseudoalteromonas Haloplanktis KMM 223 (44-1)	→4)- β-D-GlcA-(1→4)-β-D-GlcA-(1→3)-β-D-QuiNHb4NHb-(1→2)- **α-L-IdoA**-(→ 4 ↑ 1 α-D-QuiNAc4NAc	Hanniffy *et al*., 1998
Escherichia coli O112ab & *Shigella boydii* B15	β-D-GlcNAc 1 ↓ 3 →4)-α-D-GalNAc-(1→4)→α-D-Glc-(1→4)-**α-L-IdoA**-(1→3)-β-D-GalNAc-(1→	Perepelov *et al*., 2008 Liu *et al*., 2008
Halobacterium halobium	GlcA-(1→4)-GlcA-(1→4)-GlcA-(1-4)-β-D-Glc-(1→N)-Asn 3 3 3 | ↑ ↑ ↑ X OSO_3^- OSO_3^- OSO_3^- | Thr / Ser	1/3 of GlcA residues can be replaced by **IdoA** Wieland *et al*., 1985

Figure 1 – Known prokaryotic structures containing L-iduronic acid; bacterial O-antigens and the *Halobacterium halobium* glycoconjugate. GlcA, glucuronic acid; QuiNHb4N, 2,4-diamino-2,4,6-trideoxy-D-glucose (Hb, S-3-hydroxybutyryl; Ac, acetyl); IdoA, iduronic acid; GlcNAc, N-acetyl-glucosamine; GalNAc, N-acetyl-galactosamine.

More recently, two additional O-antigens have been identified in which L-iduronic acid is one of the building blocks (Fig.1). Both *Escherichia coli* type 112ab and *Shigella boydii* B15, have an identical pentasaccharide structure (Perepelov *et al*., 2008; Liu *et al*., 2009). Many more O-antigens structures of various serotypes of both *E. coli* and *S. boydii* have been resolved to date, however L-iduronic acid seems to be restricted to these two reported strains. Just like the previous reports dealing with L-iduronic acid identification in bacteria, the occurrence of this structure is highly type-specific rather than a general feature.

Archaea

L-iduronic acid has been reported in archaea only once. There is evidence of the presence of iduronic acid in a cell surface lipoprotein of *Halobacterium halobium* (Wieland *et al*., 1986). The cell wall of this archaeon is a glycoprotein based S-layer. The glycoprotein has two specific forms of N-glycosylation. First each polypeptide consists of a single glycosaminoglycan-like polysaccharide with a [-4)-GalNAc-(1-4)-GalA-(1-3)-GalNAc-(1-]$_{n10-15}$ backbone attached. Apart from that, there are 12 potential glycosylation sites where an Asn-Glc (asparaginylglucose) linkage unit is extended by two or three β(1-4) bound glucuronic acid residues. About 1/3 of these glucuronic acid moieties are replaced by an iduronic acid

(Fig.1). An identical glycoconjugate can be found at the organisms' flagellin (Wieland *et al*., 1985). Although archaeal flagellins often undergo posttranslational modification (Jarrell *et al*., 1996), until now this is the only report of iduronic acid presence in such a structure.

Non-canonical L-iduronic acid containing polymers in eukaryotes

L-iduronic acid traditionally is considered to be a component in many common animal glycosaminoglycans. Apart from these well-characterized GAGs, recently L-iduronic acid also has been identified in some atypical polymers. The eukaryotic organisms having these non-canonical polymers usually do not possess the traditional GAGs as found in animals. Possibly the formation of L-iduronic acid is the result of another C5-epimerase than the heparosan D-glucuronyl C5-epimerase. An overview of some of these rare L-iduronic acid containing structures is provided below.

Algae | *Pleurochrysis haptonemofera* is a unicellular coccolithophorid marine alga. It produces coccolith, a calcified scale. Apart from carbonate crystals, this scale contains a small amount of polysaccharide called "coccolith matrix acidic polysaccharide" (CMAP). The structure of CMAP has been determined to be composed of a repeating disaccharide structure, of which L-iduronic acid is one of the sugars (Ozaki *et al*., 2007). In addition there are reports of the existence of L-iduronic acid in specific polysaccharides in multicellular algae. The cell wall of sea lettuce (genus *Ulva*) includes four types of polysaccharides, of which the water-soluble ulvan is exclusively found in members of the Ulvales. This polysaccharide has a repetitive disaccharide of L-iduronic acid that is α(1-4) linked to a C3-sulfated rhamnose (Lahye and Robic, 2007).

Fungi | Phallic acids are specific glycuronans that can be found in the fruiting-bodies of members of the taxon Phallales. Tsuchihashi and colleagues have described the isolation of phallic acid of at least ten species, all containing L-iduronic acid. (Miyazaki *et al*., 1979). The exact structural composition is still unknown, but it has been reported that these polysaccharides are composed of β-glucuronic acid and α-iduronic acid residues that have an (1-4) linkage. The internal ratio of these two hexuronic acids varies around 2:1 to 3:1. The polysaccharide is called protuberic acid when the ratio glucuronic acid to iduronic acid is equal to 2:1 (Tsuchihashi *et al*., 1983).

Sponges | Citronamides A and B are unique products that have been isolated from the Australian sponge *Citronia astra*. Both are non-canonical tetrapeptides with a linked 3- or 4-O-(aminocarbonyl)-α-iduronic acid residue. These compounds accidently have been co-isolated with Dysinosin A, a potential serine protease (thrombin) inhibitor. Citronamides A and B are structurally not related to Dysinosin A, and the biological function of these products still needs to be clarified (Carroll *et al*., 2009).

Identification of C5-epimerases in prokaryotes

The above described polymeric structures are examples of L-iduronic acid containing polysaccharides and glycosaminoglycan-like structures in several microorganisms. The existence of L-iduronic acid does suggest D-glucuronyl C5-epimerase activity to occur in these organisms. The wide diversity of these GAG-like structures suggests the presence of candidate C5-epimerases with a different or broader substrate specificity. To date no such candidate enzyme has been identified.

In an attempt to identify candidate C5-epimerases, we screened all available prokaryotic genomes for sequences homologous to human D-glucuronyl C5-epimerase by Blast analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et al*., 1997). Multiple prokaryotic sequences were identified that have homology to the human protein sequence. Also Blasts on metagenome data reveal multiple putative prokaryotic C5-epimerases. All found candidate C5-epimerases have a well-conserved domain making them members of the pfam06662 superfamily (Bateman, *et al*., 2004) containing the consensus of the C-terminus of D-glucuronyl C5-epimerases. An overview of all prokaryotes having one or more candidate D-glucuronyl C5-epimerase gene(s) is provided in Table 1.

The occurrences of these candidate C5-epimerases seem to be type-specific rather than species-specific. This is in line with the earlier discussed reports on the identification of L-iduronic acid containing polymers in prokaryotes, that also appear to be type-specific in various species. Among the identified candidate C5-epimerases differences in size exist; however, this is mainly a result of variation of the N-terminal domain of the protein.

Table 1 – Overview candidate C5-epimerases in prokaryotes

Organism	Class	Size	Gene ID	e-value
Bacteria; Acidobacteria				
Candidatus Koribacter versatilis Ellin345	Candidatus koribacter	401 aa	94970543	9e-65
Bacteria; Actinobacteria				
Sanguibacter keddieii DSM 10542	Actinobacteridae	355 aa	269795908	1e-16
Conexibacter woesei DSM 14684	Rubrobacteridae	473 aa	284043115	3e-19
Conexibacter woesei DSM 14684	Rubrobacteridae	489 aa	284046741	1e-26
Bacteria; Bacteroidetes				
Bacteroides sp. 1_1_6	Bacteroidia	305 aa	253567783	1e-29
Bacteria; Firmicutes				
Bacillus cereus R309803	Bacillales	307 aa	229164211	2e-40
Clostridium ljungdahlii ATCC 49587	Clostridia	737 aa	300854052	6e-52
Clostridium ljungdahlii ATCC 49587	Clostridia	399 aa	300855735	9e-53
Clostridium ljungdahlii ATCC 49587	Clostridia	372 aa	300855857	7e-50
Clostridium kluyveri DSM 555	Clostridia	411 aa	153953040	1e-35
Clostridium kluyveri DSM 555	Clostridia	372 aa	153953635	2e-32
Clostridium kluyveri DSM 555	Clostridia	452 aa	153953720	1e-32
Clostridium kluyveri NBRC 12016	Clostridia	414 aa	219853696	1e-35
Ruminococcus sp. 18P13	Clostridia	302 aa	291544017	6e-34
Thermoanaerobacter tengcongensis MB4	Clostridia	312 aa	20807150	3e-58
Bacteria; Proteobacteria				
Campylobacter hominis ATCC BAA-381	Epsilonproteobacteria	458 aa	154149157	4e-34
Nautilia profundicola AmH	Epsilonproteobacteria	317 aa	224373665	2e-38
Bermanella marisrubri sp. RED65	Gammaproteobacteria	448 aa	94501207	2e-76
Escherichia coli O112ab	Gammaproteobacteria	323 aa	187880578	1e-23
Shigella boydii type 15	Gammaproteobacteria	323 aa	187880569	1e-23
Photorhabdus asymbiotica	Gammaproteobacteria	370 aa	253991788	5e-15
Photorhabdus luminescens sp. laumondii TTO1	Gammaproteobacteria	368 aa	37528614	4e-12
Vibrio cholerae bv. albensis VL426	Gammaproteobacteria	308 aa	229524794	2e-27
Vibrio cholerae st. CO845	Gammaproteobacteria	311 aa	295149004	9e-34
Archaea; Korarchaeota				
Candidatus Korarchaeum cryptofilum OPF8	Candidatus korarchaeum	337 aa	170290161	3e-28
Archaea; Euryarchaeota				
Methanothermobacter marburgensis st Marburg	Methanobacteria	381 aa	304314521	8e-40
Methanothermobacter thermautotrophicus ΔH	Methanobacteria	405 aa	15678358	2e-59
Methanocaldococcus fervens AG86	Methanococci	305 aa	241911015	5e-59
Methanocaldococcus infernus ME	Methanococci	298 aa	241904910	6e-48
Methanocaldococcus jannaschii DSM 2661	Methanococci	308 aa	15669315	5e-61
Methanocaldococcus vulcanius M7	Methanococci	305 aa	255051598	2e-51
Methanococcus voltae A3	Methanococci	330 aa	163798575	7e-10
Thermococcus barophilus MP	Thermococci	367 aa	197628867	3e-22
Thermococcus barophilus MP	Thermococci	511 aa	197629016	3e-33
Thermococcus onnurineus NA1	Thermococci	520 aa	212224590	3e-32
Thermococcus sibiricus MM 739	Thermococci	371 aa	242398189	2e-27
Thermococcus sibiricus MM 739	Thermococci	512 aa	242398190	9e-35

Most prokaryotic sequences show a similar organization as the protein sequence of animal D-glucuronyl C5-epimerases. An N-terminal signal peptide is predicted for many sequences and at the C-terminus of the protein the conserved pfam06662 domain can be found. This architecture resembles that of animal D-glucuronyl C5-epimerases. A multiple sequence alignment of the C-terminal domain of the prokaryotic candidate C5-epimerase and a selection of animal C5-epimerases is included as supplementary figure 1.

It is tempting to speculate on the role of those amino acid residues that are completely conserved. Residues possibly involved in catalysis are the conserved tyrosines and histidines. Structural data of two other types of C5-epimerases, not homologous to the heparin C5-epimerase (Valla *et al*., 2001) and the prokaryotic candidate epimerases, reveal a role of conserved histidines and tyrosines in catalysis for both functionally distinct C5-epimerases (Rozeboom *et al*., 2008; Pacheco *et al*., 2008). Although there is no homology at amino acid level, a similar catalytic mechanism of the heparin-acting C5-epimerase to these distinct C5-epimerases cannot be ruled out and could be feasible.

We constructed a phylogenetic tree (Huson; 2007) containing several eukaryotic D-glucuronyl C5-epimerases, as well as a selection of prokaryotic homologs (Fig.2). The multiple sequence alignment (Notredame *et al*, 2000; Thompson *et al*., 1997) is mostly based on the C-terminus of the genes as to be seen in supplementary figure 1.

No remarkable differences were observed when constructing a tree of the full length sequences or of the C-terminus only. Phylogenetically, the prokaryotic candidate C5-epimerase sequences cluster in a domain-specific way (Fig.2). Most deviation is observed in bacterial sequences, while archaea and eukaryotes are more alike. Obvious inter-domain substitutions can not be observed, and are not expected to have occurred recently. Few bacterial sequences do cluster with eukaryotes and archaea but these are close to the root and bootstrap values are too low to draw any conclusions.

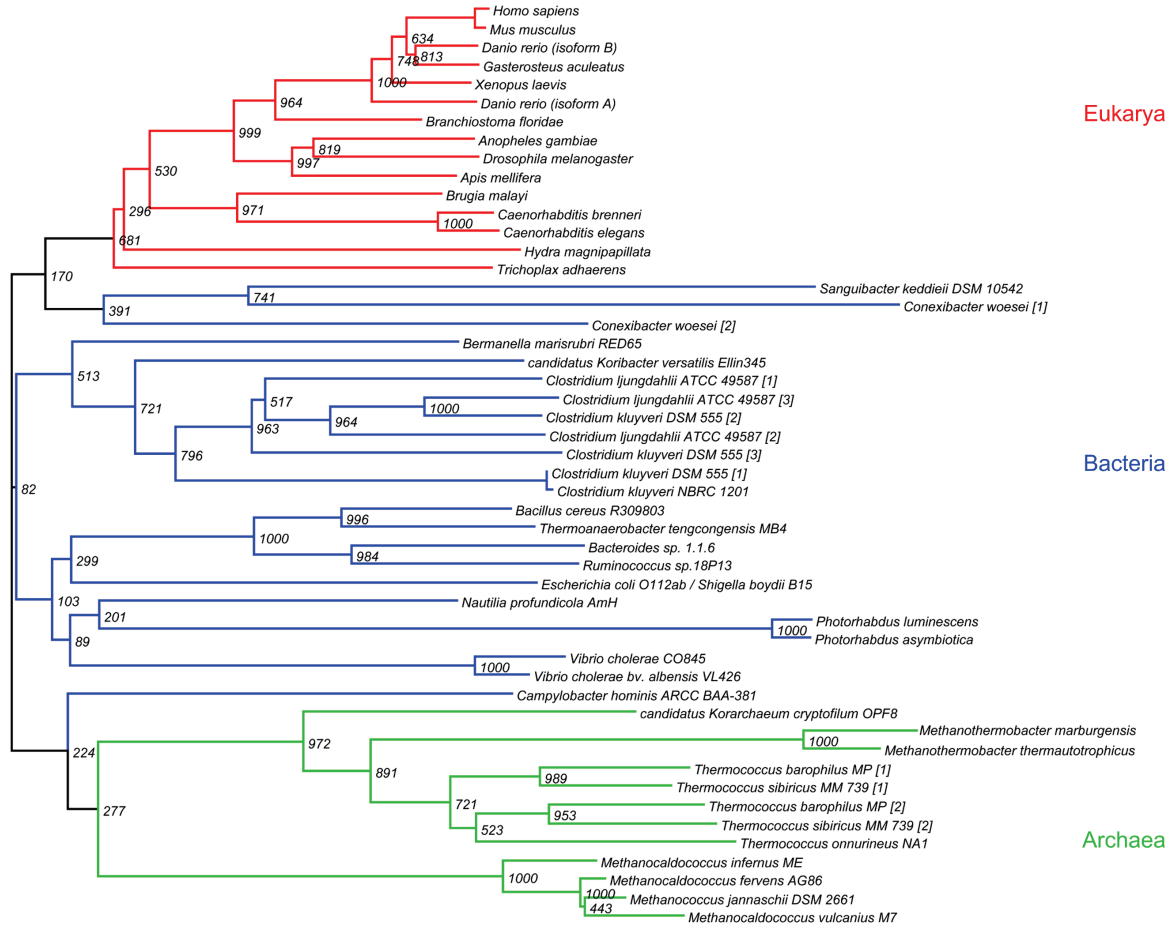


Figure 2 – Phylogenetic analysis candidate C5-epimerases. The coloring of branches is domain specific; eukaryotes in red, bacteria in blue and archaea in green. The used bootstrap value is 1000.

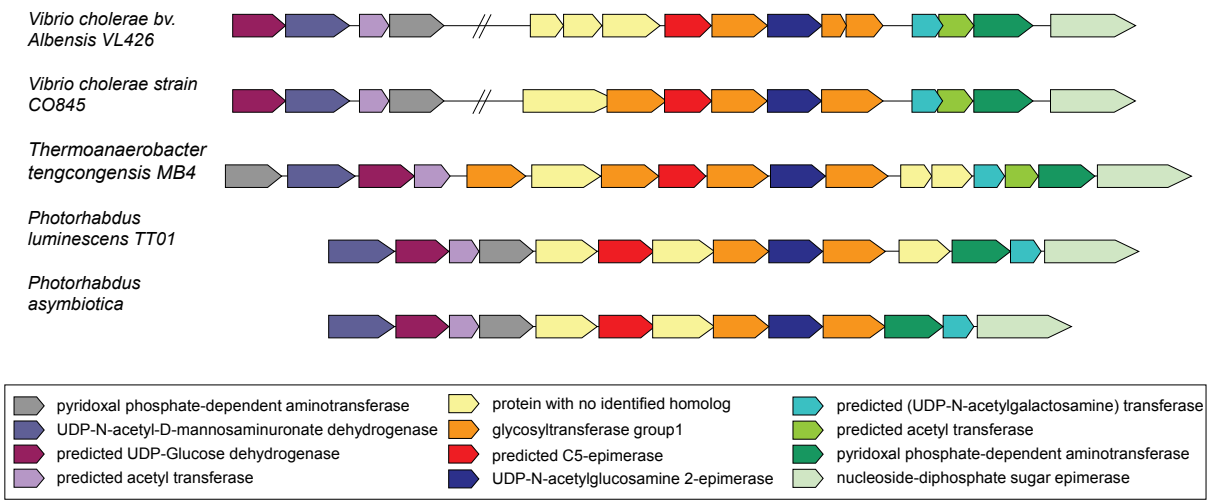


Figure 3 – Neighborhood analysis candidate C5-epimerases.

On an intra-domain level one could speculate on the clustering. Considering a confidence level of 70% or higher, the archaeal genes cluster in two clades. Surprisingly the two *Methanothermobacter* species do not cluster with the *Methano(caldo)cocci*. The subset of bacterial sequences gives rise to several clades, each not necessarily clustered in a class specific way. An example is seen for *Thermoanaerobacter, Bacillus* and *Ruminococcus* that do cluster with *Bacteroides* and not cluster with the *clostridiae* genes, even though they are all in the class of firmicutes. Instead the *clostridiae* genes cluster with an *Acidobacterium*. It is tempting to speculate that these deviations reflect the occurrence of variant enzymes (paralogs) with a different substrate specificity. This is expected because of the wide diversity that exists in bacterial cell wall polysaccharides. Most likely the candidate C5-epimerases are involved in the biosynthesis of type-specific polysaccharides.

The association of the putative candidate C5-epimerases with other (predicted) sugar modifying enzymes (e.g. glycosyltransferases) is clearly revealed upon neighborhood analysis of the involved prokaryotic genomes (Fig.3). Genes in this gene cluster are likely to be involved in strain specific O-antigen production, since some of the sequences have homology to the *wbb* operon (Derzelle *et al*., 2004), which is known to be involved in the O-antigen biosynthesis (Reeves *et al*., 1996). As LPS do not occur in gram positives, alternatively these gene clusters can also be involved in the biosynthesis of a capsular polysaccharide. This cell wall structure can occur in gram positive bacteria like *T. tengcongensis*. The exact gene function is hard to distinguish, since genes involved in polysaccharide capsule biosynthesis are sometimes embedded in the other cell wall biosynthesis related gene regions (e.g. LPS) and vice versa (Chen *et al*., 2007).

A remarkable similarity in genomic organization is seen for some of the putative C5-epimerases. Figure 3 shows the flanking genes up- and downstream the candidate C5-epimerases of two *Vibrio cholerae* strains (albensis VL426 and CO845), the thermophilic bacterium *Thermoanaerobacter tengcongensis* MB4 and two *Photorhabdus* species. Both *Photorhabdus luminescens* TT01 and *Photorhabdus asymbiotica* are symbiotic pathogens of insects, although *P. asymbiotica* is occasionally found as an opportunistic pathogen of humans as well. *V. cholera* is a well-known human pathogen, causing cholera. No virulence activity is reported for *T. tengcongensis* (Bao *et al*., 2002).

For most of these prokaryotic candidate C5-epimerases, the gene is in close proximity of various sugar modifying enzymes. The analogy in gene neighborhood organization is

remarkably similar for the above mentioned five bacteria. Other bacterial candidate C5-epimerases have a different organization of their gene neighborhood, despite the fact that they are more related to any of these five species with respect to taxonomy or candidate C5-epimerases sequence identity. A clear example is seen for the candidate C5-epimerase of *Bacillus cereus* R309803. This sequence is the best hit compared to *T. tengcongensis* (54% identity, 70% similarity). The homology of the *T. tengcongensis* sequence to the putative C5-epimerases from *Photorhabdus* (wblE) is rather low (15% identity, 31% similarity). Homology with *V. cholera* is better (23% identity, 44% similarity), but still weak compared to the best hit.

Naturally, addition of more sequences would give a better understanding of the exact sequence distribution and a better view on the number of different clades existing. However, it is obvious that within this subset of sequences a clustering in different clades exists, which is not necessarily class-specific. An explanation of this phylogenetic distribution might be the existence of several C5-epimerase paralogs. Given the enormous variety in type-specific cell wall polysaccharides, it is expected that enzymes involved in biosynthesis have a substrate optimized specificity. Variations in substrate specificity of involved D-glucuronyl C5-epimerase would certainly be feasible.

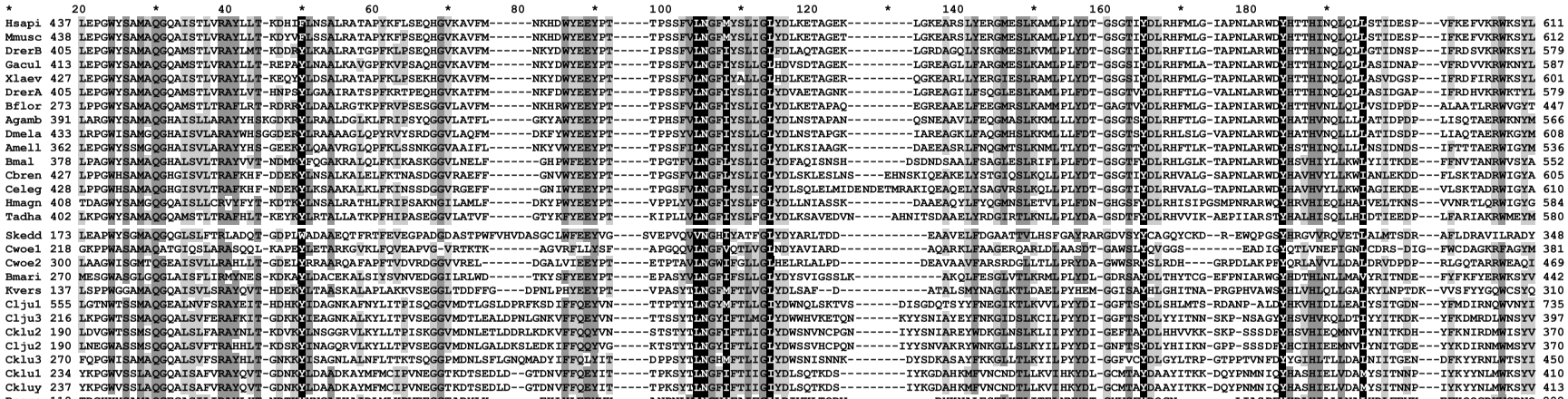
Conclusions

The C5-epimerization of D-glucuronic acid to its C5-epimer L-iduronic acid has long been considered typical for animal derived glycosaminoglycans. However, an increasing number of L-iduronic acid containing structures in microorganisms can be confidently identified in prokaryotes as well. Moreover, we found multiple candidate D-glucuronyl C5-epimerases in a wide variety of microbes by *in silico* analysis of available prokaryotic genome data. Gene neighborhood analysis of these sequences suggests a role in sugar modification, most likely in type-specific polysaccharides (e.g. capsule polysaccharides or O-antigens). Phylogenetic analysis indicates sub-clustering of the set of candidate D-glucuronyl C5-epimerases into several clades. This possibly correlates with the existence of different C5-epimerase paralogs each having a distinct substrate specificity. The exact physiological function and substrate specificity requires biochemical analysis. However, this subset of sequences and Blast analysis of metagenomes reveal the existence of multiple candidate C5-epimerase genes in

prokaryotes, supporting the conclusion that L-iduronic acid most likely is less rare in prokaryotes than expected. These putative C5-epimerases potentially may become important tools in controlled chemo-enzymatic synthesis of L-iduronic acid containing polymers like heparin and heparin-analogs.

Acknowledgements

We would like to thank Gijs van Dedem for critically reading the manuscript. This project is carried out in close collaboration with MSD Oss (formally Organon N.V.) and financially supported by the Netherlands Ministry of Economic Affairs and the B-Basic partner organizations (www.b-basic.nl) through B-Basic, a public–private NWO-ACTS program (Advanced Chemical Technologies for Sustainability).



Supplementary figure 1 - Multiple sequence alignment of C-terminal part of the candidate D-glucuronyl C5-epimerases. The positions of the first and last residues of the aligned region of the corresponding candidate C5-epimerase are indicates for each sequence. Names are abbreviated and in the same order as in the phylogenetic tree (Fig.2).

References

Altschul SF, Madden TL, Schäffer AA, Zha ng J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**, 3389-3402.

Bao Q, Tian Y, Li W, Xu Z, Xuan Z, Hu S, Dong W, Yang J, Chen Y, Xue Y, Xu Y, Lai X, Huang L,

Dong X, Ma Y, Ling L, Tan H, Chen R, Wang J, Yu J, Yang H. (2002). A complete sequence of the *T. tengcongensis* genome. *Genome Research* **12**, 689-700

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. (2004). The Pfam protein families database. *Nucleic Acids Research*. **32**, D138-141.

Carroll AR, Duffy S, Avery VM. (2009). Citronamides A and B, tetrapeptides from the Australian sponge *Citronia astra. Journal of natural products* **72**, 764-768.

Chen Y, Bystricky P, Adeyeye J, Panigrahi P, Ali A, Johnson JA, Bush CA, Morris JG Jr, Stine OC. (2007). The capsule polysaccharide structure and biogenesis for non-O1 *Vibrio cholerae* NRT36S: genes are embedded in the LPS region. *BMC Microbiology* **7**:20.

Darby GK, Jones AS, Kennedy JF, Walker RT. (1970). Isolation and analysis of the nucleic acids and polysaccharides from *Clostridium welchii. Journal of Bacteriology* **103**(1), 159-165.

DeAngelis PL, Gunay NS, Toida T, Mao W, Linhardt RJ. (2002). Identification of the capsular polysaccharides of Type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively. *Carbohydrate Research* **337**(17), 1547-1552.

Derzell S, Turlin E, Duchaud E, Pages S, Kunst F, Givaudan A, Danchin A. (2004). The PhoP-PhoQ two-component regulatory system of *Photorhabdus luminescens* is essential for virulence in insects. *Journal of Bacteriology* **186**(5), 1270-1279.

Hagner-McWhirter A, Hannesson HH, Campbell P, Westley J, Rodén L, Lindahl U, Li J. (2000). Biosynthesis of heparin/heparan sulfate: kinetic studies of the glucuronyl C5-epimerase with N-sulfated derivatives of the *Escherichia coli* K5 capsular polysaccharide as substrates. *Glycobiology* **10**(2), 159-171.

Hanniffy OM, Shashkov AS, Senchenkova SN, Tomshich SV, Komandrova NA, Romanenko LA, Knirel YA, Savage AV. (1998). Structure of a highly acidic O-specific polysaccharide of lipopolysaccharide of *Pseudoalteromonas haloplanktis* KMM 223 (44-1) containing L-iduronic acid and D-QuiNHb4NHb. *Carbohydrate Research* **307**, 291-298.

Höök M, Lindahl U, Bäckström G, Malmström A, Fransson LA. (1974). Biosynthesis of heparin. III. Formation of L-iduronic acid residues. *The Journal of Biological Chemistry* **249**, 3908-3915.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics* **8**: 460.

Jarrell KF, Bayley DP, Kostyukova AS. (1996). The archaeal flagellum: a unique motility structure. *Journal of Bacteriology* **178**(17), 5057–5064.

Kusche M, Hannesson HH, Lindahl U. (1991). Biosynthesis of heparin. Use of *Escherichia coli* K5 capsular

polysaccharide as a model substrate in enzymic polymer-modification reactions. *Biochemical Journal* **275**, 151-158.

Lahaye M, Robic A. (2007). Structure and functional properties of ulvan, a polysaccharide from green seaweeds. *Biomacromolecules* **8**, 1765-1774.

Lee L, Cherniak R. (1974). Identification of iduronic acid as a constituent of the "type-specific" polysaccharide of *Clostridium perfringens* Hobbs 10. *Carbohydrate Research* **33**, 387-390.

Lindahl U, Bäckström G, Malmström A, Fransson LA. (1972). Biosynthesis of L-iduronic acid in heparin: Epimerization of D-glucuronic acid on the polymer level. *Biochemical and Biophysical Research Communications* **46**(2), 985-991.

Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Wang Q, Reeves PR, Wang L. (2008). Structure and genetics of Shigella O antigens. *FEMS Microbiology Reviews* **32**, 627-653.

Miyasaki T, Tsuchihashi H, Yamada H, Yodomae, T. (1979). Detection and identification of L-iduronic acid in the glycuronan "protuberic acid" from *Koboyasia nipponica. Carbohydrate Research* **77**, 281-284.

Nazarenkoa EL, Komandrovaa NA, Gorshkovaa RP, Tomshicha SV, Zubkova VA, Kilcoyneb M, Savage AV. (2003). Structures of polysaccharides and oligosaccharides of some Gram-negative marine *Proteobacteria*. *Carbohydrate Research* **338**(23) 2449-2457.

Notredame C, Higgins DG, Heringa J. (2000). T-Coffee: A novel method for multiple sequence alignments. *Journal of Molecular Biology*, **302**, 205-217.

Ørskov I, Ørskov F, Jann B, Jann K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli. Bacteriological Reviews* **41**(3), 667-710.

Ozaki N, Sakuda S, Nagasawa H. (2007). A novel highly acidic polysaccharide with inhibitory activity on calcification from the calcified scale "coccolith" of a coccolithophorid alga, *Pleurochrysis haptonemofera. Biochemical and Biophysical Research Communications* **357**, 1172-1176.

Pacheco B, Maccarana M, Goodlett DR, Malmström A, Malmström L. (2008). Identification of the active site of DS-epimerase 1 and requirement of N-glycosylation for enzyme function. *The Journal of Biological Chemistry* **284**(3), 1741-1747.

Perepelov AV, Liu B, Senchenkova SN, Shashkov AS, Feng L, Knirel YA, Wang L. (2008). Structure of the O-polysaccharide of *Escherichia coli* O112ab containing L-iduronic acid. *Carbohydrate Research* **343**(3), 571-575.

Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, Rick PD. (1996). Bacterial polysaccharide synthesis and gene nomenclature. *Trends in Microbiology* **4**(12), 495-503.

Rimler RB. (1994). Presumptive identification of *Pasteurella multocida* serogroups A, D and F by capsule depolymerisation with mucopolysaccharidases. *Veterinary Record* **134**, 191-192.

Rodriguez M, Jann B, Jann K. (1988). Structure and serological characteristics of the capsular K4 antigen of *Escherichia coli* O5 : K4 : H4, a fructose-containing polysaccharide with a chondroitin backbone.

European Journal of Biochemistry **177**, 117-124.

Rosner H, Grimmecke HD, Knirel YA, Shashkov AS. (1992). Hyaluronic acid and a (1→4)-β-D-xylan, extracellular polysaccharides of *Pasteurella multocida* (Carter type A) strain 880. *Carbohydrate Research* **223**, 329-333.

Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvåg H, Holtan S, Aachmann FL, Valla S Dijkstra BW. (2008) Structural and mutational characterization of the catalytic A-module of the Mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii. The Journal of Biological Chemistry* **283**, 23819-23828.

Stack RJ, Plattner RD, Cote GL. (1988). Identification of L-iduronic acid as a constituent of the major extracellular polysaccharide produced by *Butyrivibrio fibrisolvens* strain X6C61. *FEMS Microbiology Letters* **51**, 1-5.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876-4882.

Tsuchihashi H, Yadomae T, Miyazaki T. (1983). Structure of L-idurono-D-glucuronans from the fungi *Aseroe araclmoidea* and *Pseudocolas fusiformis. Carbohydrate Research* **108**(1) 123-128.

Valla S, Li J, Ertesvåg H, Barbeyron T, Lindahl U. (2001). Hexuronyl C5-epimerases in alginate and glycosaminoglycan biosynthesis. *Biochimie* **83**(8), 819-830.

Vann WF, Schmidt MA, Jann B, Jann K. (1981). The structure of the capsular polysaccharide (K5 antigen) of urinary-tranct-infective *Escherichia coli* O10:K5:H4. A Polymer Similar to Desulfo-Heparin. *European Journal of Biochemistry* **116**, 359-364.

Whitfield C. (1995) Biosynthesis of lipopolysaccharide O antigens. *Trends in Microbiology* **3**(5), 178-185.

Wieland F, Paul G, Sumper M. (1985). Halobacterial flagellins are sulfated glycoproteins. *The Journal of Biological Chemistry* **260**(28), 15180-15185.

Wieland F, Lechner J, Sumper M. (1986). Iduronic acid: constituent of sulphated dolichyl phosphate oligosaccharides in halobacteria. *FEBS Letters* **195**, 77-81.

Chapter 3

Evidence for D-glucuronyl C5-epimerase activity in the bacterium *Bermanella marisrubri* sp. RED65

Manuscript in preparation:

Raedts J, Li JP, Lundgren M, Kengen SWM, van der Oost J.

Abstract

Glycosaminoglycans are specific biologically active polysaccharides that are widespread among the animal kingdom. However, also some type-specific prokaryotic cell wall polysaccharides have been identified which show resemblance to these animal glycosaminoglycans. Moreover, a unique constituent of glycosaminoglycans, L-iduronic acid, has been shown to exist in prokaryotes as well. However, until now there is no experimental proof of the existence of a prokaryotic D-glucuronyl C5-epimerase, the enzyme responsible for the conversion of D-glucuronic acid (GlcA) into its C5-epimer L-iduronic acid (IdoA). We cloned a selection of putative prokaryotic D-glucuronyl C5-epimerases, to test them for D-glucuronyl C5-epimerase activity. We confirmed C5-epimerase activity for one of these proteins, a hypothetical protein of the marine bacterium *Bermanella marisrubri* sp. RED65.

Introduction

Glycosaminoglycans (GAGs) are carbohydrates that are widely spread in the animal kingdom in various animal tissues, involved in many biological functions. GAG were believed to occur exclusively in animals, however today there is increasing evidence that also GAG of prokaryotic origin do exist. Typically, those bacterial GAGs are less complex and modifications as seen for animal GAGs appear to be absent in bacterial GAGs. The C5-epimerization of D-glucuronic acid to L-iduronic acid is such a modification. In animal GAGs this modification leads to increased flexibility of the GAG chain (Casu et al., 1988), facilitating binding properties for specific polysaccharide-protein interactions (Jia *et al*., 2009).

Previously, we performed an *in silico* screen on available prokaryotic genomes for putative D-glucuronyl C5-epimerases, which revealed multiple candidate C5-epimerases in both archaea and bacteria (Chapter2; Raedts *et al*., 2011). Here we describe the cloning and expression experiments of some of these candidate D-glucuronyl C5-epimerases encoding genes. The candidate C5-epimerase identified in the bacterium *Bermanella marisrubri* sp. RED65 (RED-C5-epimerase) has by far the highest similarity with animal C5-epimerases compared to the other identified candidate C5-epimerases.

Material and Methods

L-iduronic acid was purchased from Toronto Research Chemicals Inc. (North York, Canada), capsular polysaccharide of *E. coli* K5 ('K5 substrate') was ordered from Iduron (Manchester, United Kingdom). All other chemicals were bought from Sigma-Aldrich. Restriction enzymes, Pfu DNA polymerase and T4 DNA ligase were from Invitrogen and New England Biolabs.

Gene cloning and expression

Candidate C5-epimerase *Bermanella marisrubri* | A synthetic gene was designed based on hypothetical protein RED65_08024, as genomic DNA was not available. The codon usage optimized for the codon bias of *Escherichia coli*. The gene was cloned in vector pRSF-1b (Novagen) using restriction sites *Kpn*I and *Bam*HI, fusing the gene in frame with an amino-terminal polyhistidine tag. The resulting vector, named pWUR537, was used to transform *E. coli* strains DH5α and Rosetta (DE3) containing pRARE (helper plasmid rare tRNAs).

An overnight culture of *E. coli* Rosetta (DE3) containing pWUR537 was used to inoculate (1% v/v) a shake flask with 1 liter Luria Bertani medium supplemented with 50

µg/ml kanamycin, 50 µg chloramphenicol, and 1 mM $MgSO_4$. At an optical density $A_{600} \approx 0.5$, gene expression was induced by adding 0.05 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). The culture was incubated overnight at 30°C. Cells were harvested by centrifugation (5,000 x g for 15 minutes), and stored at -20°C until further use.

Other candidate C5-epimerases | A synthetic gene was designed based on hypothetical protein Gae (*Escherichia coli* serogroup O112ab; GenBank ID: 187880578). The codon usage was optimized for the codon bias of *Escherichia coli*. The other candidate C5-epimerases were PCR amplified using genomic DNA as template using the following templates and corresponding primers (introduced restriction sites underlined):

- *Escherichia coli* serotype O112ab candidate C5-epimerase (ECO-C5-epimerase):
  synthetic gene "hypothetical protein Gae";
  BG2987 (5'-GCGGAATTCATGAAAGAAGCAACCACCTCC-3') sense,
  BG2988 (5'-CGCCTGCAGTTACAGGCGAAACACCAC-3') antisense.
- *Thermoanaerobacter tengcongensis* MB4 candidate C5-epimerase (TTE-C5-epimerase):
  "hypothetical protein TTE0659" (GI:20671658; region 661205-662158);
  BG2497 (5'-GCGGGTACCTTACATGTTAACCAGGGAATTGGAAAAATTTACTC-3') sense,
  BG2206 (5'-CGCGGATCCCTACTTAATTAATACAACCTCTCCGGGTTCC-3') antisense.
- *Methanocaldococcus jannaschii* DSM 2661 candidate C5-epimerase (MJA-C5-epimerase):
  "conserved hypothetical protein MJ_1128" (GI:6626255; region 1069044-1069970);
  BG2207 (5'-GCGCCATGGCATATTTTGTTGGA GCATCTCAACCTTTATATTCTG-3') sense,
  BG2208 (5'-CGCGGATCCCTATTCTCTCCATTTTTTGTAGTATTTTAGGTATATCTC-3') antisense.
- *Methanothermobacter thermautotrophicus* ΔH candidate C5-epimerase (MTH-C5-epimerase): "conserved protein MTH_330" (GI:6626257; region 265312-266529);
  BG2209 (5'-GCGCCATGGATGACCCCCAGGAGGTAAACAGC-3') sense,
  BG2210 (5'-CGCGGATCCTTATATCCTTTCGGGTTTGAGTCCCTTCC -3') antisense.

PCR amplified genes were cloned in an expression vector (Gae in pMAL, TTE0659 in pRSF, MJ_1128 and MTH_330 in pET24d) using the introduced restriction sites. The resulting plasmids were used to transform the construction host *Escherichia coli* DH5α. The generated plasmid was mini-prepped (Qiagen) and sequenced (BaseClear) to verify correct cloning. As

an expression host *Escherichia coli* BL21 (DE3) supplemented with pSJS1244 (helper plasmid, rare tRNAs) was used. Overnight cultures of *E. coli* BL21 (DE3) containing the generated expression vectors were used to inoculate (1% v/v) a shake flask with 1 liter Luria Bertani medium supplemented with 50 µg/ml kanamycin (100 µg/ml ampicillin for ECO-C5-epimerase) and 50 µg spectinomycin. At an optical density A_{600} ≈ 0.5, gene expression was induced by adding 0.5 mM IPTG. The culture was incubated overnight at 30°C. Cells were harvested by centrifugation (3,000 rpm for 15 minutes).

Protein purification

Candidate C5-epimerase *Bermanella marisrubri* (RED-C5-epimerase) | Pelleted *E. coli* cells were resuspended in 20 mM Tris-HCl buffer (pH 7.5) containing 200 mM NaCl (buffer A). The resuspended cells were disrupted by sonication (intermitted cooling on ice) after which cell debris was removed by a centrifugation step (16,000 x g for 10 minutes). The resulting cell free extract (CFE) was filtered (0.45 µm) and applied on a Talon Cobalt affinity column (Clontech). After washing with two column volumes of buffer A and one column volume buffer A with 10 mM imidazole, RED-C5-epimerase was eluted with one column volume of bufferA containing, in a linear gradient, 10 - 250 mM imidazole. RED-C5-epimerase containing fractions were pooled and applied onto a HiPrep desalting to remove imidazole. Enzyme presence and purity was checked by SDS-PAGE analysis. A broad range protein marker (Biorad) was used to estimate the molecular mass. The presence of RED-C5-epimerase was confirmed by LC-MS analysis.

Candidate C5-epimerase *Thermoanaerobacter tengcongensis* (TTE-C5-epimerase) | The pelleted *E. coli* cells were disrupted similarly as described above. Buffer A differed in composition being a 25 mM Tris-HCl buffer (pH 7.5) containing 250 mM NaCl. After applying on the Talon cobalt affinity column and washing with two column volumes buffer A, it was washed with one column volume buffer with 1 mM imidazole, and eluted with one column volume buffer A that contained, in a linear gradient, 10 - 500 mM imidazole. Removal of imidazole, and enzyme presence and purity was performed identically as described above.

Size exclusion chromatography

The oligomeric state of both RED-C5-epimerase and TTE-C5-epimerase was determined by

size exclusion chromatography on a Superdex 200 high-resolution 10/10 column (24 ml) (Amersham Biosciences) equilibrated in 20 mM Tris-HCl (pH 7.5) containing 200 mM NaCl. Two hundred microliters of enzyme solution was loaded onto the column, using a flow rate of 0.5 ml per minute. Proteins used for calibration were blue dextran 2000 (>2000 kDa), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa) bovine serum albumin (67 kDa), ovalbuvin (43 kDa), chymotrypsinogen A (25 kDa), and ribonuclease A (13.7 kDa).

HPLC analysis

Enzyme activity was determined by HPLC-PAD (pulse amperometric detection) analysis of the formed reaction products using a Dionex system. Either cell lysate or purified protein sample of a candidate C5-epimerase was mixed with N-Acetyl-de-O-sulfated heparin (sodium salt; Sigma-Aldrich) or any other substrate of choice. The mixture was incubated at 30°C for the desired reaction time; typically overnight. Subsequently, protein was heat inactivated (100°C) and removed from the reaction mixture by a centrifugation step (16,000 x g for 10 minutes). The reaction mixture was hydrolyzed (90°C for 4 hours) by addition of one volume of hydrochloric acid to an end concentration of 1 M. Hydrochloric acid was removed by evaporation in a speedvac, and hydrolyzed samples were redissolved in Millipore water. Final traces of hydrochloric acid were removed by titration with sodium hydroxide. Samples were analyzed for their hexuronic acid content by anion exchange chromatography on a Dionex system equipped with a CarboPac-100. The used effluent was 45 mM sodium hydroxide, with an increasing gradient of sodium acetate (0-960 mM). The used detection method is based on PAD (electrocatalytically oxidation); to prevent loss of signal due to electrode poisoning, a series of pulsed potentials is applied.

Tritium labeled substrate

Either freshly prepared cell lysate or purified protein sample of a candidate C5-epimerase was mixed with tritium labelled N-Acetyl-de-O-sulfated heparin (sodium salt; Sigma-Aldrich) or K5 capsular polysaccharide. The enzyme/substrate mixture was incubated at 30°C overnight. Subsequently, reaction mixtures were mixed with a cocktail solution for biphasic separation (Li, 2010). The C5-epimerase activity could be measured by measuring tritium release using a biphasic liquid scintillation procedure as described before (Campbell, 1983).

Results

A BLAST-search (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et al*., 1997) with the human heparan sulfate D-glucuronyl C5-epimerase protein sequence as query, indicated high homology with other animal heparan sulfate D-glucuronyl C5-epimerases. However, as described in chapter 2, also homologous sequences can be identified in genomes of both bacteria and archaea, although with a relatively low sequences homology (Raedts *et al*., 2011). We selected some of these prokaryotic candidate C5-epimerase genes from different clades of the phylogenetic three, for protein expression studies. An overview of all cloned genes and constructed plasmids is given in Table 1.

Table 1 – Overview of cloned and expressed candidate C5-epimerases

Name microorganism used as template	Signal peptide predicted	N-terminal protein tag	pWUR number
Thermoanaerobacter tengcongensis MB4	yes	histidine	pWUR536
Methanothermobacter thermautotrophicus ΔH	yes	none	pWUR611
Methanocaldococcus jannaschii DSM 2661	yes	none	pWUR612
Escherichia coli serotype O112ab	no	MBP	pWUR574
Bermanella marisrubri sp. RED65	no	histidine	pWUR537

Among this group of prokaryotic sequences, the candidate C5-epimerase sequence (RED-C5-epimerase) of the marine bacterium *Bermanella marisrubri* sp. RED65 (Pinhassi *et al*., 2009) showed significantly higher homology to animal heparan sulfate D-glucuronyl C5-epimerase than any other prokaryotic sequence. Based on the sequence, RED-C5-epimerase is placed within PFAM family C5-epim_C (PF0662), representing the C-terminal domain of heparan sulfate D-glucuronyl C5-epimerase (http://pfam.sanger.ac.uk/family?entry=PF06662; Bateman *et al*., 2004). The gene coding for RED-C5-epimerase (RED65_08024) contains 1347 base pairs, resulting in a protein of 448 amino acids with a theoretical molecular mass of 51,914 Da.

We constructed a multiple sequence alignment (T-Coffee server; http://www.tcoffee.org; Notredame *et al*, 2000), to illustrate the homology between RED-C5-epimerase and the animal heparan sulfate D-glucuronyl C5-epimerases (Fig.1). The multiple sequence alignment was visualized using GeneDoc version 2.6 (Nicholas *et al*.,

1997). RED-C5-epimerase has 22% identity and 36% similarity with human heparan sulfate D-glucuronyl C5-epimerase, mostly based on high similarity at the carboxy terminus (40% identity, 59% similarity). In contrast with the animal proteins, RED-C5-epimerase has no transmembrane helix at the amino terminus. However, the overall secondary structure is predicted to be rather similar to the animal heparan sulfate D-glucuronyl C5-epimerase (PSIPRED Server http://www.psipred.net/psiform.html; McGuffin *et al*., 2000).

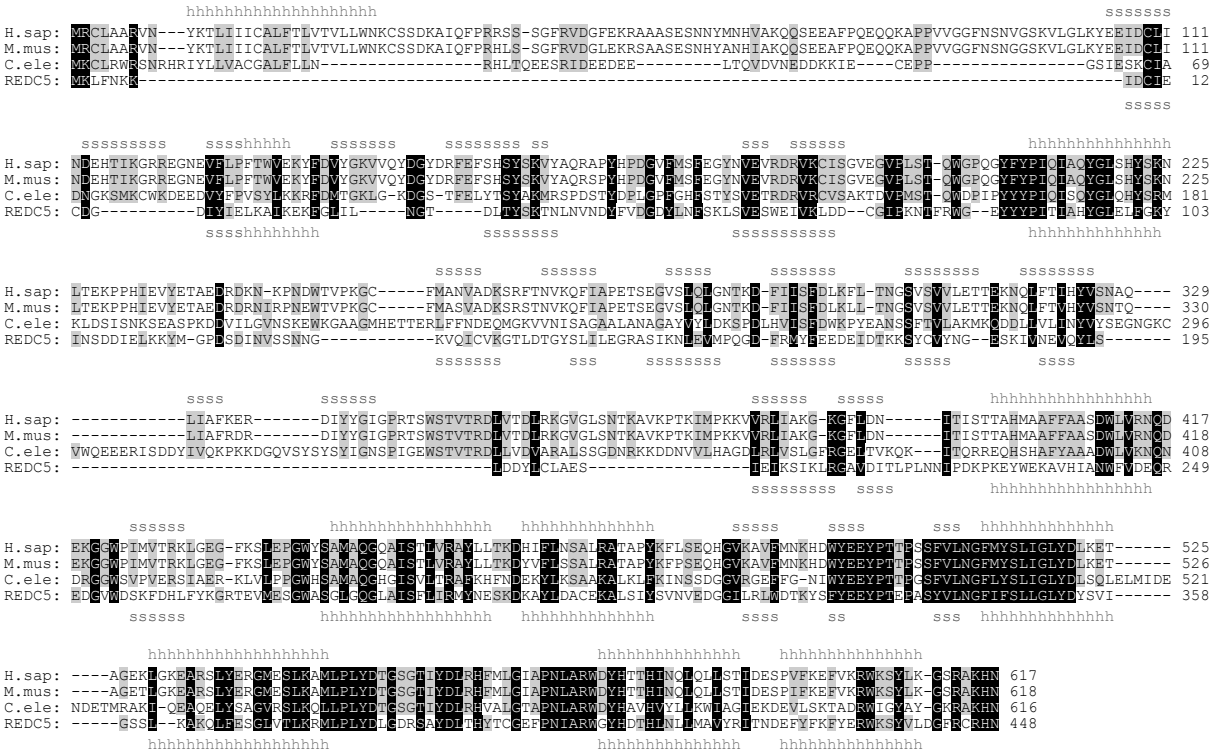


Figure 1 – Multiple sequence alignment of *Bermanella marisrubri* sp. Red65 RED-C5-epimerase (RED65) (Locus: RED_08024) compared to characterized animal heparan sulfate D-glucuronyl C5-epimerases (H.sap: *Homo sapiens*; M.mus: *Mus musculus*; C.ele: *Caenorhabditis elegans*). Secondary structure predictions are indicated for human and *B. marisrubri* (h indicates predicted α-helices, s indicates predicted β-strands).

Protein production studies

As for the animal proteins, putative signal sequences were present in many of the prokaryotic sequences (SignalP; Bendtsen *et al*., 2004). However, these have not been included in the PCR amplicons, resulting in truncated proteins starting at the predicted cleavage site. For most candidate C5-epimerases we added a protein tag to facilitate purification. Addition of such protein tag always was at the N-terminus, as research on the murine heparan sulfate D-glucuronyl C5-epimerase showed a dramatic reduction in activity

when the enzyme had a C-terminal tag, while an N-terminal tag had no obvious effect on activity (Crawford *et al*., 2001).

The constructed plasmids were used to transform *E. coli* BL21(DE3). Every candidate C5-epimerase was expressed successfully resulting in significant quantities of protein. However, we experienced the solubility of these proteins not to be very good. Nevertheless, most constructs could be partly expressed as soluble protein, taking care of tightly controlled expression conditions with relatively low expression time, temperature and IPTG concentrations. Only the candidate C5-epimerase from *E. coli* (ECO-C5-epimerase) was completely insoluble. Therefore we constructed a fusion of Maltose Binding Protein (MBP) to the N-terminus of ECO-C5-epimerase, resulting in soluble production of the fusion protein. Although highly speculative, possibly the MBP compensates for the absence of the other members of the LPS biosynthesis pathway. Enzymes involved in the biosynthesis of cell surface polysaccharides, typically are associated to the inner membrane of the cell wall as part of in a bigger complex (review: Whitfield, 1995). Based on our findings ECO-C5-epimerase is likely to be part of such a membrane spanning protein complex together with other members of the O-antigen biosynthesis pathway. This is in contrast to previous findings that propose ECO-C5-epimerase as a cytoplasmic enzyme (Liu *et al*., 2008). This hypothesis is based on the absence of a signal sequence in ECO-C5-epimerase. Indeed, no signal sequence is predicted (SignalP 3.0 Server http://www.cbs.dtu.dk/services/SignalP; Bendtsen *et al*., 2004), although a highly hydrophobic domain is predicted in the amino acid sequence, potentially involved in the association with other proteins.

Size exclusion chromatography

To determine the oligomeric state we performed size exclusion chromatography and applied purified samples of both the bacterial candidate C5-epimerases on a Superdex 200 high-resolution 10/10 column. The oligomeric state of ECO-C5-epimerase could not be determined, since the fusion protein tended to form soluble aggregates of large undefined size. The oligomeric size of TTE-C5-epimerase could be determined as monomeric as the elution occurred in one major peak with a molecular mass of 28 kDa.

The oligomeric state of RED-C5-epimerase is larger than monomeric. In freshly prepared protein samples RED-C5-epimerase is present in a dimeric/tetrameric form, they occur in about a 2:1 ratio, and appears to be a dynamic equilibrium. During storage RED-C5-

epimerase has the tendency to form larger aggregates. Non-detergent sulfobetaines (NDSB) are known to efficiently prevent protein aggregation (Collins *et al*., 2006). We used up to 1 M of NDSB201 and although we could not see any obvious effect in prevention of aggregation at 4°C, NDSB201 did prevent aggregation upon freezing. Also addition of reducing agents (2 mM Tris(2-carboxyethyl)phosphine) did help to prevent aggregation, and helped maintaining the initial oligomeric state. High levels of reducing agent (e.g. 50 mM β-mercaptoethanol) even partly reversed aggregation. This effect possibly can be explained as prevention of the formation of intramolecular and intermolecular disulfide bonds between any of the nine cysteine residues in the amino acid sequence of RED-C5-epimerase.

Neighborhood analysis

To gain insight in the physiological role of RED-C5-epimerase we performed a gene neighborhood analysis. The genomic context of RED65_08024 comprises genes involved in biosynthesis of surface polysaccharides. In bacteria these genes generally are organized in gene clusters. Studies on the organization of similar clusters in *Escherichia coli* (Clarke *et al*., 1999) or *Pasteurella multocida* (Chung *et al*., 1998), show that these clusters consist of three regions. Region 1 and 3 are conserved and involved in capsular polysaccharide export, while region 2 is involved in the biosynthesis of serotype specific polysaccharides, and therefore is highly variable (Whitfield, 2006). RED65_08024 and another nine ORFs, are flanked by these capsular polysaccharide export regions 1 and 3 (Fig.2).

We compared the genomic context of RED-C5-epimerase with the organization of similar genes in other bacteria. We observed high resemblance with the capsular gene clusters of the aquatic bacteria *Shewanella violacea* DSS12 (Aono *et al*., 2010), *Aeromonas hydrophila* PPD134/91 (Zhang *et al*., 2002) and *Burkholderia ambifaria* IOP40-10. Not only do they all possess the region 1 and 3 genes involved in capsular polysaccharide transport as seen in *Escherichia coli* K10 and *Pasteurella multocida* X-73, they also have several genes in common of the type specific region 2. However, the presence of a C5-epimerase is unique to *Bermanella marisrubri* sp. RED65. Most likely this candidate C5-epimerase is involved in a species- or type-specific sugar modification. This would be in agreement with what would be expected for a bacterial D-glucuronyl C5-epimerase. As discussed in chapter 2, all demonstrations of L-iduronic acid being present in prokaryotes are known to be species- or type-specific constituents.

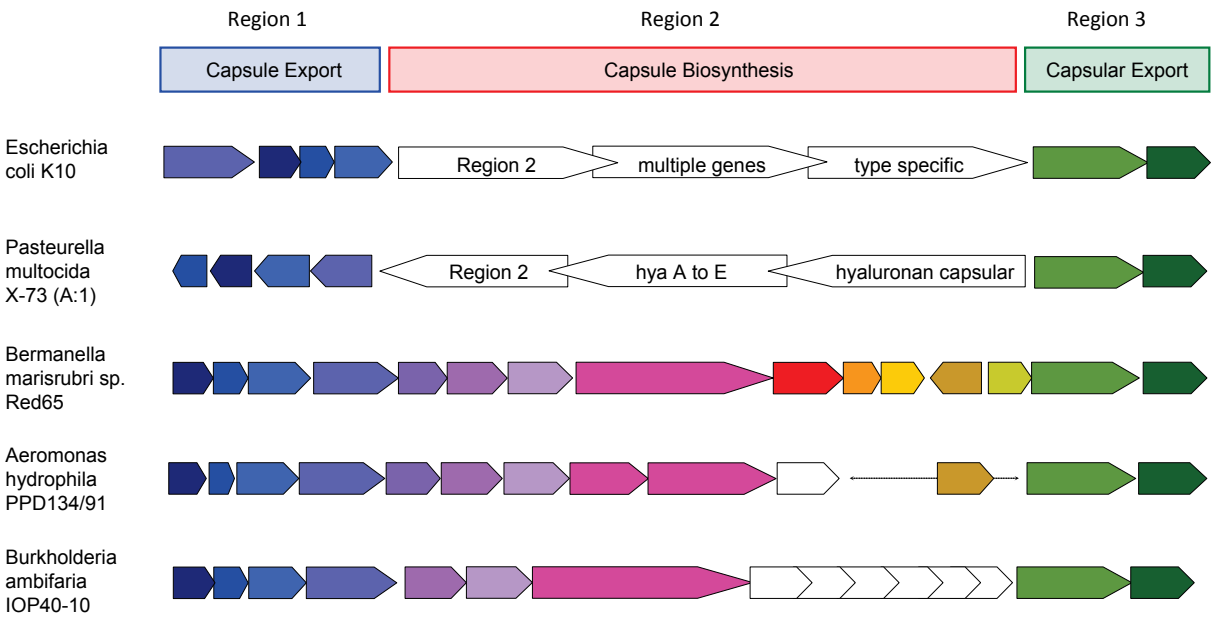


Figure 2 – Genomic context of Red_08024 (in red) in comparison with other bacterial capsular polysaccharide biosynthesis gene clusters. Flanking genes in blue (region 1; kpsM, kpsT, kpsE and kpsD) and green (region3; kpsC and kpsS) are likely to be involved in polysaccharide export. The genes in between (region 2) are likely to be involved in serotype specific capsular polysaccharide biosynthesis. The conserved genes are predicted UDP-N-acetyl glucosamine-2-epimerase (purple), UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase (violet) and glycosyltransferase (magenta). The D-glucuronyl C5-epimerase (red) is unique to *B. marisrubri* Sp. RED65.

Enzyme activity studies

Analysis of hexuronic acids constituents in polysaccharides is easiest after hydrolysis of the corresponding polysaccharide. Generally enzymatic hydrolysis is preferable over chemical hydrolysis, since less side products are formed. Several heparinases have been characterized to date, each having a specific substrate specificity (Desai *et al*., 1993). Unfortunately none of them are applicable for hexuronic acid analysis since all these enzymes are lyases rather than hydrolases (Hovingh and Linker, 1970). As a result of the beta-elimination reaction on the glycosidic bond, both D-glucuronic acid and L-iduronic acid are converted to 4,5-unsaturated hexuronic acid.

Chemical hydrolysis with HCl does result in partial hydrolysis of heparin, liberating monomeric hexuronic acids. Tight control of hydrolysis time, temperature and acid concentration are crucial to prevent substantial formation of side products and to prevent differential release of D-glucuronic acid and L-iduronic acid, due to differences in acid lability of their glycosidic bonds (Conrad, 1980). Therefore a control sample is included in each run to validate the hydrolysis conditions.

Monomeric D-glucuronic acid can be separated from L-iduronic acid by high performance anion exchange chromatography (HPAE) (Whitfield *et al*., 1991; De Ruiter *et al*., 1992). We optimized the protocol for analysis of the hydrolyzed reaction mixture on a Dionex system, monitoring hexuronic acids with an electrochemical detector using pulsed amperometric detection (PAD) (Fig.3A).

Using an anion exchange column (CarboPac PA-100) we were able to separate D-glucuronic acid and L-iduronic acid. However, it is important to keep salt-levels low since the retention time of L-iduronic acid is lowered at increasing salt concentrations. With hardly any salt effect on the retention time of D-glucuronic acid, best is to keep salt concentrations as low as possible in the samples, to prevent dramatic decrease in separation of both hexuronic acids.

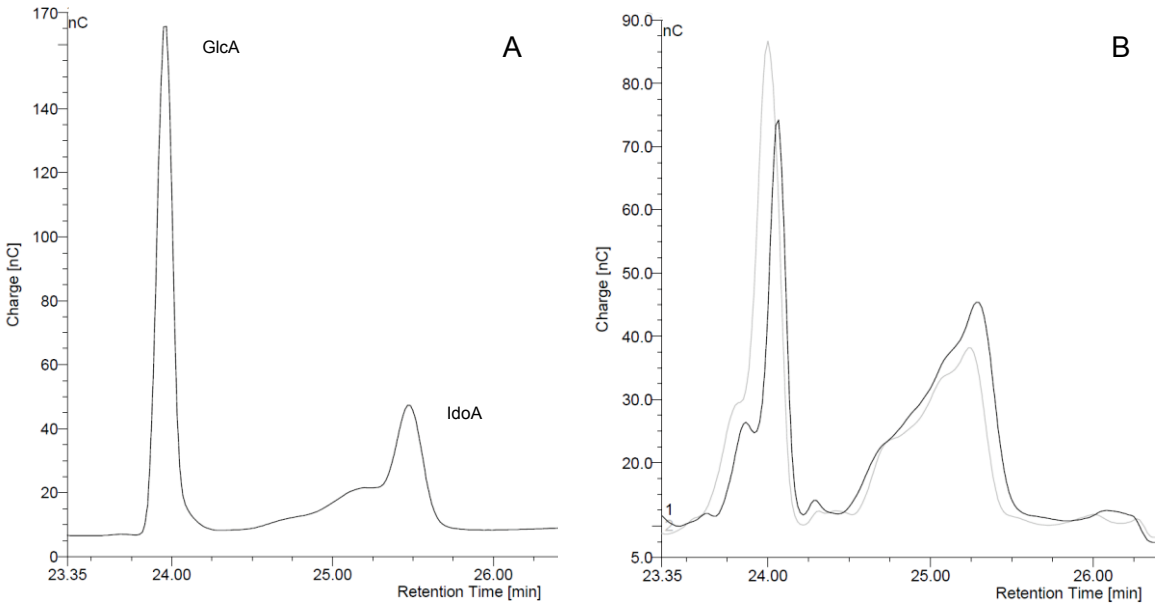


Figure 3 – Separation of D-glucuronic acid (GlcA) and L-iduronic acid (IdoA) using an HPCL-PAD system. (A) Separation of standards (B) Hydrolyzed (de-O-sulfated) heparin; black = control sample, gray = sample incubated with RED-C5-epimerase.

Using this method we tested C5-epimerase activity of the selected candidate C5-epimerases on various available D-glucuronyl containing GAGs and GAG precursors. Tested substrates included chondroitin, hyaluronan, UDP-GlcA, N-sulfated K5 and de-sulfated / re-acetylated heparin. Most of the samples did not give any significant changes in the hexuronic acid content of the hydrolysate, except for the RED-C5-epimerase. We tested the effect on de-O-

sulfated heparin, both by addition as crude cell lysate and purified enzyme. In heparin the ratio between both hexuronic acids is typically 35% D-glucuronic acid and 65% L-iduronic acid. However, overnight incubations with RED-C5-epimerase resulted in a shift in ratio of both hexuronic acids where L-iduronic acid levels drop to 45% and D-glucuronic acid increase to 55% (Fig.3B and Fig.4). Control samples were treated identically and only differ in the cell lysate composition, which lacks RED-C5-epimerase.

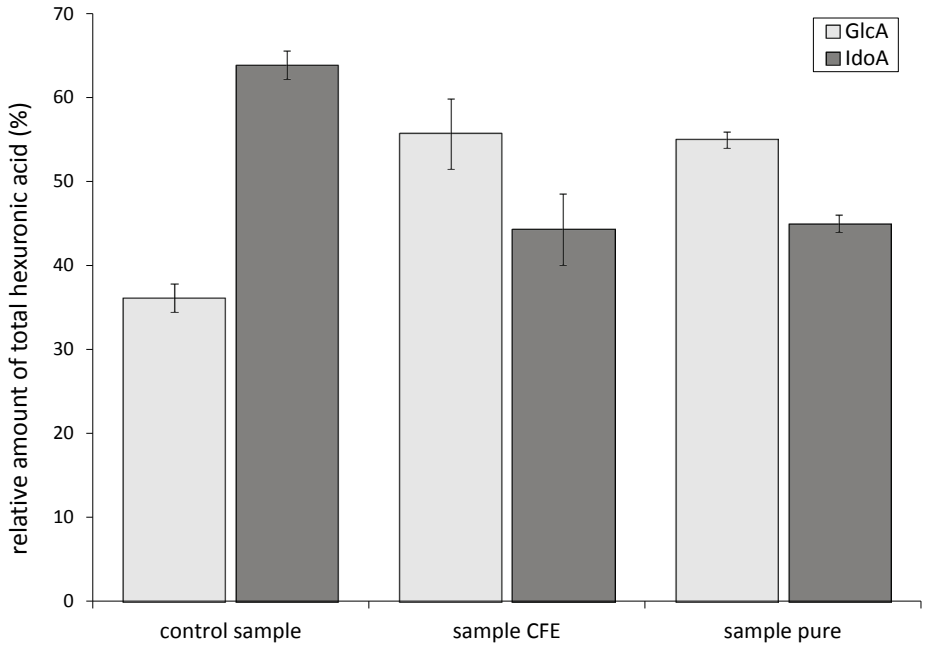


Figure 4 – Hexuronic acid ratio in heparin before and after treatment with RED-C5-epimerase. Both cell lysate and pure protein show partial epimerization of L-iduronic acid (IdoA) into D-glucuronic acid (GlcA).

We performed a time series experiment to have a better monitoring of the activity. These time series showed a gradually decrease in the conversion rate of the hexuronic acids until an equilibrium was reached with a ratio of about 55% D-glucuronic acid and 45% L-iduronic acid (Fig.5).

Elongation of the reaction time did not result in more net conversion of L-iduronic acid, neither did addition of higher amounts of enzyme. The reverse reaction was tested by mixing RED-C5-epimerase with heparonan (K5 polysaccharide). We could detect some trace levels of L-iduronic acid although the amounts were rather low, and therefore were considered not conclusive.

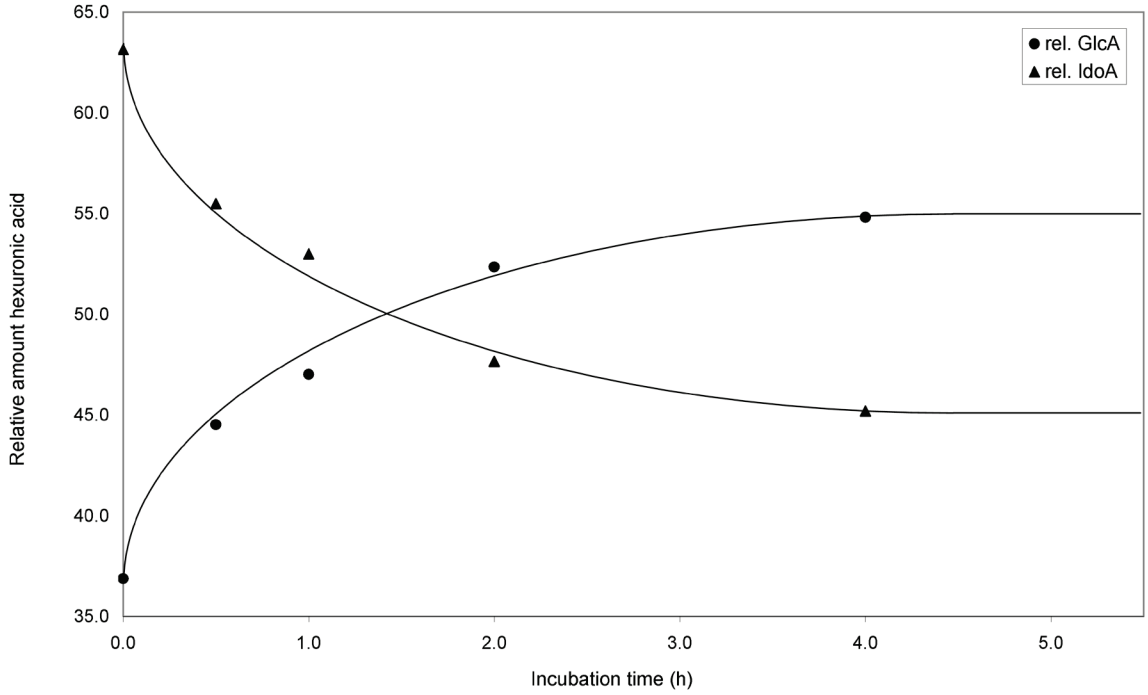


Figure 5 – Analysis of the hexuronic acid content of heparin as a result of RED-C5-activity.

To confirm this D-glucuronyl C5-epimerase activity we performed an additional screen based on the tritium release method using ^{3}H-labeled polysaccharide substrate (Campbell *et al*., 1983; Kusche *et al*., 1991). We tested RED-C5-epimerase activity on two different tritium labeled substrates, K5 polysaccharide (De-acetylated/N-sulfated) and heparin and all measurements were performed in duplo (Table 2).

Table 2 – Activity measurement of D-glucuronyl C5-epimerase mediated tritium release. When no significant activity could be determined N.A. is stated.

Substrate	Enzyme	Buffer	^{3}H release (cpm)	
Heparin	RED-C5-epimerase	HEPES	N.A.	
Heparin	RED-C5-epimerase	TRIS-HCl	64	±29
Heparin	Control (D-glucuronyl C5-epimerase)	HEPES	274	±55
K5 polysaccharide	RED-C5-epimerase	HEPES	N.A.	
K5 polysaccharide	RED-C5-epimerase	TRIS-HCl	35	±23
K5 polysaccharide	Control (D-glucuronyl C5-epimerase)	HEPES	27	±20

For both substrates we detected significant D-glucuronyl C5-epimerase activity using RED-C5-epimerase, although this activity was only observed using a TRIS-HCl buffer prepared with TCEP and NDSB-201. A HEPES buffer lacking TCEP and NDSB-201 did not give any significant activity. Activity was low compared to the murine heparan sulfate D-glucuronyl C5-epimerase which was used as a control. The fact that we were unable to detect activity of C5-epimerase in absence of TCEP and NDSB-201 is most likely due to oxygen related inactivation upon storage.

Discussion

L-iduronic is a common and essential sugar in several animal glycosaminoglycans. Previous *in silico* analyses (Chapter 2) demonstrate that L-iduronic acid not only is typical for eukaryotic GAGs, but occasionally also can be found as a constituent in type specific prokaryotic cell wall polysaccharides. From glycosaminoglycan biosynthesis we know that L-iduronic acid is synthesized via a C5-epimerization of D-glucuronic acid (review: Li, 2010), a rather common sugar that is found in most life forms. L-iduronic acid in prokaryotes is likely to be synthesized in a similar way, which implies the involvement of a D-glucuronyl C5-epimerase.

We selected five putative D-glucuronyl C5-epimerases based on a previously analysis (Raedts *et al*., 2011). The selected genes were cloned and expressed in *E. coli* and tested for D-glucuronyl C5-epimerase activity. For all but one of the produced candidate C5-epimerases, we have not been able to show conversion of D-glucuronic acid / L-iduronic acid. Point of discussion could be whether this is due to alternative substrate requirements or inadequate reaction conditions. With MBP-ECO-C5-epimerase as an exception, we have no knowledge on the capsular / lipopolysaccharide composition and as such nothing is known about the substrate specificity of the different candidate C5-epimerases. Identification of these cell wall polysaccharides could be a first step in providing insight in both the physiological role and substrate acceptance of these candidate C5-epimerases.

The candidate C5-epimerase from the marine bacterium *Bermanella marisrubri* sp. RED65 is an exception on this, as we were able to measure conversion of hexuronic acid levels in the glycosaminoglycan heparin. For hydrolyzed (de-O-sulfated) heparin / RED-C5-epimease reaction mixtures we detected a conversion of L-iduronic acid into D-glucuronic acid (Fig.4 and Fig.5). This makes us speculate that RED-C5-epimerase has similar catalytic properties and results in similar conversion ratios as the animal heparan sulfate D-glucuronyl

C5-epimerase. An epimerization reaction typically ends up in an equilibrium between the two specific epimers. For animal D-glucuronyl C5-epimerases it has been postulated that the epimerization equilibrium is slightly towards D-glucuronic acid over L-iduronic acid (Hagner-McWhirter *et al*., 2000).

In case of the animal GAGs, the subsequent O-sulfation results in irreversibility of the epimerization reaction *in vivo*, and thus in accumulation of L-iduronic acid to levels up to 90% of the total hexuronic acid amount (Hagner-McWhirter *et al*., 2004). The de-O-sulfated heparin that we used as substrate does not have such a "pull" of the equilibrium and therefore is likely to result in a state of equilibrium upon addition of a D-glucuronyl C5-epimerase. This equilibrial state is clearly shown in figure 5. The observed conversion must be an effect of the RED-C5-epimerase, as control-samples do not show the shift in hexuronic acid ratio. No further conversion was observed after prolongated incubation or after addition of more enzyme.

We were unable to proof significant L-iduronic acid levels for the RED-C5-epimerase / K5 substrate incubations using HPLC-PAD analysis. However, tritium labeled substrate assays resulted in low but significant levels of tritium release, indicating reversible D-glucuronyl C5-epimerase activity. Likely the K5 polysaccharide is a poor substrate for the enzyme; likewise for the murine heparan sulfate D-glucuronyl C5-epimerase it is known that activity on the K5 polysaccharide also is less than on heparin (Li, personal communication).

As far as we know this is the first example showing D-glucuronyl C5-epimirase activity in prokaryotes. Most likely, the existence of more prokaryotic D-glucuronyl C5-epimerases is anticipated, as several L-iduronic acid containing cell wall polysaccharides have been described (Raedts *et al*., 2011). Unfortunately the responsible D-glucuronyl C5-epimerases have not been identified yet, as genomic data is generally not available for the corresponding organisms. This study reveals that prokaryotic candidate C5-epimerases such as RED-C5-epimerase potentially are applicable as a biotechnological tool in the controlled chemo-enzymatic synthesis of L-iduronic acid containing glycosaminoglycans as heparin.

Acknowledgement

The authors would like to thank Henk Schols and Willy van den Berg (Wageningen University, the Netherlands) for their help in initiating the HPLC-PAD analysis.

References

Altschul SF, Madden TL, Schäffer AA, Zha ng J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**, 3389-3402.

Aono E, Baba T, Ara T, Nishi T, Nakamichi T, Inamoto E, Toyonaga H, Hasegawa M, Takai Y, Okumura Y, Baba M, Tomita M, Kato C, Oshima T, Nakasone K, Mori H. (2010). Complete genome sequence and comparative analysis of *Shewanella violacea*, a psychrophilic and piezophilic bacterium from deep sea floor sediments. Molecular Biosystems **6**(7), 1216-1226.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. (2004). The Pfam protein families database. *Nucleic Acids Research* **32**, D138-141.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783-795.

Casu B, Petitou M, Provasoli M, Sinaÿ P. (1988). Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans. *Trends in Biochemical Sciences* **13**(6), 221-225.

Campbell P, Feingold DS, Jensen JW, Malmström A, Rodén L. (1983). New assay for uronosyl 5-epimerases. *Analytical Biochemistry* **131**(1), 146-152.

Chung JY, Zhang Y, Adler B. (1998). The capsule biosynthetic locus of *Pasteurella multocida* A:1. *FEMS Microbiology Letters* **166**, 289-296.

Clarke BR, Pearce R, Roberts IS. (1999). Genetic organization of the *Escherichia coli* K10 capsule gene cluster: identification and characterization of two conserved regions in group III capsule gene clusters encoding polysaccharide transport functions. *Journal of Bacteriology* **181**(7), 2279-2285.

Collins T, D'Amico S, Georlette D, Marx JC, Huston AL, Feller G. (2006). A nondetergent sulfobetaine prevents protein aggregation in microcalorimetric studies. *Analytical Biochemistry* **352**(2), 299-301.

Crawford BE, Jeffrey SK, Esko JD, Pinhal MAS. (2001). Cloning, Golgi Localization, and Enzyme Activity of the Full-length Heparin/Heparan Sulfate-Glucuronic Acid C5-epimerase. *The Journal of Biological Chemistry* **276**, 21538-21543.

Conrad HE. (1980). The acid lability of the glycosidic bonds of L-iduronic acid residues in glycosaminoglycans. *Biochemical Journal* **191**, 355-363.

de Ruiter GA, Schols HA, Voragen AG, Rombouts FM. (1992). Carbohydrate analysis of water-soluble uronic acid-containing polysaccharides with high-performance anion-exchange chromatography using methanolysis combined with TFA hydrolysis is superior to four other methods. *Analytical Biochemistry* **207**, 176-185.

Desai UR, Wang HM, Linhardt RJ. (1993). Substrate specificity of the heparin lyases from *Flavobacterium heparinum*. *Archives of Biochemistry and Biophysics* **306**(2), 461-468.

Hagner-Mcwhirter A, Lindahl U, Li JP. (2000). Biosynthesis of heparin/heparan sulphate: mechanism of

epimerization of glucuronyl C-5. *Biochemical Journal* **347**, 69-75.

Hagner-McWhirter A, Li JP, Oscarson S, Lindahl U. (2004). Irreversible Glucuronyl C5-epimerization in the biosynthesis of Heparan sulfate. *The Journal of Biological Chemistry* **279**(15), 14631-14638.

Hovingh P, Linker A. (1970). The enzymatic degradation of heparin and heparitin sulfate. 3. Purification of a heparitinase and a heparinase from flavobacteria. *The Journal of Biological Chemistry* **245**(22), 6170-6175.

Jia J, Maccarana M, Zhang X, Bespalov M, Lindahl U, Li JP. (2009). Lack of L-iduronic acid in heparan sulfate affects interaction with growth factors and cell signaling. *The Journal of Biological Chemistry* **284**(23), 15942-15950.

Kusche M, Hannesson HH, Lindahl U. (1991). Biosynthesis of heparin. Use of *Escherichia coli* K5 capsular polysaccharide as a model substrate in enzymic polymer-modification reactions. *Biochemical Journal* **275**, 151-158.

Li JP. (2010). Glucuronyl C5-epimerase an enzyme converting glucuronic acid to iduronic acid in heparan sulfate/heparin biosynthesis. *Progress in molecular biology and translational science* **93**, 59-78.

Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Wang Q, Reeves PR, Wang L. (2008). Structure and genetics of Shigella O antigens. *FEMS Microbiology Reviews* **32**, 627-653.

McGuffin LJ, Bryson K, Jones DT. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*. **16**, 404-405.

Nicholas KB, Nicholas HB. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. *Distributed by the author*

Notredame C, Higgins DG, Heringa J. (2000). T-Coffee: A novel method for multiple sequence alignments. *Journal of Molecular Biology* **302**, 205-217.

Pinhassi J, Pujalte MJ, Pascual J, González JM, Lekunberri I, Pedrós-Alió C, Arahal DR. (2009). *Bermanella marisrubri* gen. nov., sp. nov., a genome-sequenced gammaproteobacterium from the Red Sea. *International Journal of Systematic and Evolutionary Microbiology* **59**(2), 373-377.

Raedts J, Kengen SW, van der Oost J. (2011). Occurrence of L-iduronic acid and putative D-glucuronyl C5-epimerases in prokaryotes. *Glycoconjugate journal* **28**(2), 57-66.

Whitfield C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annual Review of Biochemistry* **75**, 39-68.

Whitfield DM, Stojkovski S, Pang H, Baptista J, Sarkar B. (1991). Diagnostic methods for the determination of iduronic acid in oligosaccharides. *Analytical Biochemistry* **194**, 259-267.

Zhang YL, Arakawa E, Leung KY. (2002). Novel *Aeromonas hydrophila* PPD134/91 genes involved in O-antigen and capsule biosynthesis. *Infection and Immunity* **70**(5), 2326-2335.

Chapter 4

Microbial production of the murine heparan sulfate D-glucuronyl C5-epimerase

Manuscript in preparation:

Raedts J, Creutzburg S, Li JP, Lundgren M, Kengen SWM, van der Oost J.

Abstract

The epimerization of D-glucuronic acid into L-iduronic acid is an essential step in the biosynthesis of heparin. The gene coding for the responsible C5-epimerase from mouse has been identified and cloned previously (Li *et al*, 1997). Although the corresponding enzyme has *in vitro* epimerization activity, still many restrictions exist in order to use this D-glucuronyl C5-epimerase in an industrial process. Protein production levels, number of accepted substrates and *in vitro* activity are all rather low. Therefore we aimed to set up C5-epimerase production in a microbial expression system. A synthetic *Mus musculus* D-glucuronyl C5-epimerase gene was used to transform two different expression hosts. Protein expression levels were tested in both the yeast *Pichia pastoris* and the bacterium *Escherichia coli* BL21(DE3). No detectable levels of C5-epimerase expression were observed in *P. pastoris*. Successful bacterial expression was achieved. Initial solubility problems of the C5-epimerase could be overcome by fusion of the C5-epimerase to maltose bind protein (MBP). D-glucuronyl C5-epimerase activity was confirmed indicating an active fusion protein.

Introduction

Glycosaminoglycans (GAGs) are a specific class of biological active carbohydrates that are present throughout the animal kingdom. A well-known example is heparin. This GAG has a major pharmaceutical application as an anti-coagulation drug. It is a polymer that is highly negatively charged due to the presence of sulfate- and carboxyl residues. The sulfatation step can occur at multiple positions, while the carboxyl group is part of a hexuronic acid. Typically two types of hexuronic acids can be found in GAGs; D-glucuronic acid and its C5-epimer L-iduronic acid. Variation in the relative amounts of these two significantly influences the physiological properties of the carbohydrate chain. This makes the C5-epimerization an essential modification in glycosaminoglycans like heparin.

The conversion of D-glucuronic acid to L-iduronic acid is catalyzed by a single enzyme. The responsible enzyme, a D-glucuronyl C5-epimerase, is conserved throughout the animal kingdom (Fig.1). This enzyme is localized in the Golgi apparatus (Crawford *et al*., 2001) where it probably acts in a larger complex with other enzymes involved in heparin biosynthesis (Sasisekharan and Venkataraman, 2000). Previous studies have led to the (biochemical) characterization of the D-glucuronyl C5-epimerase. Early research was mainly on protein isolates from bovine tissue (Campbell *et al*., 1994). Later research shifted to recombinant M*us musculus* (mouse) D-glucuronyl C5-epimerase produced in insect cells co-transfected with baculovirus (Li *et al*., 2001) or Chinese hamster ovary cells (Crawford *et al*., 2001).

For long time production of the recombinant *M. musculus* D-glucuronyl C5-epimerase has been limited to animal or insect cell expression systems (Jalkanen *et al*., 2008). Functional expression of the enzyme in microbial expression systems has been problematic for long time, although recently progress has been made on yeast expression systems (Li, 2010) and bacterial expression systems (Li *et al*., 2010). Establishment of D-glucuronyl C5-epimerase production in a bacterial expression system would enhance rapid, large scale and less expensive production of this mammalian C5-epimerase. An additional advantage would be the possibility for enzyme engineering. Protein improvement using laboratory evolution generally needs large and easily culturable libraries, which makes *Escherichia coli* very attractive as a production host. Protein expression experiments of D-glucuronyl C5-epimerase in *E. coli* have been attempted before, resulting in production of large quantities of D-glucuronyl C5-epimerase. However, major solubility problems do exist for this protein

being expressed in this bacterial host, resulting in neither soluble nor functional protein.

Functional expression of a D-glucuronyl C5-epimerase is crucial when aiming for controlled synthesis of heparin. Most heparin modifications can be done in a chemical way, the only exception being the epimerization of D-glucuronic acid towards L-iduronic acid. To reach successful chemo-enzymatic synthesis of heparin a more robust and better expressed D-glucuronyl C5-epimerase is crucial. In this study for production of elevated quantities of a functional D-glucuronyl C5-epimerase in a microbial host. Two distinct expression systems are tested, each applicable in large scale production processes. This study describes the attempts to obtain improved solubilization and stabilization of C5-epimerase expression in *E. coli* by fusion to maltose binding protein (MBP). The MBP protein is highly soluble in *E. coli*, and is known to be a potential solubilizing factor for *E. coli* based protein production (Kapust and Waugh, 1999). Fusion of MBP to the amino terminus of D-glucuronyl C5-epimerase (MBP-MmC5) possibly increases the overall solubility/stability. In addition expression studies have been performed in the yeast *Pichia pastoris*. Based on bioinformatical predictions there are three putative N-glycosylation sites in the murine protein (Fig.1). These sites are potential bottlenecks for bacterial-based expression, since bacterial systems are generally unable to modify proteins post-translationally. The yeast *P. pastoris* is widely used as microbial expression host because of its ability for N-glycosylation similar to mammalian systems, and therefore possibly can improve protein expression.

Materials and methods

Chemicals were purchased from Sigma-Aldrich. L-iduronic acid was purchased from Toronto Research Chemicals. Restriction enzymes, Pfu DNA polymerase and T4 DNA ligase were from Invitrogen and New England Biolabs. All plasmids were constructed in *E. coli* DH5α. The used bacterial expression host was *E. coli* BL21 (DE3) (Novagen), containing the tRNA helper plasmid pSJS1244. The used bacterial expression vectors were pET24d (Novagen) and pMAL-C2 (New England Biolabs). For yeast expression we used strain *P. pastoris* GS115 and the expression vector pPIC9 (invitrogen).

Bioinformatics

Sequences coding for heparosan D-glucuronyl C5-epimerases were identified by performing BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et al*., 1997) using protein

sequences of the characterized C5-epimerases from mouse, human and *C. elegans*. Retrieved sequences were aligned in a multiple sequence alignment using the T-Coffee server (http://www.tcoffee.org; Notredame *et al*, 2000). The alignment was manually checked and adjusted when considered appropriate. The multiple sequence alignment was visualized using GeneDoc version 2.6 (Nicholas *et al*., 1997). A bootstrapped phylogenetic tree corrected for multiple substitutions, was generated with ClustalX version 1.83 (Thompson *et al*., 1997). The tree was displayed using the neighbor-joining method with TreeView version 1.6.5. (Page *et al*., 1996).

These sequences were scanned for conserved motifs against multiple databases by using InterProScan (http://www.ebi.ac.uk/Tools/InterProScan; Zdobnov and Apweiler, 2001). The amino terminus was screened for putative signal sequences using the SignalP 3.0 Server (http://www.cbs.dtu.dk/services/SignalP; Bendtsen *et al*., 2004). Putative N-glycosylation sites were identified using NetNGlyc 1.0 (Gupta *et al*., 2004). The protein secondary structure was predicted using the PSIPRED Server (http://www.psipred.net/psiform.html; McGuffin *et al*., 2000). Conserved protein folds were recognized by using the PHYRE Protein Fold Recognition Server (http://www.sbg.bio.ic.ac.uk/phyre/; Kelley and Sternberg, 2009).

Gene optimization and cloning

A synthetic gene was designed based on *M. musculus* (mouse) D-glucuronyl C5-epimerase (MmC5) mRNA (GI: 133892759; base pair 7-1863). The gene was optimized towards the codon bias of *P. pastoris* and generated by BaseClear (Leiden, the Netherlands). Restriction sites *Xho*I / *Eco*RI, introduced flanking the gene, were used to clone the synthetic gene in expression vector pPIC9. *E. coli* DH5α was transformed with the ligation mix to amplify the resulting plasmid named pWUR543. The truncated genes used for *E. coli* based expression were created by PCR using the previously mentioned synthetic MmC5 gene as template.

Three different sense primers used were for the amplification; BG2853 (truncation1; 5'-CCCGAATTCAGCGACAAAGCAATCCAATTTCCTCG-3'), BG3143 (truncation2; 5'-CGCGAATTCGTAGAAGTCCGTGACAGAGTCAAATGTATAAG-3') and BG3144 (truncation3; 5'-CGCGAATTCACTGCCCACATGGCTGCTTTCTTTGC-3'). The antisense primer was BG2854 (5'-CCCCTGCAGTTAGTTATGCTTTGCCCTACTTCCTTT-3'). The introduced flanking restriction sites *Eco*RI / *Pst*I (underlined) were used to clone the gene in the bacterial expression vector pMal-C2 (New England Biolabs). All plasmids were transformed into *E. coli* DH5α.

MmC5 production in *Pichia pastoris*

P. pastoris was transformed with pWUR543 according to the supplier 's protocol. Successful integration of the murine MmC5 gene in the *P. pastoris* genome was confirmed by PCR analysis. A starter culture was grown overnight at 30°C in a 100 ml baffled Erlenmeyer flask with 10 ml BMGY medium (Buffered Minimal Glycerol Yeast extract). Biomass was collected via a centrifuge step, and used to inoculate 50 ml of fresh BMMY medium (Buffered Minimal Methanol Yeast extract). To achieve sufficient aeration a 500 ml baffled flask was used, that was shaken briskly. The MmC5 gene is under control of the AOX1 promoter, therefore expression was induced by addition of 0.5% v/v methanol after which cultivation was continued for 24 hours at 30°C. After the induction phase the medium was separated from the biomass by a centrifugation step. Excreted proteins in the medium were enriched using a centrifugal protein concentrator with a cut-off of 30 kDa. A 10 mM HEPES buffer (pH7) containing 50 mM NaCl was used to wash the protein fraction. Both the enriched secreted protein fraction as a total cell pellet fraction were checked on protein content by SDS-PAGE analyses. As a control identical fractions of wild type *P. pastoris* were used. Five protein bands that size-wise could be the C5-epimerase were isolated from the gel, and analyzed by mass spectrometry for protein identification.

MmC5 production in *Escherichia coli*

Expression vectors pWUR544 (no tag), pWUR575 (truncation1), pWUR576 (truncation 2) and pWUR577 (truncation 3) were used to transform *E. coli* BL21(DE3) harboring pSJS1244. For each construct a single colony was used to prepare a starter culture by inoculating Luria-Bertani medium (LB) containing 100 µl/ml ampicillin and 50 µl/ml spectinomycin. The starter culture was grown overnight at 37°C while shaking. A larger culture was started by inoculation of 100 ml LB (similar antibiotic content) with 1% starter culture. The culture was grown at 37°C, while shaking, until OD_{600} = 0.6 was reached. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM, and growth was continued for another 3 hours while shaking at a temperature of 30°C. Cells were harvested by centrifugation, and resuspended in 10 ml of 10 mM HEPES buffer (pH7) with 50 mM NaCl. The cells were disrupted by sonication with intermitted cooling steps on ice. The crude cell lysate was centrifuged (16,000 x g) to remove cell debris. The supernatant was applied on a gravity flow column of amylose resin, and purified according

the manufacturer's protocol. All elution fractions that contained MBP-MmC5 were pooled, and checked by SDS-PAGE analysis. Purified MBP-MmC5 fusion proteins were used for inductively coupled plasma atomic emission spectroscopy (ICP-AES) and for activity measurements. Activity measurements were based on HPLC-PAD analysis and tritium release from labeled substrate as described in Chapter 3.

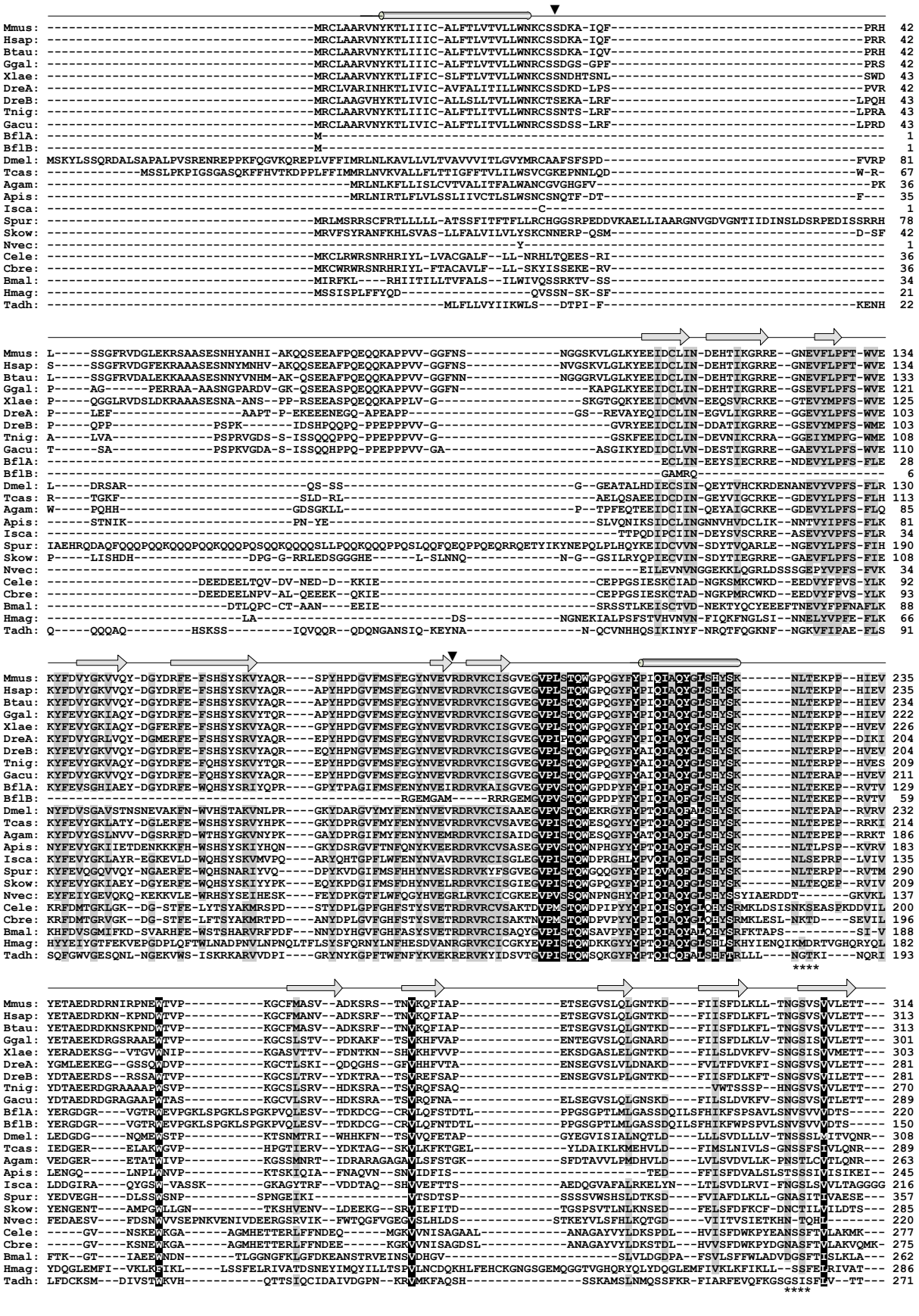
Results

Bioinformatics

To get a better understanding of the occurrence and conserved residues of heparosan D-glucuronyl C5-epimerases, we performed BLAST searches with protein sequence human C5-epimerase. These reveal candidate genes in almost all available animal genomes. In analogy to heparin, the presence of candidate C5-epimerase is not restricted to vertebrates that have a vascular blood system, but also can be found in many invertebrates (e.g. insects, nematodes, protozoa). An example where no homologous gene was found is the genome of the sponge *Amphimedon queenslandica*. While being virtually ubiquitous in animals, the C5-epimerase seems absent in plants and fungi, where no single candidate could be identified.

A number of sequences of evolutionary diverse animals have been selected to compose a multiple sequence alignment (Fig.1). Sequence homology is not uniform throughout the amino acid sequence. Least homology is seen at the amino-termini of the candidate C5-epimerases. SignalP analysis reveals an N-terminal signal peptide for most sequences. Therefore the first 32 amino acids where not included in the constructed recombinant proteins (see below). For the central part of the sequence three glycosylation sites are predicted; for the murine sequence these sites are Asp-225, Asp-304, and Asp-394 (Fig.1). Possibly this N-glycosylation is needed structurally.

In contrary to the more variable N-terminus, the carboxy-terminus of the C5-epimerases are highly conserved. All members have such a conserved C-terminal domain and are included in the family PFAM06662 (Bateman, *et al*., 2004). Secondary structures predictions indicate seven putative α-helixes in this domain. Based on predictions with PHYRE these helixes possibly are arranged in an alpha/alpha toroid fold. The alpha/alpha toroid fold comprises up to seven alpha-hairpins are arranged in closed circular array. Members that are proven to have such a fold include glycosidases and lyases. It is likely that the D-glucuronyl C5-epimerase protein shares structural homology with such a protein.



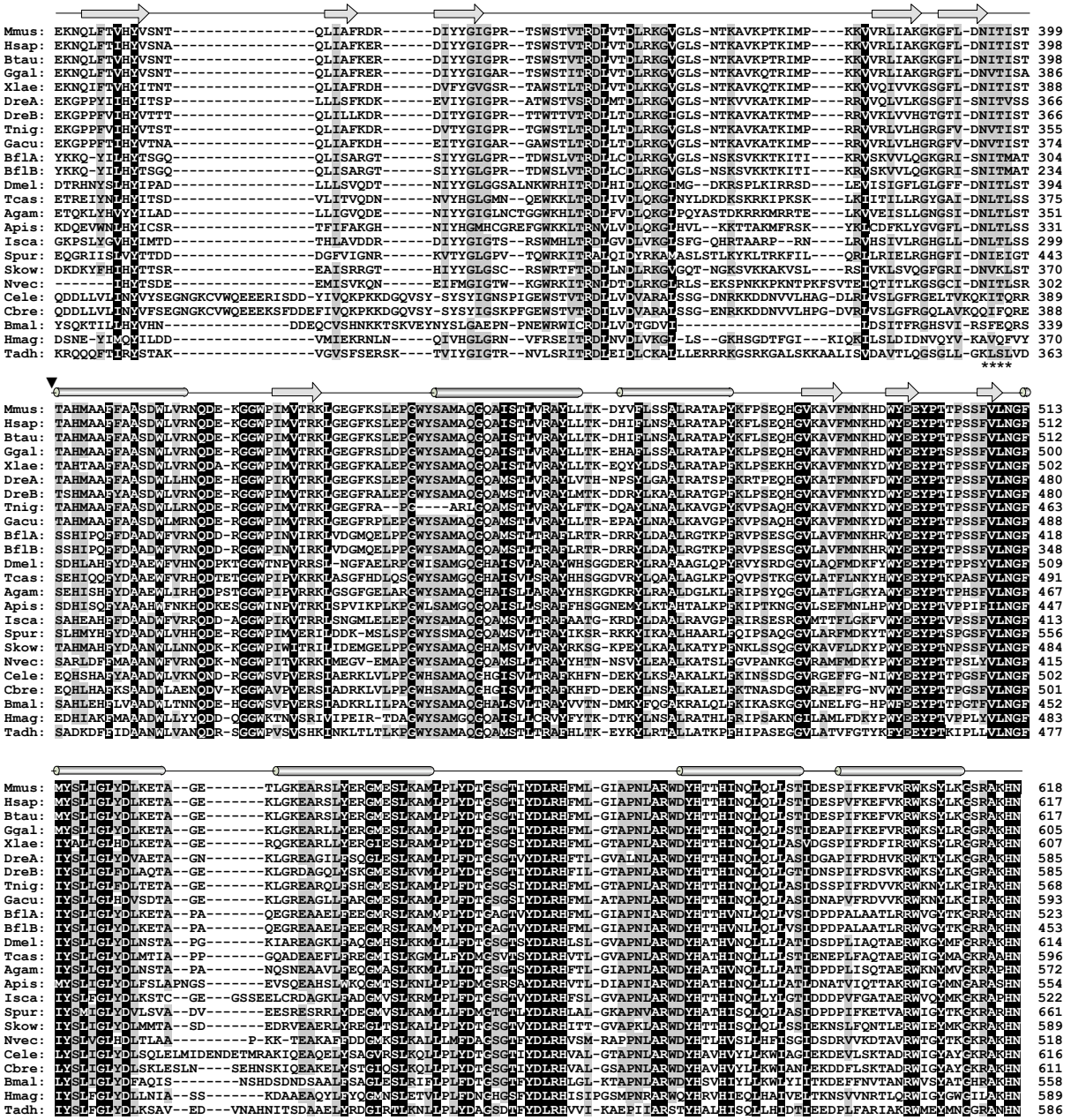


Figure 1 – Multiple sequence alignment of animal D-glucuronyl C5-epimerases. Putative N-glycosylation sites are indicated with asterisks (below sequence). The start positions of all made truncations are indicated as triangles. Secondary structure predictions are indicated as cylinders (α-helixes) and arrows (β-strands). Abbreviations: Mmus: *Mus musculus* (mouse); Hsap: *Homo sapiens* (human); Btau: *Bos taurus* (cow); Ggal: *Gallus gallus* (chicken); Xlae *Xenopus laevis* (frog); DreA: *Danio rerio* isoform A (zebrafish); DreB: *Danio rerio* isoform B (zebrafish); Tnig: *Tetraodon nigroviridis* (pufferfish); Gacu: *Gasterosteus aculeatus* (three spined stickleback); BflA: *Branchiostoma floridae* isoform A (lancelet fish); BflB: *Branchiostoma floridae* isoform B (lancelet fish); Dmel: *Drosophila melanogaster* (fruit fly); Tcas: *Tribolium castaneum* (red flour beetle); Agam: *Anopheles gambiae* (malaria mousquito); Apis: *Acyrthosiphon pisum* (pea aphid); Isca: *Ixodes scapularis* (black-legged tick); Spur: *Strongylocentrotus purpuratus* (purple sea urchin); Skow: *Saccoglossus kowalevskii* (Acorn worm); Nvec: *Nematostella vectensis* (starlet sea anemone); Cele: *Caenorhabditis elegans* (nematode); Cbre: *Caenorhabditis brenneri* (nematode); Bmal: *Brugia malayi* (filarial roundworm); Hmag: *Hydra magnipapillata* (freshwater polyp hydra); Tadh: *Trichoplax adhaerens* (Placozoa).

Studies based on polymannuronan- and chondroitin C5-epimerases indeed indicate structural homology of these epimerases with specific lyases (Pacheco *et al*., 2008). Unfortunately, no known structure in the Protein Data Bank (PDB) has sufficient homology with heparosan D-glucuronyl C5-epimerases to generate a reliable 3D model.

Throughout the sequence there are multiple highly conserved tyrosine and histidine residues, of which most are located in the C-terminal domain. Some of these residues may be functional analogs of the catalytic residues in the polymannuronan- and chondroitin C5-epimerase. For chondroitin D-glucuronyl C5-epimerase 1, evidence has been provided that one tyrosine (Tyr-261) and two histidine residues (His-205 and His-450) are involved in catalysis (Pacheco *et al*., 2008). A model has been proposed in which His-450 is the C5-proton abstractor, Tyr-261 mediates a β-elimination reaction, and His-205 eventually re-protonates the C5 from the opposite site. The proposed catalytic mechanism of alginate C5-epimerase AlgE4 also involves a conserved histidine and tyrosine. The tyrosine (Tyr-149) probably acts as proton abstractor and a histidine (His-154) as proton donor (Rozeboom *et al*., 2008).

MmC5 production in *P. pastoris*

To achieve optimal codon usage we constructed a synthetic gene optimized towards the codon bias of *P. pastoris.* This resulted in a 3.5% decrease of the GC content to 42.4%, which is an ideal level for *P. pastoris*. We could successfully integrate this synthetic murine MmC5 gene in the *P. pastoris* genome by homologous recombination. Successful integration of the gene was confirmed by PCR analysis.

Growth conditions could be optimized for batch-culture growth in baffled Erlenmeyer flasks. It is known that sufficient aeration is crucial for good expression; this could be achieved by heavy shaking and using relative little medium per flask providing ample headspace. We did grow multiple batches of the recombinant *P. pastoris* strain with variations in medium composition and expression time.

Generally, multiple proteins are being secreted by *P. pastoris* to the medium. Protein analysis of the medium by SDS-PAGE confirmed the presence of secreted protein (Fig.2). Levels of expressed protein in the medium fractions were generally low, and were only detectable after silver staining. In comparison with a control strain, protein excretion levels were higher for the epimerase strain. To isolate sufficient protein the medium fraction was

concentrated and washed with buffer, after which fractions were separated by gel filtration. Of the secreted proteins that had the highest expression levels, several candidate protein fractions (numbered 1-5) were isolated from the SDS-PAGE gel.

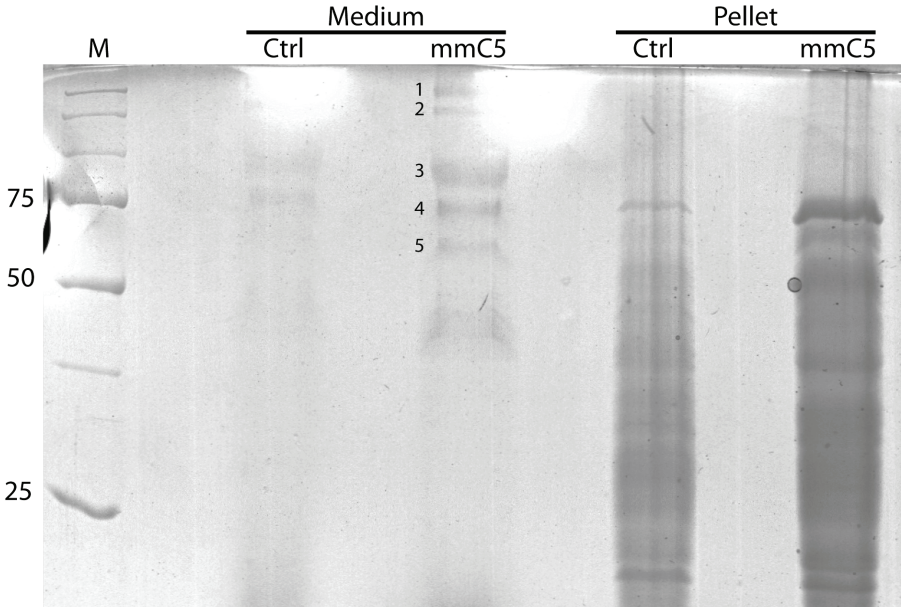


Figure 2 – Protein production of MmC5 in *P. pastoris,* both excreted (Medium) and total cells (Pellet). Indicated protein bands (1-5) have been isolated and analyzed by mass spectrometry. MmC5 is expected to be ≥65kDa. None of the analyzed bands was MmC5. Strain of *P. pastoris* GS115 was used as a background control (Ctrl).

These fractions were Trypsin digested and analyzed by mass spectrometry to reveal the protein identity. No murine C5-epimerase sequence was identified in the analyzed fractions. Hence, we were not able to proof C5-epimerase expression in *P. pastoris*.

We tried variations in medium composition and expression time, however these never resulted in detectable C5-epimerase expression. Possibly, the enzyme is not secreted successfully. For the non-secreted protein fraction we noticed no obvious difference between the control *P. pastoris* and the MmC5 strain and therefore we did not select any non-secreted protein fraction for mass-spectrometry analysis. However, gel analysis cannot exclude low expression levels of non-secreted protein. A solubilizing factor potentially can be a solution to solve these possible *in vivo* solubility problems, for example maltose binding protein (MBP). Recent research showed the potential of MBP as a solubilizing factor in *P. pastoris* based recombinant enzyme production (Dälken *et al*., 2010).

MmC5 production in *E. coli*

In parallel with C5-epimerase expression in *P. pastoris* we tested C5-epimerase production in the bacterium *Escherichia coli*. We cloned the synthetic mouse MmC5 gene successfully in several different bacterial expression vectors, generating several different constructs (Fig.3).

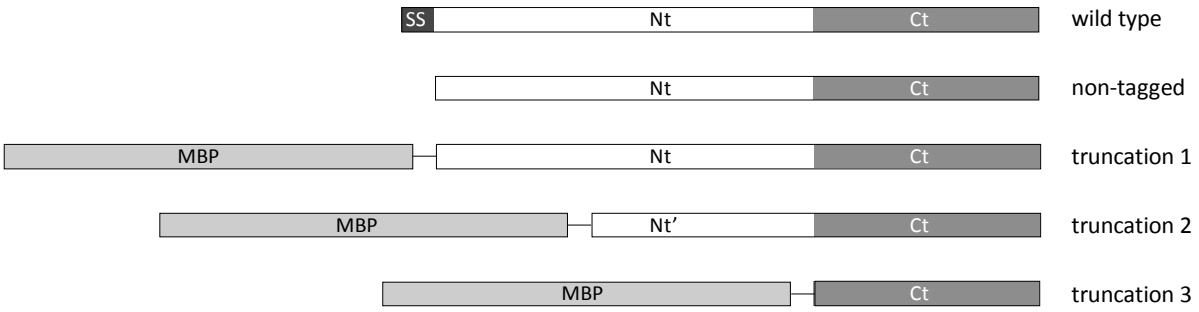


Figure 3 – Schematic representation of constructed C5-epimerases for truncation studies. Used abbreviations: SS - signal sequence, Nt - amino terminus, Ct - carboxyl terminus, MBP - maltose binding protein.

Protein expression levels were examined for each construct. The D-glucuronyl C5-epimerase showed major solubility problems when expressed in *E. coli* (Fig.4A). We could not detect any soluble protein expression, while large quantities of insoluble protein were observed as inclusion bodies. Numerous variations in expression conditions were tried but none had any positive effect. To tackle this solubility problem we created several fusion proteins in which the *E. coli* Maltose Binding Protein (MBP) is fused to the N-terminus of the murine D-glucuronyl C5-epimerase (Fig.3). In total three different fusion constructs were tested in which fragments of the poorly conserved amino terminus of D-glucuronyl C5-epimerase are removed. Positions where truncations start are based on secondary structure predictions (Fig.1), to prevent disruption of secondary structures. The truncation studies showed increased solubility upon removal of the N-terminus of the C5-epimerase (Fig.4B).

Another important parameter for protein solubility is the culture induction time. Maximum protein solubility appears to be reached after 3 hours of protein expression. After 24 hours of expression the amount of soluble protein decreases to levels hardly detectable. The amount of insoluble protein produced after 24 hours is massive. Protein solubility is severely affected in time; long expression time leads to formation of inclusion bodies (not shown).

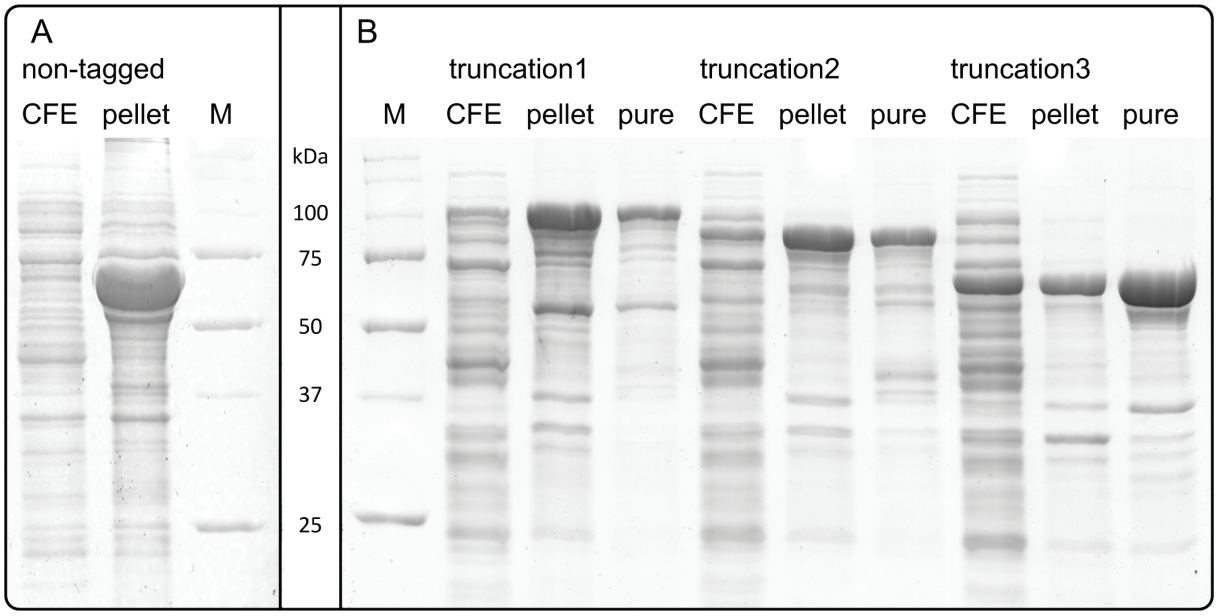


Figure 4 – Murine D-glucuronyl C5-epimerase production in *E. coli*. (A) Non-tagged C5-epimerase is produced massively (size 67kDa). (B) N-terminal fusion with Maltose Binding Protein (MBP; 43 kDa) results in soluble protein in the cell free extract (CFE). The three N-terminal truncation are 110 kDa (truncation1), 95 kDa (truncation2) and 70 kDa (truncation3) in size.

All three truncated fusion proteins do bind to the amylose-resin. The MBP-fusion constructs could be enriched from cell lysate. However the enriched protein was not pure, even after extensive washing. Many *E. coli* proteins apparently associate with the truncated fusion proteins, most likely via non-specific, hydrophobic interactions. Size exclusion chromatography analysis (using a Superdex 200) of the 'purified' soluble MBP-MmC5 fusion protein did not result in one single defined oligomeric state but rather in soluble aggregates of high molecular weight. More examples are known of MBP fusion constructs that result in large soluble aggregates (Zanier *et al*., 2007). The native fold of such a fusion protein is not necessarily affected. However, generally one can expect major problems in crystallization trials (Smyth *et al*., 2003) and therefore we did not proceed with structure elucidation to confirm the predicted alpha/alpha toroid fold.

In contrast to other C5-epimerases, the D-glucuronyl C5-epimerase seems to be Ca^{2+} independent (Valla *et al*., 2001). To confirm the absence of calcium and other divalent ions we performed divalent metal analysis using inductively coupled plasma atomic emission spectroscopy (ICP-AES). No significant amounts of potential cofactors were observed in MBP-MmC5 samples, except for detection of Zn^{2+} in some enzyme samples. Affinity to zinc

has never been reported for the murine D-glucuronyl C5-epimerase. We tested the addition of Zn^{2+} to the medium or reaction buffer but we did not see any effect on protein production levels or activity (see below) when added. Apart from zinc, no other metal was found to be associated with MBP-MmC5.

To screen the generated fusion proteins for D-glucuronyl C5-epimerase activity, we incubated both purified enzyme and cell lysate with K5 capsular polysaccharide. After hydrolysis of the polysaccharide, the samples of monosaccharides were analyzed for their hexuronic acid content using a HPLC-PAD system. Significant amounts of monomeric glucosamine and D-glucuronic acid were observed, indicating proper hydrolysis. However formation of L-iduronic acid for any of the constructs was not too obvious. We could detect some trace levels of L-iduronic acid although the amounts were rather low, and therefore not conclusive.

To confirm this D-glucuronyl C5-epimerase activity we performed an additional screen based on the tritium release method using ^{3}H-labeled polysaccharide substrate (Campbell *et al*., 1983; Kusche *et al*., 1991). We tested MBP-MmC5 activity both as freshly prepared cell lysate and as purified protein. Two substrates were tested, K5 polysaccharide (De-acetylated/N-sulfated) and heparin, and all measurements we done in duplo (Table 1).

Table 1 – Activity measurements using MBP-MmC5 mediated tritium release. The abbreviation N.A. means no significant activity observed.

Substrate	Enzyme	Quantity	^{3}H release (cpm)	
Heparin	MBP-MmC5	50	1487	±199
Heparin	MBP-MmC5	5	634	±206
Heparin	MBP-MmC5 truncation 2	50	N.A.	
Heparin	MBP-MmC5 truncation 3	50	N.A.	
Heparin	Control	2	172	±47
K5 polysaccharide	MBP-MmC5	50	167	±25
K5 polysaccharide	MBP-MmC5	5	114	±19
K5 polysaccharide	MBP-MmC5 (overnight)	5	26	N.D.
K5 polysaccharide	MBP-MmC5 truncation 2	50	N.A.	
K5 polysaccharide	MBP-MmC5 truncation 3	50	N.A.	
K5 polysaccharide	Control	2	13	±7

For both substrates we detected significant C5-epimerase activity using the freshly prepared cell lysate, although a significant reduction of activity was observed upon overnight storage. We were unable to detect activity of the purified protein, most likely due to inactivation upon storage. We also tested both truncated variants of MBP-MmC5 for activity; no activity could be determined for these two truncated C5-epimerases.

In parallel with our studies, similar truncation studies have been performed by Liu and colleagues (University of North Carolina). They claimed functional protein expression of an MBP fusion protein with the human D-glucuronyl C5-epimerase (Muñoz *et al*., 2006). They showed partial epimerization of the D-glucuronic acid residues in N-sulfo-heparosan upon mixing with their fusion protein and an MBP fusion protein of 2-O-sulfotransferase (MBP-2-OST). The addition of MBP-2-OST appears to enhance epimerase activity as shown previously for 2-OST (Pinhal *et al*., 2001). Just recently they confirmed their finding in another report (Li *et al*., 2010). They prepared six N-terminal truncated (human D-glucuronyl C5-epimerase) constructs each as a fusion protein with MBP. Also all their constructs could be expressed. They claim activity for their longer constructs, while C5-epimerase functionality is lost after truncation of the region N92-W203. This leads them to the conclusion that the catalytic site is partly located in this region of the protein.

Discussion

We did not succeed in obtaining C5-epimerase expression in *Pichia pastoris*, as no significant amounts of C5-epimerase were detected. It is unclear whether expression levels are below threshold or completely zero. Possibly there are problems in the α-Factor secretion signal preventing popper C5-epimerase secretion. The addition of an affinity-tag can help in selective protein enrichment and in detecting possible low levels of C5-epimerase. When antibodies are available, western blot could provide a more sensitive, and selective method to prove expression. Alternatively one could check transcription levels of the D-glucuronyl C5-epimerase gene. A highly sensitive activity assay could give ultimate prove of possible low levels of C5-epimerase activity, unfortunately no such assay is available yet.

Apart from the discussion whether or not there is expression, it has become clear that D-glucuronyl C5-epimerase expression in *P. pastoris* does not work in small scale batch

approach. An alternative setup for expression could possibly result in more successful protein expression. Previous findings by others using a fed-batch expression systems did show production of functional C5-epimerase (Weegar *et al*., 2003), when *P. pastoris* is grown in a laboratory scale fermentor. However, this activity was limited to a narrow time span. Their system completely lost C5-epimerase activity in time, possibly due to proteolytic degradation of the epimerase. Oxygen deficiency is one possible reason leading to enhanced proteolysis. As variations in medium pH can have a positive effect in protease deactivation, we tried buffered media of different pH values. However, we never observed a detectable effect on protein expression.

Alternatively, and based on our findings, the loss of epimerase activity could also be a result of lack of protein stabilization. Medium enrichment with a higher amino acid content did not have an effect. Possibly, and in parallel with the observed *E. coli* based expression, D-glucuronyl C5-epimerase needs the stabilizing effect of other protein residues. This supports the hypothesis that this specific enzyme acts in a larger protein complex and needs other members of the heparin biosynthesis pathway for stabilization and activity.

The synthetic gene designed for *P. pastoris* could successfully be expressed in *E. coli*. However, here again the native C5-epimerase showed severe solubility problems. Fusion of the C5-epimerase to a MBP-tag greatly improved the solubility. Also removal of the animo-terminus of the D-glucuronyl C5-epimerase resulted in increased solubility. Larger truncations result in more soluble MBP-MmC5 in the CFE, and less insoluble MBP-MmC5. Protein purification by means of amylase resin results in enrichment for all different MBP-MmC5 truncations, however some *E. coli* proteins stably associate to truncation 1-3. The hypothesis that the animal D-glucuronyl C5-epimerase acts in a larger protein complex would be in agreement with this observation. The enzyme needs the stabilizing effect of maltose binding protein to reach soluble protein expression, allowing the production of large quantities of soluble MBP-MmC5.

We could observe trace levels of L-iduronic acid formation using HPLC-PAD analysis. However, since these levels were low these results alone were not conclusive. To confirm C5-epimerase activity we used the ^{3}H release assay, using tritium labeled substrate. The constructed full length fusion protein not only has an increased solubility but also still is active. The two truncated C5-epimerases were also tested for activity, but turned out to be

completely inactive. It is tempting to speculate that the truncated N-terminus is essential for activity, as both truncated C5-epimerases are completely inactive. However, it cannot be excluded that the MBP possibly blocks activity of the two truncated fusion-proteins.

In parallel with our findings, similar results have been reported for parallel experiments done with the human C5-epimerase upon fusion with MBP. In contrast to our analyses, these studies did include the 2-O sulfotransferase, which is proven to enhance the C5-epimerase activity substantially (Li *et al*., 2010). They used a novel activity assay based on an engineered 2-OST, only able to act on L-iduronic acid. Via a coupled enzyme assay they can indirectly measure C5-epimerase activity. Using this method they could demonstrate activity of the MBP-hsC5-epimerase. They have confirmed their method by using disaccharide RPIP-HPLC analysis. Both their and our findings indicate that soluble and function D-glucuronyl C5-epimerase expression in *E. coli* is feasible. This facilitates alternative production methods of D-glucuronyl C5-epimerases as well it enables the construction of large mutant libraries allowing enzyme optimization studies.

Acknowledgements

The authors are grateful to Emile Wolbert and Jan Springer for their help in the transformation of the *P. pastoris* strain. We also like to thank Bram Snijders for the mass spectrometry analysis and Peter Nobels for his help with the inductively coupled plasma atomic emission spectroscopy (ICP-AES).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. (2004). The Pfam protein families database. *Nucleic Acids Research* **32**, D138-141.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783-795.

Campbell P, Feingold DS, Jensen JW, Malmström A, Rodén L. (1983). New assay for uronosyl 5-epimerases. *Analytical biochemistry* **131**(1), 146-152.

Campbell P, Hannesson HH, Sandbäck D, Rodén L, Lindahl U, Li J. (1994). Biosynthesis of heparin/heparan sulfate. Purification of the D-glucuronyl C-5 epimerase from bovine liver. *The Journal of Biological Chemistry* **269**(43), 26953-26958.

Crawford BE, Jeffrey SK, Esko JD, Pinhal MAS. (2001). Cloning, Golgi Localization, and Enzyme Activity of the Full-length Heparin/Heparan Sulfate-Glucuronic Acid C5-epimerase. *The Journal of Biological Chemistry* **276**, 21538-21543.

Dälken B, Jabulowsky RA, Oberoi P, Benhar I, Wels WS. (2010). Maltose-Binding Protein Enhances Secretion of Recombinant Human Granzyme B Accompanied by In Vivo Processing of a Precursor MBP Fusion Protein. *PLoS One*. **5**(12), e14404.

Gupta R, Jung E, Brunak S. (2004). Prediction of N-glycosylation sites in human proteins (in preparation - http://www.cbs.dtu.dk/services/NetNGlyc).

Jalkanen M, El Darwish K, Lindahl U, Li J. (2008). Glucuronyl C5-epimerase, DNA encoding the same and uses thereof. *United States Patent 7399626*

Kapust RB, Waugh DS. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**(8), 1668–1674.

Kelley LA, Sternberg MJE. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* **4**, 363 – 371.

Kusche M, Hannesson HH, Lindahl U. (1991). Biosynthesis of heparin. Use of *Escherichia coli* K5 capsular polysaccharide as a model substrate in enzymic polymer-modification reactions. *Biochemical Journal* **275**, 151-158.

Li J, Hagner-McWhirter Å, Kjellén L, Palgi J, Jalkanen M, Lindahl U. (1997). Biosynthesis of Heparin/Heparan Sulfate. cDNA cloning and expression of D-glucuronyl C5-epimerase from bovine lung. *The Journal of Biological Chemistry* **272**(44), 28158-28163.

Li J, Gong F, El Darwish K, Jalkanen M, Lindahl U. (2001). Characterization of the D-glucuronyl C5-epimerase involved in the biosynthesis of heparin and heparan sulfate. *The Journal of Biological Chemistry* **276**(23), 20069-20077.

Li J. (2010). Glucuronyl C5-Epimerase: An Enzyme Converting Glucuronic Acid to Iduronic Acid in Heparan Sulfate/Heparin Biosynthesis. *Progress in molecular biology and translational science.* **93**, 59-78.

Li K, Bethea HN, Liu J. (2010). Using engineered 2-O-sulfotransferase to determine the activity of heparan sulfate C5-epimerase and its mutants. *The Journal of Biological Chemistry* **285**, 11106-11113.

McGuffin LJ, Bryson K, Jones DT. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*. **16**, 404-405.

Muñoz E, Xu D, Avci F, Kemp M, Liu J, Linhardt RJ. (2006). Enzymatic synthesis of heparin related polysaccharides on sensor chips: Rapid screening of heparin–protein interactions. *Biochemical and biophysical research communications* **339**, 597-602.

Nicholas KB, Nicholas HB. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. *Distributed by the author*

Notredame C, Higgins DG, Heringa J. (2000). T-Coffee: A novel method for multiple sequence alignments. *Journal of Molecular Biology* **302**, 205-217.

Page, RDM. (1996). Treeview: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.

Pacheco B, Maccarana M, Goodlett DR, Malmström A, Malmström L. (2008). Identification of the active site of DS-epimerase 1 and requirement of N-glycosylation for enzyme function. *The Journal of Biological Chemistry* **284**(3), 1741-1747.

Pinhal MAS, Smith B, Olson S, Aikawa J, Kimata K, Esko JD. (2001). Enzyme interactions in heparan sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **98** (23), 12984-12989.

Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvåg H, Holtan S, Aachmann FL, Valla S Dijkstra BW. (2008) Structural and mutational characterization of the catalytic A-module of the Mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii. The Journal of Biological Chemistry* **283**, 23819-23828.

Sasisekharan R, Venkataraman, G. (2000). Heparin and heparan sulfate: biosynthesis, structure and function. *Current Opinion in Chemical Biology* **4**, 626-631.

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Science* **12**(7), 1313-1322.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876-4882.

Valla S, Li J, Ertesvåg H, Barbeyron T, Lindahl U. (2001). Hexuronyl C5-epimerases in alginate and glycosaminoglycan biosynthesis. *Biochimie* **83**(8), 819-830.

Weegar J, Dahlbacka J, Åkerlund B, Ollikka P, Fagervik K. Production of C-5 epimerase in a laboratory scale bio-reactor with *Pichia pastoris*. Poster presented as part of the 11th European Congress on Biotechnology, Basel, Switzerland, 24-29 August 2003.

Zanier K, Nominé Y, Charbonnier S, Ruhlmann C, Schultz P, Schweizer J, Travé G. Formation of well-defined soluble aggregates upon fusion to MBP is a generic property of E6 proteins from various human papillomavirus species. *Protein expression and purification* **51**(1), 59-70.

Zdobnov EM, Apweiler R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**(9), 847-848.

Chapter 5

Improved production of *Pasteurella multocida* hyaluronan synthase in *Escherichia coli*

Manuscript in preparation:

Raedts J, Sloothaak J, Nag A, Kengen SWM, van der Oost J.

Abstract

Maltose Binding Protein (MBP) has been developed as a general tool to enhance the production of heterologous proteins as MBP fusion in *Escherichia coli*. The target protein of the present study is the hyaluronan synthase from *Pasteurella multocida* (PmHAS). We created a MBP-PmHAS fusion protein, and compared expression levels in *E. coli* with a control PmHAS (PmHAS-His). The fusion with MBP resulted in a three-fold increase in PmHAS protein solubility. Although MBP is a rather large protein-tag, it does not block HAS activity of the MBP-PmHAS fusion protein. Compared with the control protein there is no difference in specific activity of the MBP-PmHAS fusion protein.

Introduction

Glycosaminoglycans are biological active polysaccharides widely spread among the animal kingdom. Hyaluronan, also referred to as hyaluronic acid, is a widely used and well-studied glycosaminoglycan. It is a linear, negatively charged polysaccharide that, compared to other glycosaminoglycans, is rather large with as much as 2,000-25,000 disaccharides of D-glucuronic acid (GlcA) and N-acetyl-D-glucosamine (GlcNAc) (Kogan *et al*., 2007). These disaccharides are joined alternately by β(1-3) and β(1-4) glycosidic bonds, forming a uniformly repetitive structure (Fig.1). Unlike other glycosaminoglycans, hyaluronan biosynthesis does not require post-polymerization modifications like epimerization or sulfation.

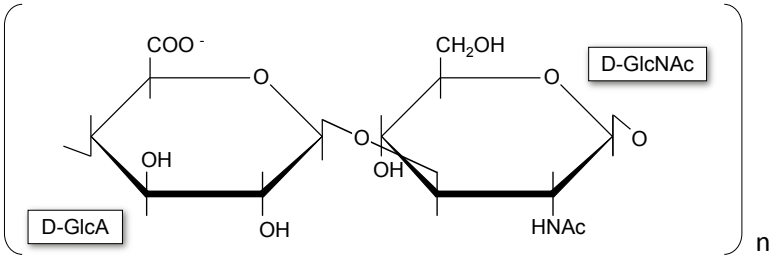


Figure 1 – Repetitive disaccharide of hyaluronan comprising D-glucuronic acid and N-acetylglucosamine

Hyaluronan is involved in numerous biological processes, one of them reflected by its name. Hyaluronan is derived from 'hyaloid', Greek for the vitreous of the eye. Moreover, hyaluronan is also present in the synovial fluid of articular joints and the intercellular space of the epidermis. This wide variety of biological functions results in a large number of applications that exist for hyaluronan in the field of medicine and cosmetics, with a worldwide market for hyaluronan estimated close to $1 billion (Chong *et al*., 2005). Hyaluronic acid is isolated from animal tissue, a good source is rooster comb that has hyaluronan concentrations up to 0.75% v/v (Laurent *et al*., 2002). Another frequently used source of hyaluronan are the capsules of certain (pathogenic) bacteria where they play a role as a virulence factor (Wessels *et al*., 1991). Although both are good sources of hyaluronan, major drawbacks exist in (1) possible co-purification of immunogenic contaminants and (2) the undefined molecular sizes of the isolated hyaluronan fractions. Both these features are highly undesirable when aiming at pharmaceutical and cosmetic applications. A way to prevent either problem would be the controlled enzymatic synthesis of size-defined hyaluronan.

The biosynthesis of hyaluronan involves specific glycosyltransferases, pertinently called hyaluronan synthases (HASs; EC 2.4.1.212). Unlike other glycosyltransferases these HASs are able to transfer two different UDP-sugars, UDP-GlcA and UDP-GlcNAc, to the growing polymer. Multiple HASs have been identified the past two decades (DeAngelis, 1999; Weigel and DeAngelis, 2007). Apart from vertebrates HASs, also a viral and multiple bacterial HASs have been identified (Weigel, 2002).

Comparison of all these different HASs has led to the division of these enzymes into two different classes (DeAngelis, 1999). Almost all HASs are so-called Class I HASs. These are integral membrane proteins containing multiple membrane domains present throughout the protein sequence. One glycosyltransferase domain is responsible for the transfer of both UDP-sugars. The hyaluronan synthase from the gram negative bacterium *Pasteurella multocida* (PmHAS) is the only example of a Class II HAS (DeAngelis, 1996). PmHAS is a peripheral membrane protein that is membrane associated via its carboxyl-terminus. Differently than Class I HASs, PmHAS has two glycosyltransferase domains, each responsible of the transfer of one of the UDP-sugars (DeAngelis and Padgett-McCue, 2000; Jing and DeAngelis, 2003; Kooy, 2010). These UDP-sugars are added at the non-reducing end of the growing hyaluronic acid chain (Fig.2).

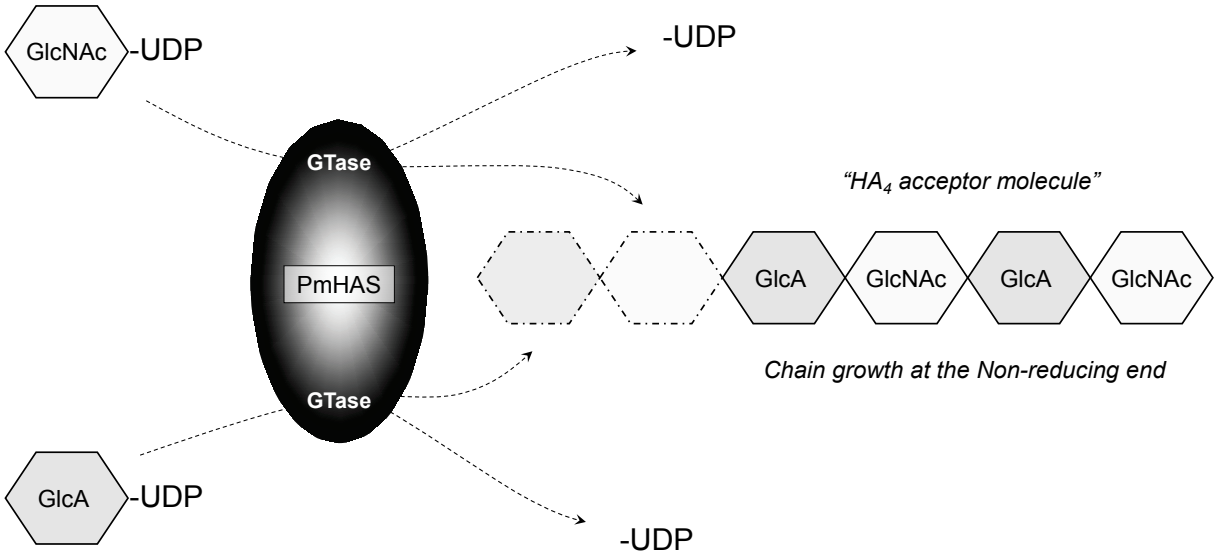


Figure 2 – Schematic representation PmHAS mediated hyaluronan biosynthesis. Two active sites (GTase) are present each responsible for the transfer of one UDP-sugar to the non-reducing end of an acceptor molecule. UDP is released upon chain elongation.

The initiation step is very slow, addition of an acceptor molecule (typically a tetramer; HA_4) increases chain elongation speed. Addition of HA_4 also reduces the polydispersity, the resulting polymers vary less in size (Jing and DeAngelis, 2004).

The *PmHAS* gene has been cloned and heterologically expressed in *Eschericia coli* previously (DeAngelis, 1998). The C-terminal membrane association domain could be removed, resulting in a cytoplasmatic protein, without losing HAS activity (Jing and DeAngelis, 2003). Although the current levels of PmHAS production may be is sufficient for biotechnological purposes, increased production levels as well as improved protein solubility are desirable. We aim for this by fusing PmHAS to maltose binding protein (MBP), resulting in a MBP-PmHAS fusion protein (Fig.3). In this study we describe the stabilizing effect of MBP on PmHAS, in comparison with a PmHAS lacking such stabilizing tag.

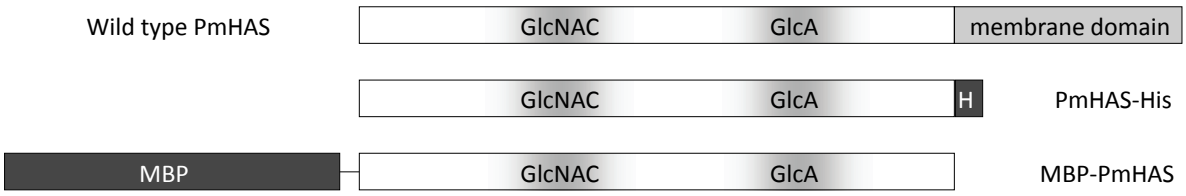


Figure 3 – Schematic representation of wild type PmHAS (top), and constructed fusion proteins. Abbreviations used are GlcNAC / GlcA = glycosyltransferase domains, MBP = maltose binding protein, H = polyhistide-tag.

Materials and methods

Chemicals, UDP-sugars and the pyruvate kinase and lactate dehydrogenase were purchased from Sigma-Aldrich. HA_4 was a kind gift of Floor Kooy (A&F Wageningen UR). Used restriction enzymes, Pfu DNA polymerase and T4 DNA ligase were from Invitrogen Amylose Resin was purchased from New England Biolabs, the Talon metal Affinity Resin was purchased from BD Biosciences.

The used bacterial expression vectors were pET101/D-TOPO (Invitrogen) and pMAL-C2 (New England Biolabs). The His-tagged construct (pET101) was a kind gift from Jan Springer (A&F Wageningen UR), the MBP-tagged construct (pMAL-C2) was constructed in *E. coli* DH5α. The used bacterial expression host was *E. coli* BL21 (DE3) (Novagen), containing the tRNA helper plasmid pSJS1244.

Gene cloning

The MBP-tagged fusion protein was created by PCR using the His-tagged construct as template. The sense primer used was BG2650 (5'-CCCGAATTCAATACATTATCACAAGCAATAAAAGCATATAACAGC-3'), the antisense primer was BG2651 (5'-CCCCTGCAGTTAAATATCTTTTAAGATATCAATCTCTTCTTGATATTCAGC-3'). The introduced flanking restriction sites EcoRI / PstI (underlined) were used to clone the gene in the bacterial expression vector pMal-C2. *E. coli* DH5α was transformed with the ligation mix to amplify the resulting plasmid named pWUR539.

Protein expression

Expression vector pWUR539 (MBP-tag) was used to transform *E. coli* BL21(DE3) harboring pSJS1244. For each construct one single colony was used to prepare a starter culture by inoculating Luria-Bertani medium (LB) containing 100 µl/ml ampicillin (and 50 µl/ml spectinomycin for pSJS1244). The starter culture was grown overnight at 37°C while shaking. A larger culture was started by inoculation of 100 ml LB (similar antibiotic content) with 1% starter culture. The culture was grown at 37°C, while shaking, until OD_{600} = 0.6 was reached. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM, and growth was continued for another 3 hours while shaking at a temperature of 37°C. Cells were harvested by centrifugation, and resuspended in 5 ml of 20 mM TRIS-HCl buffer (pH7.5) with 50 mM KCl. The cells were disrupted by sonication with intermitted cooling steps on ice. The crude cell lysate was centrifuged (16,000 x g) to remove cell debris. Protein expression levels were checked by SDS-PAGE analysis, quantities were measured after staining with Coomassie Brilliant Blue (Bradford analysis and analysis with Quantity ONE® Bio-Rad). For protein purification the supernatant was applied on a gravity flow column (Bio-Rad) of amylose resin (MBP-PmHAS) or Talon resin (PmHAS-His), and purified according the manufacturers' protocol.

PmHAS activity measurements

Protein samples always were freshly prepared and kept on ice until used. Hyaluronan production could be confirmed by gel electrophoresis (Ikegami-Kawai and Takahashi, 2002). The reaction mixture contained 10 µg PmHAS, 10 mM of each UDP-sugar and 0.1 mM HA_4 in a TRIS-HCl buffer (50 mM pH 7.1, 5 mM $MnCl_2$, 1.1 M ethylene glycol). The reaction was

performed overnight at 30°C analyzed on a 6% TBE polyacrylamide gel (invitrogen) and stained with Stains-All as described previously (Kooy et al., 2009).

The glycosyltransferase activity of PmHAS was measured using a coupled enzyme assay (Krupa *et al*., 2007). This assay is based on the UDP release upon hyaluronan elongation reaction, which is directly linked to decrease in NADH, which could be spectrophotometrically measured at 340 nm (Equation 1-3).

(1) *n* UDP-GlcA + *n* UDP-GlcNAc + HA_4 –[*PmHAS*]→ $(GlcA\text{-}GlcNAc)_n\text{-}HA_4$ + 2*n* UDP

(2) UDP + phosphoenolpyruvate –[*pyruvate kinase*]→ UTP + pyruvate

(3) pyruvate + NADH –[*L-lactate dehydrogenase*]→ L-lactate + NAD^+

The standard reaction condition buffer was composed of 20 mM TRIS-HCl (pH 7.5), 1 M of ethylene glycol, 50 mM KCl, 13 mM $MnCl_2$, 1.4 mM phosphoenolpyruvate (PEP), 28 µM UDP-GlcA, 1.32 mM UDP-GlcNAc and 48 µM HA_4. The two coupled enzymes were present in excess, knowing 30 U/ml pyruvate kinase and 37 U/ml lactate dehydrogenase. In total 0.15 mM NADH was used to have the measurements in the linear range (ε_{340} = 6.3 $mM^{-1}cm^{-1}$). The typical volume of added PmHAS was 40 µl. All assays were performed using a Hitachi U2001 spectrophotometer with a temperature controlled cuvette holder set at 30°C.

Results and Discussion

Two tagged PmHASs were tested for their expression levels and protein solubility. Aiming for a better soluble PmHAS, we created a MBP-PmHAS fusion protein (Fig.3). Maltose binding protein (MBP) is well known for its stabilizing effect in *Escherichia coli* (Kapust and Waugh, 1999). As a reference we used a previously constructed His-tagged PmHAS, as a polyhistidine tag does facilitate easy protein purification but it generally does not result in any stabilizing effect. Both PmHAS fusion proteins are truncated (lacking C-terminal docking sequence) in order to gain maximal solubility.

Both PmHAS fusion proteins could be successfully expressed in *E. coli* BL21(DE3). We noticed that protein solubility is severely affected by elongated expression time, resulting in large quantities of insoluble protein. This could be partially prevented by having an expression time of 3 hours, although this did not completely prevent PmHAS ending up in

inclusion bodies. Comparison of protein expression of both fusion proteins clearly shows differences in protein solubility levels (Fig.4).

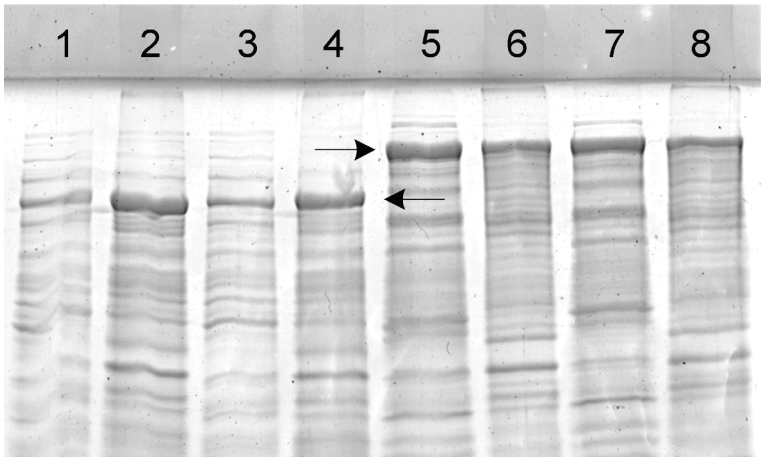


Figure 4
Comparison recombinant PmHAS fusion protein expression levels in *E. coli* BL21(DE3). The arrows indicate the size of PmHAS-His (lane 1-4; 81 kDa) and MBP-PmHAS (lane 5-8; 126kDa). Odd numbers represent soluble protein, even numbers insoluble protein. Marker (M) sizes are indicated on the left in KDa.

Using identical expression conditions, obvious differences in protein solubility levels can be observed. The PmHAS-His and MBP-PmHAS cultures were grown in parallel experiments to ensure identical expression conditions. Multiple independent expression experiments show an increase of solubility levels from 20% for PmHAS-His to almost 60% for MBP-PmHAS (Fig.5). Also the total amount of MBP-PmHAS produced generally is larger than PmHAS-His, although absolute quantities vary too much per batch to be conclusive.

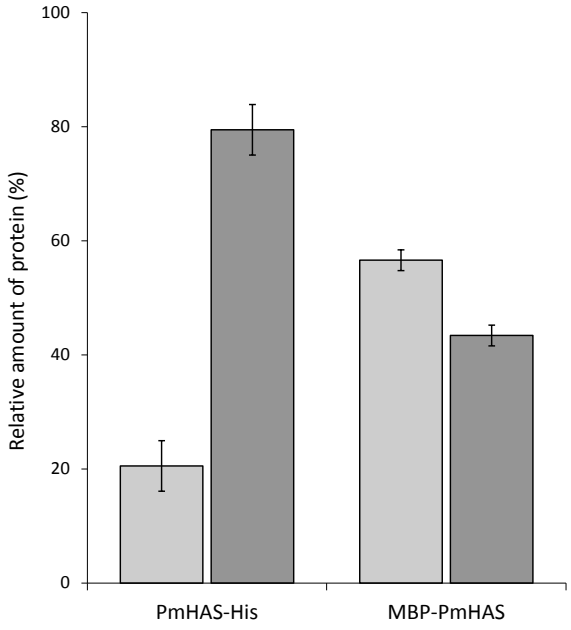


Figure 5
Comparison recombinant PmHAS fusion protein expression levels in *E. coli* BL21(DE3). Indicated is the relative amount of soluble enzyme (light grey bars) compared to insoluble enzyme (dark grey bars). Total amount of each PmHAS fusion product is set at 100%.

Obviously PmHAS solubility is greatly improved by fusion to MBP. However, since maltose binding protein is a rather large protein, it might be well possible that such large body blocks the active site. Hence MBP-PmHAS was tested for its hyaluronan synthesizing capacity, using both MBP-PmHAS in cell free extract and purified enzyme. One batch freshly prepared and one batch of three weeks old MBP-PmHAS were analyzed to test protein stability. The reactions were incubated overnight to ensure complete conversion, ethylene glycol was added to stabilize the enzyme. The produced hyaluronan was checked on a TBE gel (Fig.6).

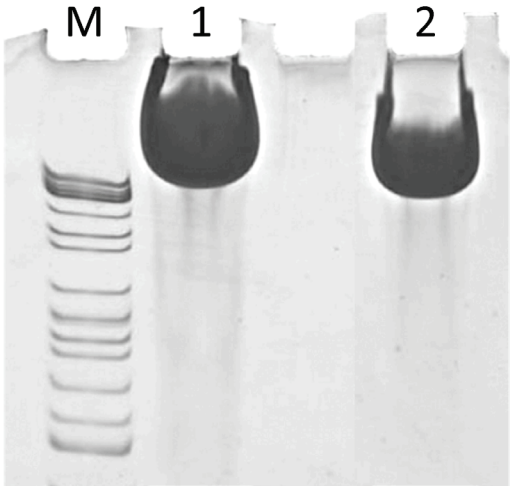


Figure 6 – TBE-gel based analysis of hyaluronan synthesis by using MBP-PmHAS. Lane M give an indication of the molecular mass. Lane 1 represent hyaluronan production using cell lysate containing MBP-PmHAS, lane 2 represent hyaluronan production by using purified MBP-PmHAS..

Overnight incubations of MBP-PmHAS with both-UDP sugars and HA_4, resulted in the production of high molecular weight hyaluronan for both MBP-PmHAS in cell lysate and for purified enzyme. The old batch of MBP-PmHAS lost activity. This corresponds with untagged PmHAS, which has similar stability problems when stored at 4°C (unpublished data).

No inhibitory effect was observed of the MBP molecule on HAS activity. However, to exclude any effect of the MBP-tag on the activity, we did some more in-depth comparison between MBP-PmHAS and PmHAS-His. Therefore we compared the specific activity of both fusion proteins. We measured multiple batches of freshly prepared and purified MBP-PmHAS and PmHAS-His using a coupled enzyme assay (equation 1-3) on a temperature controlled spectrophotometer. Since MBP is a rather large tag, the molecular mass of MBP-PmHAS increases significantly. Therefore, to correct for the change in protein-size we not only expressed activity in per mg (Fig.7A), but also in the activity per mole (Fig.7B).

We could not observe an obvious difference in specific activity of both fusion proteins after corrected for the increased mass of MBP-PmHAS. The measured activities are

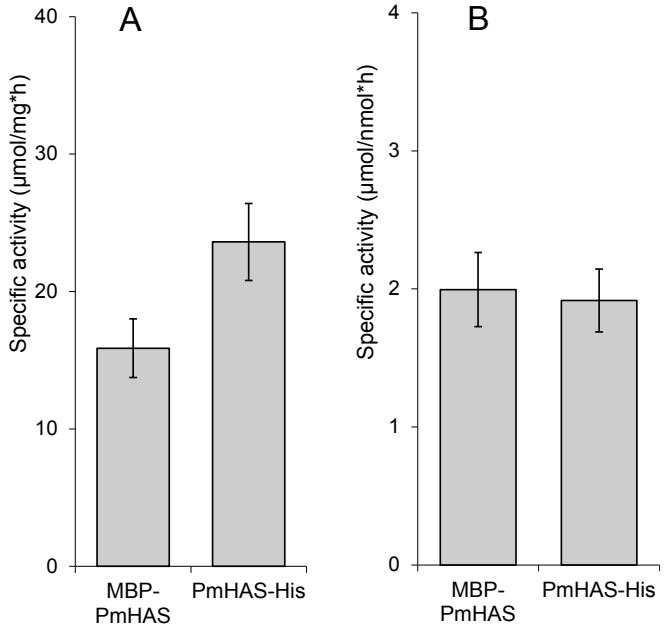


Figure 7 – (A) Comparison of the specific activity of both MBP-PmHAS and PmHAS-His expressed as µmol per mg per hour. (B) Specific activity expressed in µmol per nmol per hour.

in good agreement with values found in literature (Kooy, 2010). Although protein solubility is greatly improved, a better protein stability is still desirable. We noticed loss of activity over time for both fusion proteins. Stability of MBP-PmHAS seems increased compared to PmHAS-His, but more detailed research is needed to be conclusive.

Conclusions

The constructed fusion protein potentially is a great tool for enzymatic hyaluronan synthesis, as large quantities of protein easily can be produced. Even though MBP is a fairly large domain, no negative effect is observed on HAS activity. This suggests that PmHAS can be N-terminally coupled to a carrier material without losing activity. Such a carrier coupled PmHAS could greatly improve controlled hyaluronan synthesis. PmHAS mutants that have lost affinity of one the UDP-sugars have been created by others (Jing and DeAngelis, 2003). Step wise addition and subsequent easy removal of such a coupled PmHAS mutant from the reaction mixture potentially is a tool to synthesize hyaluronan oligosaccharides of a controlled chain length, greatly reducing hyaluronan polydispersity.

Acknowledgements

We would like to thank Gerrit Eggink and Jan Springer for providing the PmHAS gene, and are grateful to Emile Wolbert and Floor Kooy for technical assistance and useful discussions.

References

Chong BF, Blank LM, Mclaughlin R, Nielsen LK. (2005) Microbial hyaluronic acid production. *Applied microbiology and biotechnology* **66**, 341-351.

DeAngelis PL. (1996). Enzymological characterization of the *Pasteurella multocida* hyaluronic acid synthase. *Biochemistry* **35**(30), 9768-9771.

DeAngelis PL. (1999). Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cellular and molecular life sciences : CMLS.* **56**(7-8), 670-682.

DeAngelis PL, Jing W, Drake RR, Achyuthan AM. (1998). Identification and molecular cloning of a unique hyaluronan synthase from *Pasteurella multocida*. *The Journal of biological chemistry* **273**(14), 8454-8458.

Ikegami-Kawai M, Takahashi T. (2002). Microanalysis of hyaluronan oligosaccharides by polyacrylamide gel electrophoresis and its application to assay of hyaluronidase activity. *Analytical biochemistry* **311**(2), 157-165.

Jing W, DeAngelis PL. (2000). Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. *Glycobiology* **10**(9), 883-889.

Jing W, DeAngelis PL. (2003). Analysis of the two active sites of the hyaluronan synthase and the chondroitin synthase of *Pasteurella multocida*. *Glycobiology* **13**(10), 661-671.

Jing W, DeAngelis PL. (2004). Synchronized chemoenzymatic synthesis of monodisperse hyaluronan polymers. *Journal of biological chemistry* **279**(40), 42345-42349.

Kapust RB, Waugh DS. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**(8), 1668–1674.

Kogan G, Soltés L, Stern R, Gemeiner P. (2007). Hyaluronic acid: a natural biopolymer with a broad range of biomedical and industrial applications. *Biotechnology letters.* **29**(1), 17-25.

Kooy FK, Ma M, Beeftink HH, Eggink G, Tramper J, Boeriu CG. (2009). Quantification and characterization of enzymatically produced hyaluronan with fluorophore-assisted carbohydrate electrophoresis. *Analytical biochemistry* **384**(2), 329-336.

Kooy FK. (2010). Enzymatic production of hyaluronan oligo- and polysaccharides. *Ph.D. Thesis Wageningen University*, Wageningen, the Netherlands. ISBN: 978-90-8585648-1.

Krupa JC, Shaya D, Chi L, Linhardt RJ, Cygler M, Withers SG, Mort JS. (2007). Quantitative continuous assay for hyaluronan synthase. *Analytical biochemistry* **361**(2), 218-225.

Laurent TC, Fraser JR. (2002). Hyaluronan. *The FASEB journal* **6**(7), 2397-2404.

Weigel PH. (2002). Functional characteristics and catalytic mechanisms of the bacterial hyaluronan synthases. *IUBMB life* **13**(10), 661-671.

Weigel PH, DeAngelis PL. (2007). Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *The Journal of biological chemistry* **282**(51), 36777-36781.

Wessels MR, Moses AE, Goldberg JB, DiCesare TJ. (1991). Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. *Proceedings of the National Academy of Sciences of the United States of America* **88**(19), 8317-8321.

Chapter 6

Correlated mutation analysis as a tool for smart library design to improve *Pyrococcus furiosus* phosphoglucose isomerase activity

Manuscript in preparation:

Raedts J, Feras Almourfi, Hendriks S, Joosten HJ, Schaap P, Kengen SWM, Baker P, van der Oost J.

Abstract

The use of enzymes as catalysts in industrial processes is steadily increasing. Enzymes are, however, optimized to function optimally in their natural environment, which often differs from the conditions used in industrial settings. To enable a more rational approach in engineering such enzymes, various bioinformatics tools are being developed for 'smart library design'. Here we used multiple sequence alignments and a correlated mutation analysis tool (CMA) to identify targets for mutagenesis. CMA analysis of the RmlC-like cupin super-family revealed a set of correlated amino acids. Using the phosphoglucose isomerase from *Pyrococcus furiosus* (PfPGI) as model enzyme, we subsequently varied correlated residues Pro27 and Tyr28 by saturation mutagenesis. Although this amino acid pair is located in a conserved loop distant from the active site, their predicted relevance could be confirmed by activity measurements of the generated PfPGI mutants. The generated library showed a positive relationship between the prevalence of a certain amino acid pair in the superfamily and the activity of the corresponding PfPGI mutant. We hypothesize that the observed differences in PfPGI activity are caused by subtle structural changes in metal coordination. Obtained crystal structures of four selected PfPGI variants are currently under investigation.

Introduction

Enzymes have been recognized and implemented as useful catalysts for numerous industrial purposes. An enormous variety of enzymes can be found in nature, providing a rich source of potential biocatalysts. However, in evolution these enzymes are optimized to function optimally in *in vivo* environments, which may differ from *in vitro* industrial conditions. Therefore there is often a need for further protein optimization for applicability in industrial settings. Generally this is achieved by the generation of large libraries of variants, from which a few mutants with improved features are selected. As the screening of large libraries typically is costly and time inefficient, reductions in library-size by "smart library design" would be an appreciated step forward (Chica *et al*., 2005; Lutz, 2010).

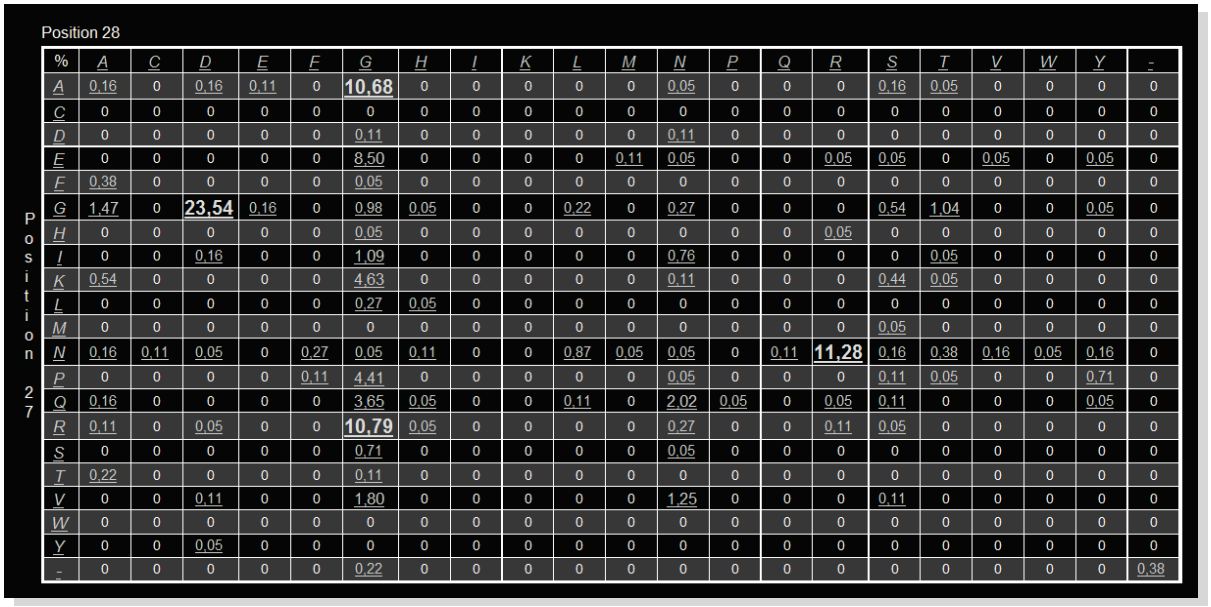
Selection of specific amino acid residues for mutagenesis, requires identification of key residues. Such amino acids either can be identified through experimental analyses, or via functional predictions using bioinformatics. Comulator, an extension of the 3DM software suite, is a recently developed bioinformatics tool that uses a correlated mutation analyses (CMA) algorithm to identify coevolved residues in (large) multiple sequence alignments alignments (MSAs) (Kuipers *et al*., 2009; Jochens and Bornscheuer, 2010).

Subject of study is the RmlC-like cupin super-family; a large group of structurally related proteins present in all three domains of life (Dunwell *et al*., 2000) which includes isomerases, dioxygenases, oxidoreductases and storage proteins. Although members of this family cover a broad spectrum of different functions, all members of the family share the same β-barrel structure composed of β-strands. The name cupin reflects this structure and has been derived from the latin word "cupa" (small barrel). Much structural data has been collected for different members of the cupin superfamily (Dunwell *et al*., 2004).

One of the best characterized members of the cupin superfamily is the cupin type phosphoglucose isomerase of the hyperthermophilic archaeon *Pyrococcus furiosus* (PfPGI; E.C.5.3.1.9). PfPGI is a glycolytic enzyme, that catalyses the reversible isomerization of glucose-6-phosphate to fructose-6-phosphate (Verhees *et al*., 2001; Hansen *et al*., 2001). Multiple crystal structures of this homodimer (monomeric subunit is 21.5 kDa) have been solved, and ample insight in the catalytic activity of the enzyme has been gained (Berrisford *et al*., 2003; Berrisford *et al*., 2004; Berrisford *et al*., 2006). The enzyme can be well expressed in *E. coli*, is very stable *in vitro* and activity can be monitored easily. All this makes

PfPGI an ideal candidate for our studies.

In this study we want to obtain a proof of principle that correlated mutation analyses can help in identification of structural / functional important sites, and can be used as a tool to identify potential "hot spots" for mutagenesis. We generated a library of substitutions at such a hot spot, by varying two correlated amino acids identified by CMA (Fig.1). The predicted relevance of the correlated residues could be validated by activity measurements. By use of smart library design a small set of promising candidates might be equally effective to improve enzyme activity, as a complete (random) library is.



Position 28

Position 27	%	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	-
	A	0,16	0	0,16	0,11	0	**10,68**	0	0	0	0	0	0,05	0	0	0	0,16	0,05	0	0	0	0
	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D	0	0	0	0	0	0,11	0	0	0	0	0	0,11	0	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	8,50	0	0	0	0	0,11	0,05	0	0	0,05	0,05	0	0,05	0	0,05	0
	F	0,38	0	0	0	0	0,05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1,47	0	**23,54**	0,16	0	0,98	0,05	0	0	0,22	0	0,27	0	0	0	0,54	1,04	0	0	0,05	0
	H	0	0	0	0	0	0,05	0	0	0	0	0	0	0	0	0,05	0	0	0	0	0	0
	I	0	0	0,16	0	0	1,09	0	0	0	0	0	0,76	0	0	0	0	0,05	0	0	0	0
	K	0,54	0	0	0	0	4,63	0	0	0	0	0	0,11	0	0	0	0,44	0,05	0	0	0	0
	L	0	0	0	0	0	0,27	0,05	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,05	0	0	0	0	0
	N	0,16	0,11	0,05	0	0,27	0,05	0,11	0	0	0,87	0,05	0,05	0	0,11	**11,28**	0,16	0,38	0,16	0,05	0,16	0
	P	0	0	0	0	0,11	4,41	0	0	0	0	0	0,05	0	0	0	0,11	0,05	0	0	0,71	0
	Q	0,16	0	0	0	0	3,65	0,05	0	0	0,11	0	2,02	0,05	0	0,05	0,11	0	0	0	0,05	0
	R	0,11	0	0,05	0	0	**10,79**	0,05	0	0	0	0	0,27	0	0	0,11	0,05	0	0	0	0	0
	S	0	0	0	0	0	0,71	0	0	0	0	0	0,05	0	0	0	0	0	0	0	0	0
	T	0,22	0	0	0	0	0,11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	V	0	0	0,11	0	0	1,80	0	0	0	0	0	1,25	0	0	0	0,11	0	0	0	0	0
	W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Y	0	0	0,05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	-	0	0	0	0	0	0,22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,38

Figure 1 – Residue pair frequency table of amino acid couple 27 and 28. For wild type PfPGI these residues are Pro132 and Tyr133. Occurrences are relative to the number of unique sequences in the super family alignment.

Experimental procedures

Yeast glucose-6-phosphate dehydrogenase was purchased from MP biomedicals. Chemicals were purchased from Sigma-Aldrich and Roche. The PfPGI mutant library was created by BaseClear (The Netherlands), the genes were cloned in expression vector pET24d (Novagen).

PfPGI mutant library

The cloning of the gene *pgiA* has been described previously (Verhees *et al.*, 2001). A site saturation library was designed and created based on CMA using the Comulator software. The constructed library consisted of *pgiA* variants that had alterations in two strongly

correlated amino acids; proline 132 and tyrosine 133. The corresponding numbering in the 3DM alignment was Pro27 and Tyr28 (Fig.1). The created *pgiA* mutants were cloned in expression vector pET24d and used to transform *Escherichia coli*.

PfPGI expression and purification

Starter cultures of the *pgiA* mutants were inoculated from a glycerol stock and grown overnight in Luria Bertani medium supplemented with 50 µg/ml kanamycin (LB/Km) in a 37°C shaker. The overnight culture was used to inoculate (0.2% v/v) sterile glass tubes of 10 milliliter LB/Km medium. When the optical density of the culture reached A_{600} = 0.5, gene expression was induced by addition of 0.1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Growth was continued overnight at 37°C, after which the cells were harvested by centrifugation (4,600 x g for 15 minutes). Pelleted *E. coli* cells were resuspended in 20 mM Tris-HCl buffer (pH 8.0) and disrupted by sonication. DNAse was added to degrade the DNA in the cell lysate to reduce viscosity. Cell debris was removed by centrifugation (16,000 x g for 15 minutes). *E. coli* proteins were denatured by heating the cell free extract at 70°C for 30 minutes, and removed by centrifugation (16,000 x g for 15 minutes). The result was a heat treated cell free extract containing mainly PfPGI. Its purity was checked by SDS-PAGE. Protein concentrations were determined by Coomassie Brilliant Blue G250 (Bradford, 1976), with bovine serum albumin as reference and analysis on SDS-PAGE (Quantity One®, Bio-Rad).

PfPGI was purified to homogeneity using FPLC as described before (Kuipers *et al*., 2009). Heat treated cell free extract was diluted to lower the salt concentration, filtered through a 0.45 µm filter and loaded on a Q-sepharose fast flow column (Amersham Pharmacia Biotech). The column was equilibrated with 20 mM Tris-HCl (pH 8.0). PGI activity eluted at 180 mM of NaCl during a linear gradient of 0 to 1 M NaCl. The fraction with the highest activity was loaded on a pre-equilibrated Superdex 200 GL column, and eluted in 20 mM Tris-HCl (pH 7.0) containing 100 mM NaCl, protein concentrations and purity were determined. The purified enzyme fraction was used for activity assays. Metal analysis was performed on samples of the same fraction by using ICP-AES.

PfPGI activity assay

Divalent metals was stripped from purified PfPGI using 50 mM ETDA incubated at 50°C for 20

min just prior to the activity measurement. PfPGI activities were determined by measuring NADPH formation in a coupled enzyme assay with yeast glucose-6-phosphate dehydrogenase. This enzyme was present in excess to ensure that the detection of NADPH absorbance at 340 nm (ε = 6.3 $mM^{-1}cm^{-1}$) corresponded to PfPGI activity. The assay mixture contained 0.5 mM NADP, 5 mM D-fructose-6-phosphate and 0.35 units of D-glucose-6-phosphate dehydrogenase, all in 20 mM Tris-HCl buffer (pH 7.0). All assays were performed using a Hitachi U2001 spectrophotometer with a temperature controlled cuvette holder set at 50°C. The optimal activity was measured after titration with $MnCl_2$, while an excess of this salt resulted in enzyme inhibition.

Results and Discussion

By using the Comulator CMA algorithm, as described previously (Kuipers *et al*., 2009), predictions were made to generate a PfPGI library. We used a refined multiple sequence alignment of the cupin super-family, containing a total of 1824 sequences. The amino acids with the highest pair-wise correlated mutation score, were those at residue position 27 and 28 (3DM-numbers) (Fig.1). This amino acid pair is located in a structurally conserved surface loop (Kuipers *et al*., 2009) that can be found in most members of the cupin super-family including PfPGI (Fig.2).

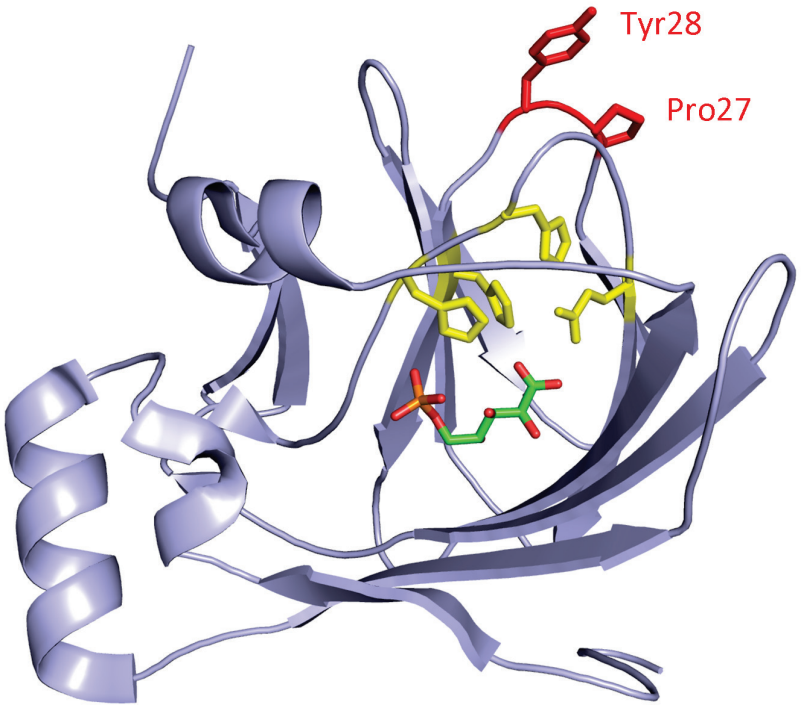


Figure 2 – Cartoon representation of the PfPGI 3D-structure (PDB number: 1x82). The correlated amino acid pair P27 / Y28 is indicated in red. In yellow those residues that are involved in metal ion binding. Inhibitor 5-phospho-D-arabinonate (5PA) is shown as a stick model.

Previous experiments have shown that a PfPGI double mutant, exhibited elevated PGI activity-levels while both single mutants were less active than wild type PfPGI (Kuipers *et al*., 2009). This result is not obvious since this surface loop is not in close proximity to the catalytic residues.

To examine the effect on the PGI activity of the correlated residues Pro27 and Tyr28 in more detail, we selected fifteen mutants (out of the 400 possible) that correspond to amino acid pairs that are either highly abundant or (almost) absent within the refined cupin superfamily alignment (Fig. 1 and Table 1).

Table 1 – Selected amino acid pairs, their natural occurrence and metal ion dependency according literature.

Amino acid pair	Occurrence in alignment (%)	Molecular function involved proteins	Metal ion bound
G27 / D28	23.54	oxidoreductase	Fe
N27 / R28	11.28	dioxygenase, oxidoreductase	Fe
R27 / G28	10.79	dioxygenase, oxidoreductase, isomerase	Fe, Mn
A27 / G28	10.68	dioxygenase, storage protein, isomerase	Fe, Ni, Mg, Mn
E27 / G28	8.50	isomerase, storage protein	Mn
K27 / G28	4.63	isomerase	Mn
P27 / G28	4.41	storage protein, isomerase, dioxygenase	Mn, Ni
V27 / G28	1.80	isomerase	Mn
G27 /G28	0.98	dioxygenase, isomerase	Cu
S27 / G28	0.71	storage protein	unknown
P27 / Y28	0.71	isomerase (including **PfPGI**)	Fe
A27 / D28	0.16	dioxygenase, oxidoreductase	Fe
G27 / Y28	0.05	oxidoreductase	Fe
V27 / Y28	0.0	-	-
P27 / R28	0.0	-	-
A27 / Y28	0.0	-	-

Cultures of these mutants could be grown as described previously and PfPGI expression could be induced successfully. As a control we included a mutant harboring an empty pET24d, to have a correction for background protein concentrations and to exclude possible background activity. Most *E. coli* proteins could be removed from the cell lysate by a heat treatment step and subsequent centrifugation. The heat stable cell free extract was used for

PfPGI activity measurements, to compare activity of the selected mutants with wild type (WT) PfPGI (Fig.3). In parallel to previous experiments PfPGI was stripped with EDTA to remove any bound divalent cations and subsequently titrated with Mn^{2+} as cofactor, as this cation results in highest *in vitro* activity (Berrisford *et al*., 2003).

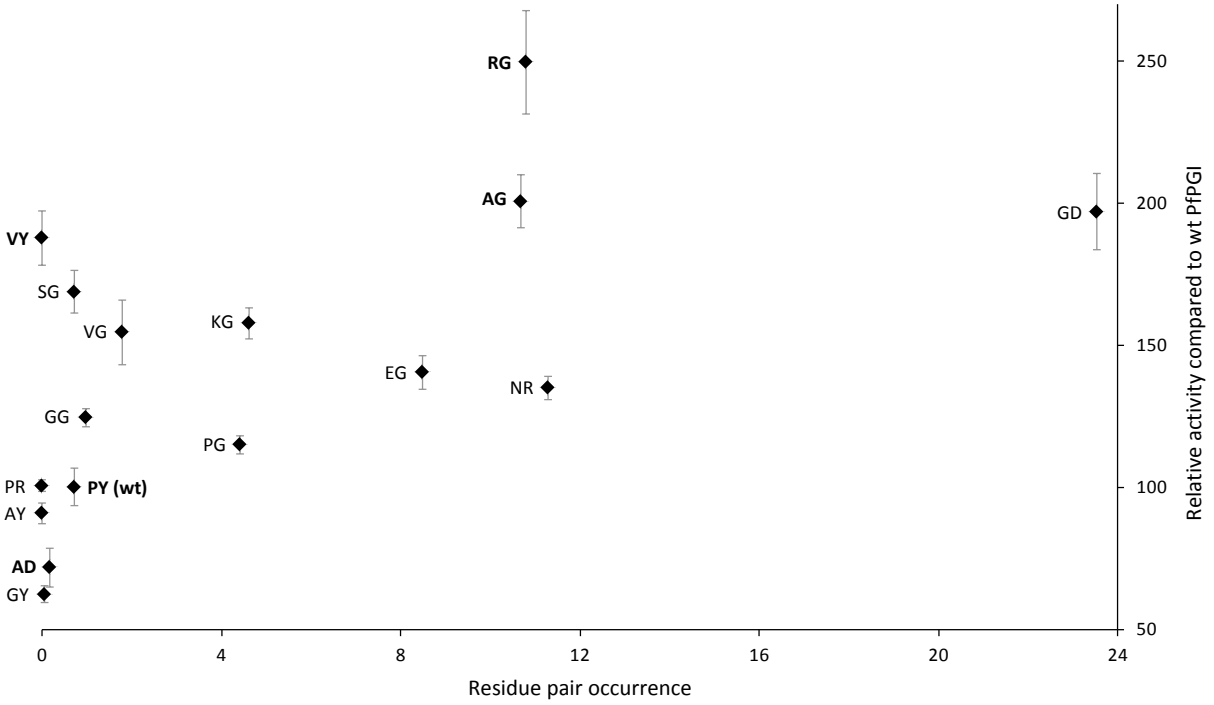


Figure 3 – Graphical representation of the relative PfPGI mutant activity (Y-axis) compared to residue pair occurrence according Comulator predictions (X-axis). All activities are relative to the wild type PfPGI ("PY").

PfPGI activity was detected in lysates of all fifteen mutants, while activity in the lysate of the negative control strain was completely absent. Despite the fact that the selected residues are localized in a surface loop rather than in the catalytic site, significant differences were detected in PGI activity of the different mutants; none of the PfPGI mutants completely lost activity. Interestingly, we observed elevated activities for those amino acid combinations that are abundant in the protein family alignment (pair frequencies; Figs. 1 and 3). Those correlations that are less abundant or absent in the CMA, typically have activity levels that are comparable or lower than wild type PfPGI (amino acid pair "PY"). These findings suggests a positive correlation between the natural prevalence of an amino acid pair, and the activity of the corresponding mutant.

To validate these values we selected some of these mutants for further purification to homogeneity, to enable a more precise analysis of the PfPGI specific activity. Therefore a total of five large batch cultures were grown; WT PfPGI and PfPGI mutants P27A/Y28G, P27R/Y28G, P27A/Y28D and P27V. The first two mutants have a high occurrence compared to wild type, while the other two mutants represent an amino acid combination that is not found in the CMA. Heat stable cell lysate of these five PfPGI mutants was produced and further purified to homogeneity by two additional FPLC steps as described previously. The pure PfPGI samples were used to determine the specific activity of each of these PfPGI mutants (Fig.4).

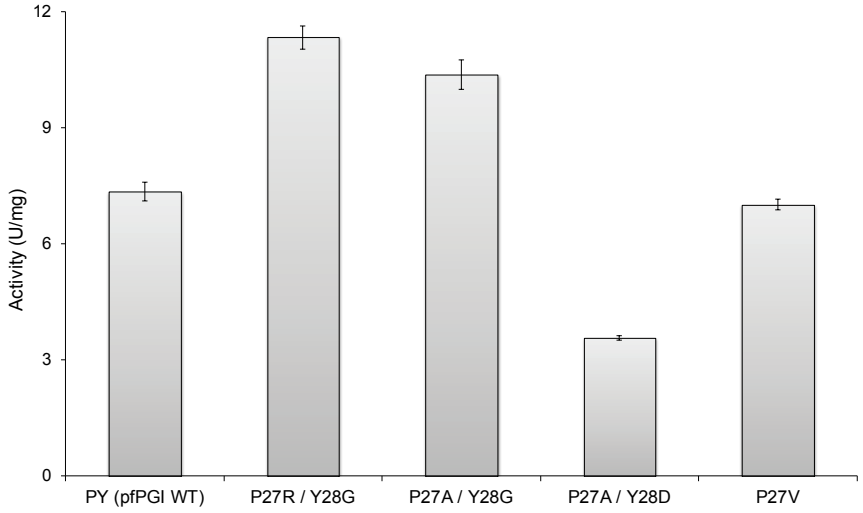


Figure 4 – Specific activity of wild type PfPGI compared to selected single and double mutants.

The specific activities measured for these purified PfPGIs are in good agreement with the previously obtained values as described in Figure 3, although the relative activity presented there was slightly overrated, probably due to difficulties in obtaining accurate protein concentrations in cell lysate. Again we observed an increased activity for both P27R/Y28G and P27A/Y28D in comparison with the wild type protein, while both P27V and P27A/Y28D have respectively a similar or decreased specific activity compared to WT PfPGI.

An important matter to be addressed is the molecular basis of the observed differences in activity. As mentioned, the correlated amino acid pair is located in a surface loop. Despite the fact that this loop is conserved within the cupin family, it is also located rather far from

the catalytic site. It is tempting to speculate that conformational changes in the PfPGI structure could lead to the observed differences in activity.

In contrast to conventional non-cupin PGIs, catalysis of a cupin-type PGI is metal dependent. Removal of divalent metals results in complete loss of activity, which can be restored by the addition of divalent metals. Most divalent cations can function as cofactor, although affinities are different for each specific cupin type PGI (Hansen *et al.*, 2005). Analysis of the 3DM data reveal the possibility that these differences at position 27 and 28 are the result of a difference in cofactor binding, either with respect to specificity or affinity. Although there is no absolute correlation, typically those amino acid pairs that have a higher activity than the wild type, bind manganese (Table 1). This mainly is true for those that have an isomerase function according to literature.

Iron is identified to be associated with PfPGI when isolated both as recombinant protein from *E. coli*, and as native protein from *Pyrococcus furiosus* (Berrisford *et al*., 2003). However, also metals other than Fe^{2+} have been found associated with PfPGI. *In vitro* experiments of PfPGI show highest activity with Mn^{2+} as cofactor (Berrisford *et al*., 2003), which is therefore used as the cation of choice for our experiments. Possibly this is an *in vitro* artefact, as the wild type PfPGI obviously is optimized to function *in vivo*. As long as PfPGI is not rate limiting in the glycolysis of *P. furiosis*, there will not be any selection pressure for protein optimization.

To identify possible differences in metal binding of wild type PfPGI and the developed mutant PfPGIs, we performed metal analysis (ICP-AES) on PGI samples that were not treated with EDTA. Previous studies on wild type PfPGI already indicated that under physiological conditions the occupation of the metal binding sites is likely to be a mixture of divalent metals (Hansen *et al.*, 2005). Comparison of wild type PfPGI with double mutant P27A/Y28G revealed a clear difference in the associated divalent metal. ICP-AES results suggests predominantly manganese for the double mutant, while the wild type manganese levels are eight-fold lower and predominantly Mg^{2+} is bound.

The results suggest that the metal coordination of the different mutants is affected, resulting in different activities. To prove this hypothesis we have initiated crystallization trials of both wild type PfPGI and the four PfPGI mutants. The crystallizations are set up with manganese as incorporated cofactor. Multiple well diffracting crystals (1.5 - 2.1 Å) have been obtained (Table 2). Structure comparison of PfPGI wild type and mutants is ongoing.

Table 2 – PfPGI mutants for which well-diffracting crystals were obtained

PfPGI crystal	Cofactor / substrate present	Resolution
P27R / Y28G	manganese	2.0 Å
P27R / Y28G	manganese and fructose-6-phosphate	1.75 Å
P27A / Y28G	manganese	1.5 Å
P27A / Y28G	manganese and fructose-6-phosphate	1.4 Å
P27A / Y28D	manganese	1.6 Å
P27A / Y28D	manganese and fructose-6-phosphate	2.2 Å
P27V	manganese	1.9 Å

Conclusions

Enzymatic conversions have great potential as green alternative for the chemical catalysis in biotechnological applications. However, natural enzymes are not optimized for industrial biocatalysis. Multiple examples exist of successful enzyme engineering. Mostly these optimizations have been achieved in an evolutionary approach, by random introduction of mutations and subsequent selection for the desired optimized features. Successful application of laboratory evolution urges high-throughput selection methods; however, lack of an appropriate screen restricts the chance on success dramatically.

Rational mutagenesis approaches are an alternative way of protein engineering, where only one or a few site specific mutations are introduced at carefully selected amino acid positions. Correct prediction of a specific substitution at a certain position is very difficult and often has been proven a less successful approach. In the present study, we used correlated mutation analysis (CMA) as a tool to identify correlated amino acids residues, which likely have a high chance of affecting the specific activity of an enzyme. The observed correlations represent hot spots that are targets for variation in 'smart library' design (Jochens and Bornscheuer, 2010). The residues identified by CMA are limited in their combinatory freedom. Alterations of these amino acid residues have been shown to result in improved features. The work we describe here supports this principle, as we observed a positive link in the prevalence of an amino acid combination in the super family multiple sequence alignment and the activity level of a mutant.

We hypothesize that the observed change in activity originates from a change in metal affinity or specificity. We observed a change in associated divalent metal in double

mutant P27A/Y28G in comparison with wild type PfPGI. We hope to get refined crystal structures of the created mutants in the near future revealing potential changes in cofactor binding, contributing in better understanding of PfPGI and general knowledge in CMA based protein optimization.

Acknowledgement

We would like to thank Peter Nobels (Wageningen, the Netherlands) for his assistance in the ICP-AES experiments.

References

Berrisford JM, Akerboom AP, Turnbull AP, de Geus D, Sedelnikova SE, Staton I, McLeod CW, Verhees CH, van der Oost J, Rice DW, Baker PJ. (2003). Crystal structure of *Pyrococcus furiosus* phosphoglucose isomerase: Implications for substrate binding and catalysis. *Journal of Biological Chemistry* **278** 33290-33297.

Berrisford JM, Akerboom J, Brouns S, Sedelnikova SE, Turnbull AP, van der Oost J, Salmon L, Hardré R, Murray IA, Blackburn GM, Rice DW, Baker PJ. (2004). The structures of inhibitor complexes of *Pyrococcus furiosus* phosphoglucose isomerase provide insights into substrate binding and catalysis. *Journal of Molecular Biology* **343**(3), 649-657.

Berrisford JM, Hounslow AM, Akerboom J, Hagen WR, Brouns SJ, van der Oost J, Murray IA, Michael Blackburn G, Waltho JP, Rice DW, Baker PJ. (2006). Evidence supporting a cis-enediol-based mechanism for *Pyrococcus furiosus* phosphoglucose isomerase. *Journal of Molecular Biology* **358**(5), 1353-1366.

Bradford MM. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **7**, 248-254.

Chica RA, Doucet N, Pelletier JN. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology* **16**(4), 378-384.

Dunwell JM, Khuri S, Gane PJ. (2000). Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiology and Molecular Biology Reviews* **64**(1), 153-179.

Dunwell JM, Purvis A, Khuri S. (2004). Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**(1), 7-17.

Hansen T, Oehlmann M, Schönheit P. (2001). Novel type of glucose-6-phosphate isomerase in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Journal of Bacteriology* **183**, 3428-3435.

Hansen T, Schlichting B, Felgendreher M, Schönheit P. (2005). Cupin-type phosphoglucose isomerases (Cupin-PGIs) constitute a novel metal-dependent PGI family representing a convergent line of PGI evolution. *Journal of Bacteriology* **187**(5), 1621-1631.

Hansen T, Schönheit P. (2005). *Escherichia coli* phosphoglucose isomerase can be substituted by members of the PGI family, the PGI/PMI family, and the cPGI family. *FEMS Microbiology Letters* **250**, 49-53.

Kuipers RKP, Joosten H, Verwiel E, Paans S, Akerboom J, van der Oost J, Leferink NGH, van Berkel WJH, Vriend G, Schaap PJ. (2009). Correlated mutation analyses on super-family alignments reveal functionally important residues. *Proteins: Structure, Function, and Bioinformatics* **76**(3), 608-616.

Lutz S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology* **21**(6), 734-743.

Jochens H, Bornscheuer UT. (2010). Natural diversity to guide focused directed evolution. *Chembiochem :*

a European journal of chemical biology **11**(13), 1861-1866.

Verhees CH, Huynen MA, Ward DE, Schiltz E, de Vos WM, van der Oost J. (2001). The phosphoglucose isomerase from the hyperthermophilic archaeon *Pyrococcus furiosus* is a unique glycolytic enzyme that belongs to the cupin superfamily. *Journal of Biological Chemistry* **276**, 40926-40932.

Chapter 7

Summary and general discussion

Raedts J, Kengen SWM, van der Oost J.

Introduction

Glycosaminoglycans like heparin and heparan sulfate have a long history as useful pharmaceuticals. Traditionally these glycosaminoglycans are isolated from livestock derived (waste) tissue. Although this is a rich source of material, several disadvantages related to these materials exist as well, as outlined in **Chapter 1**. Hence, the need for chemo-enzymatic glycosaminoglycan synthesis is obvious. Although much research has been performed on glycosaminoglycan-acting enzymes like PmHAS and the murine heparan sulfate D-glucuronyl C5-epimerase, it has become clear that there are still several hurdles to get over to before industrially applicable enzymes can be developed. To reach successful optimization of the D-glucuronyl C5-epimerase we identified three bottlenecks and made attempts to overcome these. In this concluding chapter, our major achievements are summarized and discussed, and approaches to solve the remaining problems are proposed. Also the achievements in the optimization of PmHAS are described.

Bottleneck one – Stability of glycosaminoglycan modifying enzymes

In this thesis we describe the attempts to improve enzyme production-levels and enzyme stability for two GAG modifying enzymes: the heparan sulfate murine D-glucuronyl C5-epimerase (MmC5) in **Chapter 4** and the *Pasteurella multocida* hyaluronan synthase (PmHAS) in **Chapter 5**. Attempts of heterologous expression of both genes in an industrially suitable host like *Escherichia coli* gave rise to serious limitations in protein production levels and protein stability. As discussed for both constructs, most likely, this is a result of the protein topology *in vitro*.

The heparan sulfate D-glucuronyl C5-epimerase is known to be associated *in vivo* with the 2-O-sulfotransferase (Pinhal *et al*., 2001). It is hypothesized that also other enzymes involved in heparosan chain modification associate with these two enzymes (Sasisekharan and Venkataraman, 2000). Regardless of the exact composition of the *in vivo* complex, it has become clear that heterologous expression of the heparan sulfate D-glucuronyl C5-epimerase is not feasible in *Escherichia coli*. In order to overcome stability issues of individual sub-complex members like the heparan sulfate D-glucuronyl C5-epimerase, we constructed an MBP-MmC5 fusion protein. MBP (maltose binding protein) is known to have a positive effect on the soluble production of heterologous proteins in *E. coli* (Kapust and

Waugh, 1999). Expression experiments show massive production of the non-tagged control MmC5 in *E. coli*, however all protein ends up in inclusion bodies (**Chapter 4**). We made many variations in expression conditions and buffer compositions, however none did result in any solubilization of MmC5. On the other hand our constructed MBP-MmC5 fusion protein was successful as substantial amounts of MBP-MmC5 could be produced in a soluble form. Activity measurements confirm D-glucuronyl C5-epimerase activity in the soluble fraction, providing an alternative way of production of this enzyme (**Chapter 4**).

The second glycosaminoglycan modifying enzyme that we worked on was PmHAS. This enzyme is membrane-associated via a membrane docking sequence. Removal of this domain results in 'solubilization' of PmHAS (Jing and DeAngelis, 2003). However, still most of the produced PmHAS is present in inclusion bodies. Similarly to MmC5, we could increase protein production-levels of soluble PmHAS by fusing the enzyme to the solubilizing MBP-tag (**Chapter 5**). Protein solubility levels could be increased almost threefold, while hyaluronan synthase activity is still present and no deleterious effect on the protein activity was observed.

Protein engineering

Although protein production levels have been increased successfully for both enzymes, further protein optimization may be desirable. Characteristics like enzyme stability and substrate specificity are examples of potential engineering steps. Basically, there are two different approaches for protein mutagenesis; "rational design" and "directed evolution".

Rational design | Rational design (often computational design) is based on site-specific changes of carefully selected amino acid residues, having a predicted relation with particular enzyme characteristics. Changes often are limited to relatively easily identifiable amino acid residues like the active site or other substrate interacting residues (for an example see Machielsen *et al*., 2009). Ideally a 3D-model exists of the structure, resulting in more reliable predictions. However, alternative *in silico* methods to identify important residues gradually arise, of which **Chapter 6** gives a good example of comulator based mutagenesis of PfPGI .

As structural data of the D-glucuronyl C5-epimerase is lacking, no solid predictions for site-specific changes can be made; however, some speculation of possible key residues are provided below. In order to allow for future structural analyses, we made two truncated

variants of the MBP fused MmC5, providing indications on the relevance of the N-terminal domain (**Chapter 4**).

Directed evolution | When a 3D-model of the enzyme of interest is lacking, or for improvement of poorly understood residues, protein improvement is still possible by a strategy, known as laboratory evolution or directed evolution. This method is based on the generation of a large set (library) of variants of a gene of interest. There are various ways to achieve this, including randomly introduction of a limited number of point mutations by for example error prone PCR (Cadwell and Joyce, 1992), or recombination of DNA fragments by gene shuffling of natural variants (Stemmer, 1994).

A crucial aspect of the 'laboratory evolution' approach, is a suitable screening and selection method (review: Boersma *et al*., 2009). A commonly used phrase regarding directed evolution is "you get what you screen for". To screen for an improved parameter of choice, the required method needs to be both selective and high-throughput. Depending on the screening/selection method used, libraries easily can be expanded from thousands to millions of mutants.

Knowing these described techniques, we aimed to improve the enzyme stability of PmHAS and MmC5. The idea was, in analogy to the MBP-tag, to use a colored protein tag like mCherry or GFP (green fluorescent protein) instead, facilitating easy monitoring of soluble protein (Waldo, 2003). By challenging the mutant library with conditions that denature the wild type, mutants with improved stability should remain soluble. Unfortunately, both the mCherry and GFP experiments failed due to lack of solubility of both tags in *E. coli* when fused to PmHAS/MmC5 (unpublished results). However, this method could still be a useful tool for enzyme optimization, since we demonstrated that the addition of an N-terminal protein tag had no negative effect on either protein. Crucial is sufficient solubility of the protein of interest / colored protein-tag fusion product in *E. coli*. Alternatively, the fusion protein of choice can provide a selection advantage, for example a fusion protein with an antibiotic resistance marker, providing antibiotic resistance as long as the protein of choice is produced in a soluble form.

Bottleneck two – Screen for D-glucuronyl C5-epimerase activity

Absolutely essential for the improvement of an enzyme is a reliable and reproducible method to assay the enzyme activity. Preferably the assay is fast, cheap and easy. For the D-glucuronyl C5-epimerase no such assay exists to date. We tested a variety of possible methods. Some have been described in earlier chapters, here we give a complete overview of considered and tested assays.

Hexuronic acid coloring assay

Many high-throughput enzyme assays are based on the development or change in color upon the enzymatic reaction. The resulting changes quickly and easily are observed visually or using spectrophotometry. Examples of such assays are described in **Chapter 5** and **Chapter 6** of this thesis, using a spectrophotometric method to monitor the decrease in NADH levels.

Also coloring reactions specifically for hexuronic acids do exist. They have been developed and described already 65 years back (Dische, 1946) and are based on the reaction of hexuronic acids with carbazole and sulfuric acid, resulting in a pink color development. The method has been used to identify hexuronic acids in glycosaminoglycans (Hoffman *et al*., 1956). Differences in color development are claimed; D-glucuronic acid (GlcA) gives three times more intense coloring than L-iduronic acid (IdoA). Over time the assay conditions have been further optimized (Radhakrishnamurthy and Berenson, 1963). The addition of borate to the concentrated sulfuric acid results in better coloring (Bitter and Muir, 1962). Substitution of carbazole by meta-hydroxydiphenyl (mHDP) (Blumenkrantz and Asboe-Hansen, 1973) and addition of sulfamate (Filisetti-Cozzi and Carpita, 1991), results in reduced interference from neutral sugars, and thus in better specificity for hexuronic acids.

We tested hexuronic acid detection in 96-well plates, based on earlier reports (van den Hoogen *et al*., 1998; Cesaretti *et al*., 2003). Using different sugars (1 mg/ml each) we tested for mHDP based color formation (Fig.1). Indeed the coloring is rather specific for hexuronic acids. Unfortunately, we could not observe differences in color formation between heparin (mainly IdoA) and GlcA. Therefore this assay can be considered adequate to screen for formation of hexuronic acids, but insufficient for detection of hexuronic acid epimerization.

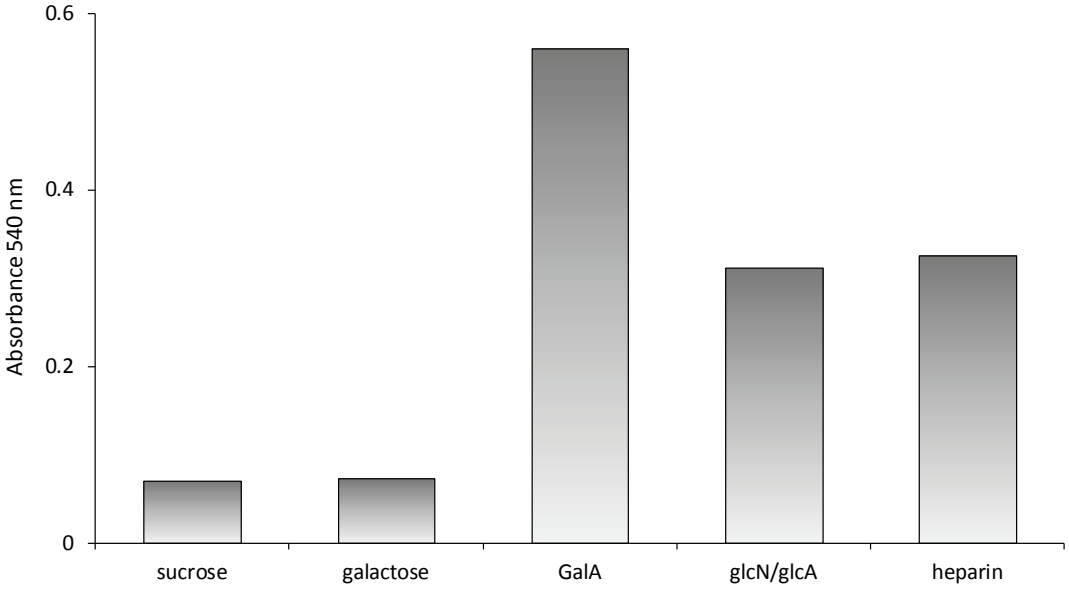


Figure 1 – Meta-hydroxydiphenyl (mHDP) based color formation of different sugars (1mg/ml)

Tritium labeled substrate

A widely used method to assay D-glucuronyl C5-epimerase activity is based on the exchange of the C5-proton with water upon the epimerization reaction. This method makes use of a tritium labeled substrate and enzyme activity results in exchange of tritium with water resulting in $^{3}H_2O$ formation. Upon biophasic separation of the $^{3}H_2O$ from the labeled substrate, the substrates remain in the water phase, while $^{3}H_2O$ is extracted to an organic phase. Any accumulation of radioactivity in this organic phase is the result of D-glucuronyl C5-epimerase activity (Campbell *et al*., 1983; Kusche *et al*., 1991). This accumulation of tritium in the organic phase can be measured using a scintillation counter.

This method has been used in **Chapter 3** and **Chapter 4** to measure RED-C5-epimerase and MBP-MmC5 D-glucuronyl C5-epimerase activity, respectively. Although this method is not exactly high-throughput, measurement of multiple samples is certainly feasible. Another advantages include the methods' sensitivity and the long history of using this method, which allows comparison with earlier results. An absolute disadvantage, and therefore the biggest bottleneck of this method, is the availability of labeled substrate. Every single substrate that should be tested needs to be tritium labeled, which is a laborious and expensive task. This makes this method not very suitable for testing a wide variety of different substrates or D-glucuronyl C5-epimerase orthologs with alternative substrate specificities.

HPLC-PAD

A method independent of the used substrate, has been described in **Chapter 3**. Here we show a method to separate D-glucuronic acid from L-iduronic acid by high performance anion exchange chromatography (HPAEC). We have optimized a protocol for analysis of a glycosaminoglycan hydrolysis mix on a Dionex system (Fig.2A/B), by which we monitored hexuronic acids with an electrochemical detector using pulsed amperometric detection (PAD). This sensitive detection method is based on electrocatalytical oxidation of carbohydrates at a gold electrode by application of a positive potential and a alkalic environment. Proportional to the carbohydrate concentration a current is generated, that is used to detect and quantify carbohydrates. To prevent loss of signal due to electrode poisoning, a series of pulsed potentials is applied.

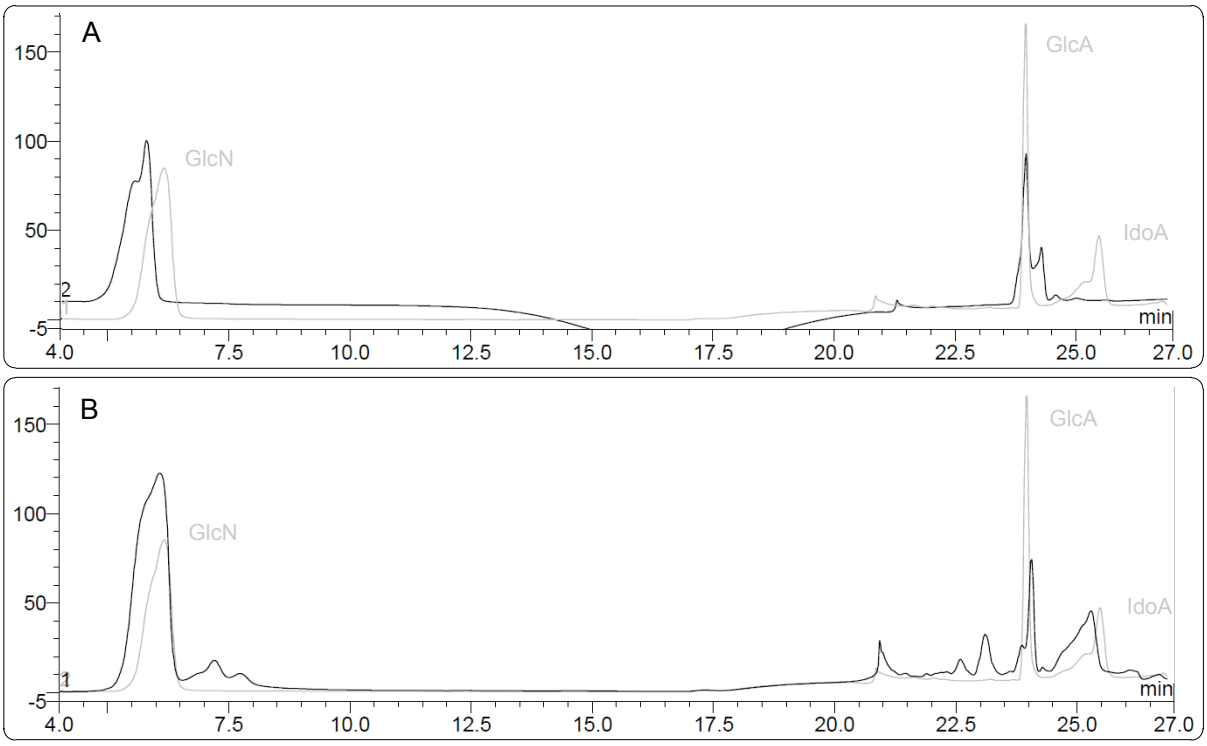


Figure 2 – Chromatogram HPLC-PAD separation of glucosamine (GlcN), D-glucuronic acid (GlcA) and L-iduronic acid (IdoA). (A) Addition of salt disturbs retention-time of mainly IdoA. (B) Hydrolysis profile of de-O-sulfated (N-sulfated) heparin. Standards are shown in grey, samples are shown in black.

Although this method is highly sensitive and virtually any substrate can be used, it also requires tightly controlled depolymerization conditions of the used substrates and requires proper control samples. Since the depolymerization involves high concentrations of

hydrochloric acid at high temperatures, unwanted degradation products are readily formed. Parameters as hydrolysis time, temperature and pH need to be tightly controlled, as well as the buffer and salt concentration. This method is also rather laborious. Glycosaminoglycans are typically hard to hydrolyze, resulting in long hydrolysis times at elevated temperatures. Upon hydrolysis, the hydrochloric acid needs to be completely removed by an evaporator. Both steps take several hours, implying that this method is not useful for screening large libraries of C5-epimerase variants.

Nitrous acid based depolymerization

Depolymerization of heparin / heparan sulfate is not an easy task since these glycosaminoglycans are known to be remarkably resistant to acid hydrolysis. However, it has been known for over half a century that glycosidic bonds of amino sugars with an unsubstituted amino group (e.g. glucosamine) can be easily cleaved by nitrous acid (Conrad, 1993). Addition of nitrous acid (pH 1.5 at 100°C) removes all N-sulfate groups instantly, while 2-3 h incubation is sufficient to remove all N-acetyl groups (Shively and Conrad, 1976). Subsequent nitrosation of the unsubstituted amino group leads to loss of N_2, accompanied with the conversion of glucosamine to 2,5-anhydro-D-mannose (Fig.3A). As a result the glycosidic bond with the neighboring hexuronic acid is disrupted, leading to depolymerization.

We have tried to implement this way of depolymerization in the previously described HPLC-PAD method. However, two problems were encountered. Firstly there is a need of both 2,5-anhydro-D-mannose / hexuronic acid disaccharide standards. Secondly, the use of nitrous acid results in an accumulation of salt in the samples which is known to disturb the separation of D-glucuronic acid and L-iduronic acid (Fig.2A). Although both problems can be overcome, the benefits versus hydrochloric acid based hydrolysis were considered insufficient to continue with this method.

Heparin Lyases

Several lyases have been described that are able to depolymerize heparin. The bacterium *Flavobacterium heparinum* has three different heparin lyases (or heparinases), each with a typical substrate specificity (Desai *et al*., 1993). Heparinase II has the broadest substrate specificity and can cleave substrates which contain D-glucuronic acid as well as L-iduronic

acid. Heparinase I (EC 4.2.2.7) and Heparinase III (EC 4.2.2.8) only specifically cleave the linkage GlcN-IdoA and GlcN-GlcA, respectively.

The substrate specificity of Heparinase I could make this lyase a potential tool to assay C5-epimerization activity. Cleavage mediated by lyases proceeds via a β-elimination reaction on the glycosidic bond. As a result both GlcA and IdoA are converted to a Δ4,5-unsaturated hexuronic acid (Fig.3B). The double bond between C4 and C5 of the reaction product can be detected easily by UV spectroscopy at 232 nm. It has been reported that Heparinase I is less active on heparin with reduced amounts of (especially 2-O) sulfation (Desai *et al*., 1993).

Since heparosan is completely O-unsulfated after the epimerization step (only the glucosamine is deacetylated / resulfated), it is questionable whether Heparinase I is able to accept this product as a substrate. We performed a pilot experiment to test activity of Heparinase I (Sigma-Aldrich) on epimerized N-sulfated heparosan (K5NSepi; Iduron). We could not detect any cleavage, and therefore did not proceed. However, potentially this method is useful in a combined assay with 2-O-sulfotransferase, giving rise to a rather fast and easy assay.

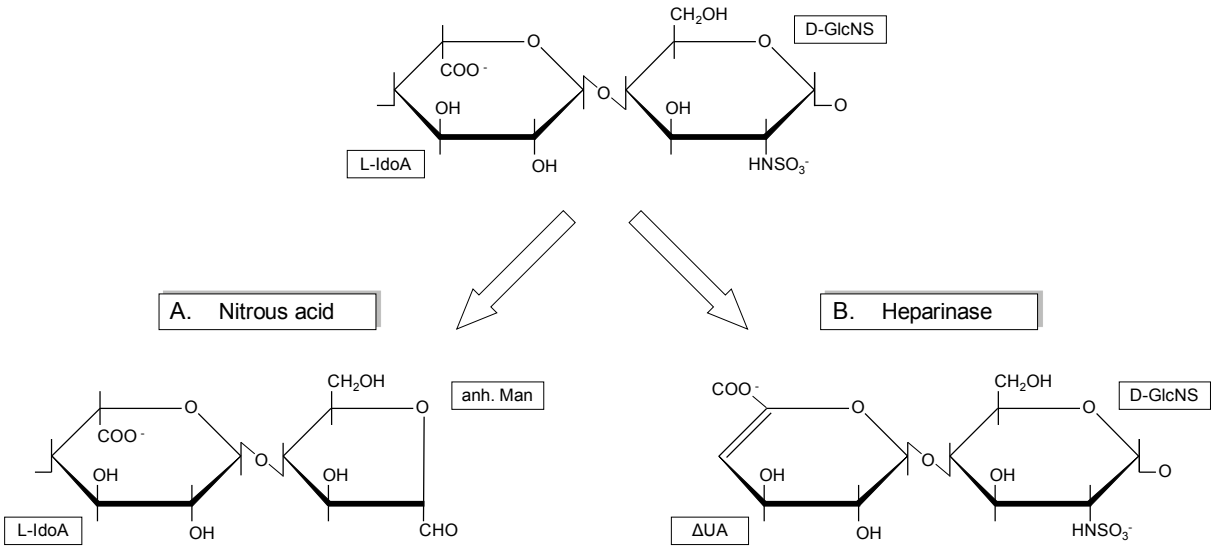


Figure 3 – Two methods leading to heparin depolymerization. (A) Nitrous acid efficiently removes the sulfate group from the glucosamine, which subsequently is converted to 2,5-anhydro-D-mannose (anh. Man), resulting in cleavage of the glycosydic bond. (B) Heparinases catalyze a β-elimination reaction on the glycosidic bond also resulting in cleavage. Both GlcA and IdoA are converted to the same Δ4,5-unsaturated hexuronic acid (ΔUA).

Antibodies

Antibodies are known for their ability to recognize epitopes in a highly specific way, making them very useful for specific detection. Although heparin / heparan sulfate are non-alien molecules in mammals and thus typically raising little immunogenic reaction, antibodies successfully have been generated using phage display (Dennissen *et al*., 2002). An antibody has been described that recognizes N-sulfated-D-glucosamine linked to L-iduronic acid, while D-glucuronic acid and O-sulfation are no epitopes (Kurup, 2007). We tested specificity binding properties of this antibody (HS4E4) for heparan sulfate (HS) and K5 polysaccharide, N-sulfated K5 (K5NS) and K5NS that is C5-epimerized (K5NSepi).

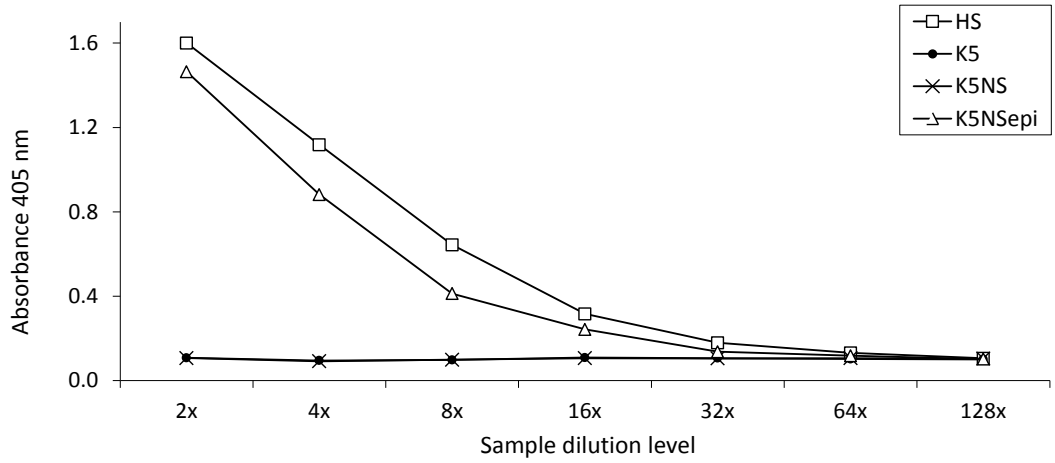


Figure 4 – Recognition and sensitivity of antibody HS4E4 on different heparin-like structures

Figure 4 shows good discrimination between non-epimerized substrates (K5 and K5NS) versus epimerized K5NS. Antibody HS4E4 therefore potentially can be used as a tool to identify D-glucuronyl C5-epimerase activity in future experiments. The sensitivity on partially epimerized substrate still is subject for further testing. However, the potential as a high-throughput screening method of such an antibody is evident.

Biosensors

All above mentioned assays are based on *in vitro* activity measurements of the enzyme. An alternative approach would be the detection of D-glucuronyl C5-epimerase activity *in vivo*. A possible solution we have considered, is based on the development of a so-called biosensor. The basics of such a tool are based on an binding module that specifically senses L-iduronic acid, while D-glucuronic acid is not recognized. Additionally this sensing capacity must be

coupled to some kind of reporter module. The binding of L-iduronic acid to the binding module should result in a conformational change, which activates the reported module. A schematic representation of such system can be seen in figure 5.

An example of such ligand-binding biosensor has been demonstrated by Guntas and co-workers (Guntas *et al*., 2005). They successfully fused MBP to the β-lactamase (BLA) gene. Their resulting (periplasmatic) biosensor facilitates growth of those cells that bind sucrose, while those that cannot bind this sugar are restricted in growth by the presence of ampicillin. The strength of such tool results from its base of high-throughput selection rather than screening, facilitating the use of very large libraries.

The construction of such tool however is very laborious and risky. Not only should the binding module be very and exclusively selective for binding the formed product, also an adequate conformational change is needed to activate the reporter module. An additional problem for glycosaminoglycans might be the substrate uptake. All described D-glucuronyl C5-epimerases act on the polymer heparosan. *E. coli* will not be able to transport the needed oligomeric sugars across the cell membrane. As a result either selection needs to be for a D-glucuronyl C5-epimerase acting on monomeric D-glucuronic acid, or metabolic engineering steps are required such that the needed substrate can be produced by the cell itself. An approach would be the introduction of the heparosan biosynthesis operon of the pathogen *E. coli* K5, containing all genes required for the production of heparosan (Petit *et al*., 1995; Sieberth *et al*., 1995; Hodson *et al*., 2000; Sugiura *et al*., 2010).

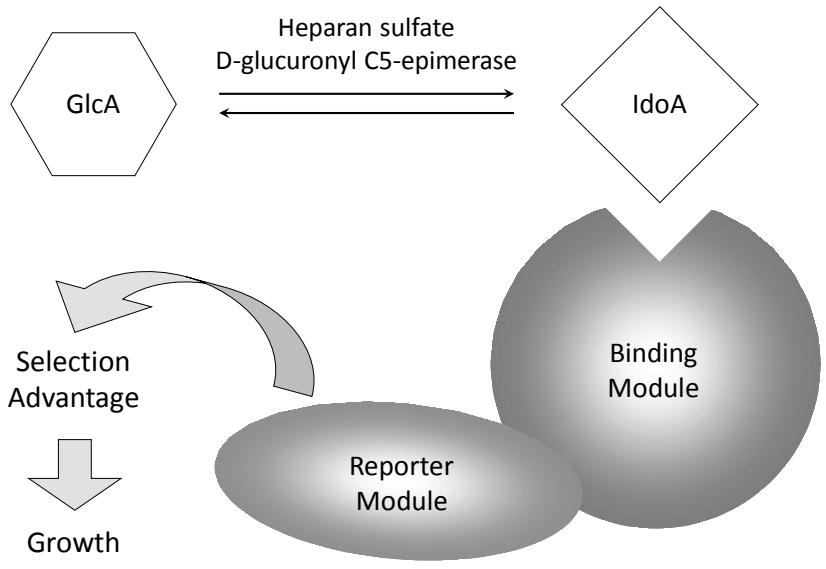


Figure 5
Schematic representation of proposed biosensor. The product formed (IdoA) can be specifically bound by a binding module. The resulting conformational change results in the activation of the reporter module, introducing a selection advantage for the cells harboring D-glucuronyl C5-epimerase activity.

Bottleneck three – Knowledge on glycosaminoglycan modifying enzymes

Since the identification of heparan sulfate D-glucuronyl C5-epimerase as the biocatalyst responsible for L-iduronic acid production in heparin / heparan sulfate, much research has been done, expanding our general understanding of this specific enzyme. Although much knowledge has been gained on the biochemistry, structural knowledge on this enzyme is still lacking.

Key residues of heparan sulfate D-glucuronyl C5-epimerase

Understanding the catalytic mechanism of any enzyme, requires knowledge on the amino acid residues that are involved in catalysis. Close examination of natural variations in D-glucuronyl C5-epimerase homologs may help to identify key residues of the enzyme. Ultimate proof of a key role of residues is typically obtained by site directed mutagenesis of candidate amino acids. These residues generally are selected since they are conserved and/or based on structural data. Site directed mutagenesis of such catalytic residues to dissimilar amino acids, will result in loss of catalytic activity.

Many animal heparan sulfate D-glucuronyl C5-epimerase sequences are available. Almost all have a similar domain distribution and can be well aligned, as seen in **Chapter 4**. An exception is the heparan sulfate D-glucuronyl C5-epimerase gene sequence (accession: XP_002189772) from *Taeniopygia guttata* (zebra finch), which shows a remarkable recombination, with the very C-terminus positioned at the N-terminal part of the protein. However, based on the large number of undetermined nucleotides in the gene sequence data, it cannot be ruled out that this is an artefact due to misassembly of the genome data. Based on the multiple sequence alignment of the other animal heparan sulfate D-glucuronyl C5-epimerase sequences, this paragraph deals with some speculation on possible key residues.

Heparin binding site | Many reports have appeared discussing residues involved in heparin and glycosaminoglycan binding. Based on numerous heparin binding proteins, several heparin-binding consensus sequences have been proposed (Table 1). Essential in glycosaminoglycan binding are basic amino acids (Capila and Linhardt, 2002). Mostly these basic amino acids are lysine (K) or arginine (R), although occasionally it also might be a histidine (H). Arginine is known to bind tighter to GAGs than lysine, while histidine binds

significantly less tightly than the other two. Also glutamine (Q) and tyrosine (Y) residues are enriched in known heparin / heparan sulfate binding sites, they are believed to play a role in inaction via hydrogen bonding. For example for the antithrombin III binding site it has been suggested that the phenolic hydroxyl group of tyrosine might interact with the N-acetyl group (Bae *et al*., 1998).

Table 1 – Consensus glycosaminoglycan binding-sites. Abbreviations: B = basic amino acid; X = hydropathic amino acid; T = turn. Apo = apolipoprotein; AT = antithrombin; FGF = fibroblast growth factor.

GAG binding consensus	Examples binding protein	Reference
XBBXBX	Vitronectin, Fibronectin	Cardin and Weintraub, 1989
XBBBXXBX	Apo E, Apo B, ATIII	Cardin and Weintraub, 1989
XBBBXXBBBXXBBX	von Willebrand Factor	Sobel *et al*., 1992
TXXBXXTBXXXTBB	FGF-1, FGF-2, FGF β-1	Hileman *et al*., 1998

Moreover, molecular modeling studies have shown that these basic residues are distributed in such way, that their corresponding secondary structural conformation typically leads to orientation at one side of the α-helix (XBBBXXBX) or β-strand (XBBXBX). These basic residues (B) will be faced towards the glycosaminoglycan chain, while the hydropathic (neutral/hydrophobic) residues (X) are facing the protein core.

Examination of the murine heparan sulfate D-glucuronyl C5-epimerase amino acid sequence reveals several basic amino acids, also conserved in heparan sulfate D-glucuronyl C5-epimerase homologs in other animals. Of the mentioned GAG binding sequences (Table 1), the motif XBBXBX is also present in the murine protein (residue 603-607). However, multiple sequence alignment analysis indicates that this motif is only conserved in vertebrates. Additionally conflicting are the secondary structure predictions that indicate an α-helix at this position, while the XBBXBX motif is seen for β-strands. Therefore it is likely that other residues are involved in binding instead, in an alternative binding motif. Highly conserved lysine and arginine residues, possibly involved in heparin binding, of known D-glucuronyl C5-epimerases are indicated in Figure 6. Helical wheel diagrams of some of the α-helices are included as Supplementary Figure 1, showing orientation of these conserved residues at one side of the helix.

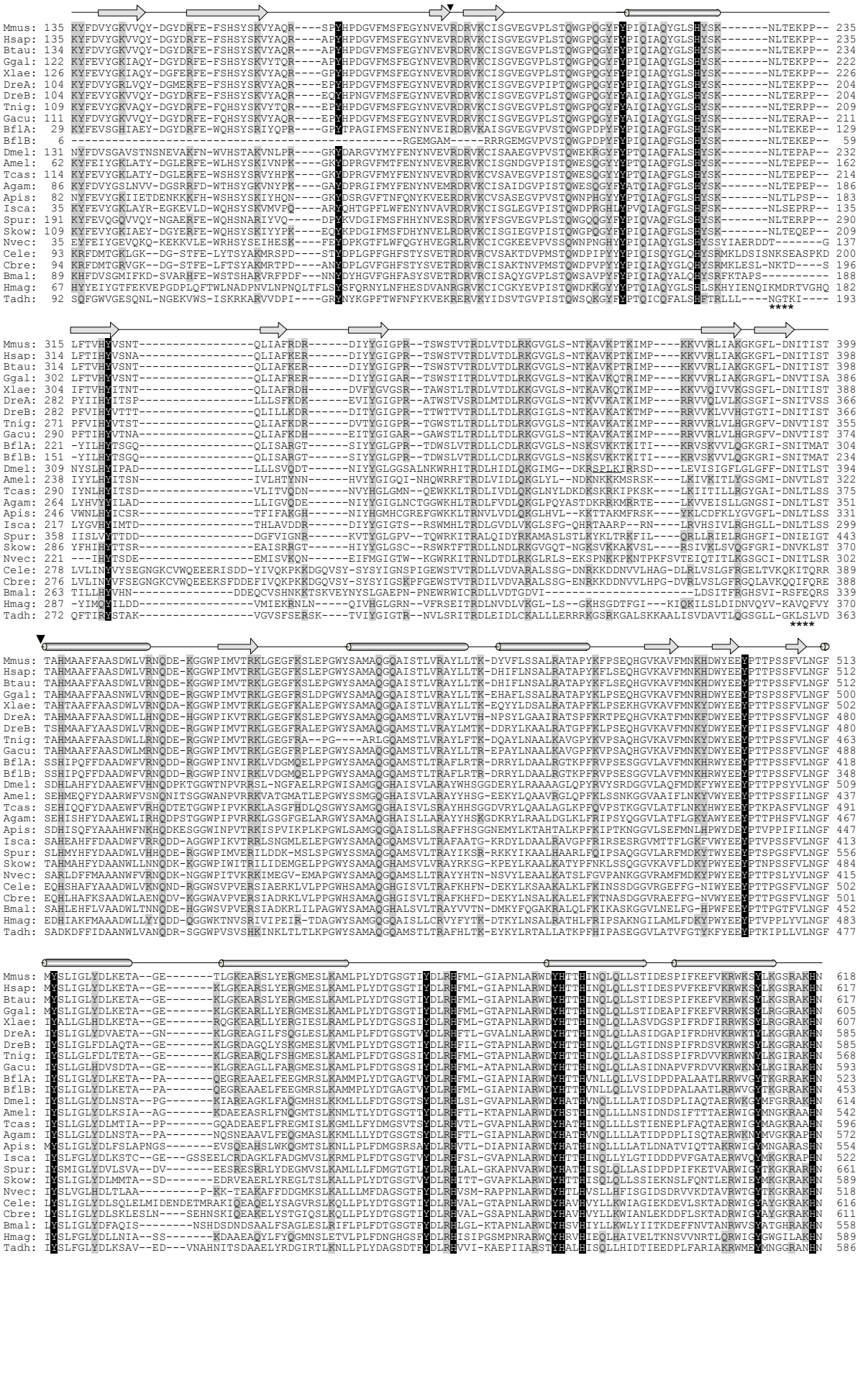


Figure 6 – Multiple sequence alignment animal D-glucuronyl C5-epimerases; conserved residues putatively involved in heparin binding (grey shading) or residues that could be involved in catalysis (black shading). Putative N-glycosylation sites are indicated with asterisks (below sequence). The start positions of all made truncations are indicated as triangles. Secondary structure predictions are indicated as cylinders (α-helices) and arrows (β-strands). Abbreviations: Mmus: *Mus musculus* (mouse); Hsap: *Homo sapiens* (human); Btau: *Bos taurus* (cow); Ggal: *Gallus gallus* (chicken); Xlae *Xenopus laevis* (frog); DreA: *Danio rerio* isoform A (zebrafish); DreB: *Danio rerio* isoform B (zebrafish); Tnig: *Tetraodon nigroviridis* (pufferfish); Gacu: *Gasterosteus aculeatus* (three spined stickleback); BflA: *Branchiostoma floridae* isoform A (lancelet fish); BflB: *Branchiostoma floridae* isoform B (lancelet fish); Dmel: *Drosophila melanogaster* (fruit fly); Amel: *Apis mellifera* (honey bee); Tcas: *Tribolium castaneum* (red flour beetle); Agam: *Anopheles gambiae* (malaria mousquito); Apis: *Acyrthosiphon pisum* (pea aphid); Isca: *Ixodes scapularis* (black-legged tick); Spur: *Strongylocentrotus purpuratus* (purple sea urchin); Skow: *Saccoglossus kowalevskii* (Acorn worm); Nvec: *Nematostella vectensis* (starlet sea anemone); Cele: *Caenorhabditis elegans* (nematode); Cbre: *Caenorhabditis brenneri* (nematode); Bmal: *Brugia malayi* (filarial roundworm); Hmag: *Hydra magnipapillata* (freshwater polyp hydra); Tadh: *Trichoplax adhaerens* (Placozoa).

Catalytic residues | Experimental proof suggests the C5-epimerization reaction to proceed via the abstraction and re-addition of the proton at the C5-atom (**Chapter 1**). This requires catalytic residues acting as a proton acceptor and a proton donor. The re-added proton is not the C5-abstracted proton, but a proton from the medium (water). Likely there are at least two residues involved in catalysis in a two-base mechanism; one for deprotonation and one for reprotonation.

While the mechanism of the heparan sulfate D-glucuronyl C5-epimerase still needs to be elucidated, molecular mechanisms have been proposed for the other hexuronyl C5-epimerases. Dermatan sulfate D-glucuronyl C5-epimerase are believed to proceed via a β-elimination via a 4,5-unsaturated hexuronic intermediate step (Pacheco *et al*., 2009), similarly as seen for heparinase cleavage. However, in contrast to the β-elimination step in lyase activity, the C5-epimerization is not finished at this intermediate, but instead the glycosidic linkage is restored and the C5-atom is reprotonated. Two histidine residues (His-205 and His-450) are believed to act as general base and general acid in the reversible epimerization reaction. A third residue is tyrosine residue Tyr261 and is believed to be essential in mediating the β-elimination reaction. This is based on a equivalently positioned tyrosine in the structure of heparinase II (Shaya *et al*., 2006).

A similar mechanism has been proposed for alginate D-mannuronan C5-epimerases (Jerga *et al*., 2006). However, structural data for AlgE4A suggests another mechanism, in which the glycosidic linkage remains intact (Rozeboom *et al*., 2008). This mechanism is

similar to the proposed mechanism of some polygalacturonic acid lyases (Charnock, 2002). Instead of β-elimination, the removal of the C5-proton leads to the formation of a carbanion, with a delocalized negative charge at the carboxyl-group. For the polygalacturonic acid lyases this negative charge is stabilized by the divalent cation Ca^{2+}, but for the alginate C5-epimerase the carboxylate group is likely to be protonated by two aspartate residues, enhancing stability and reactivity instead. A tyrosine residue (Tyr-149) acts as a general base, responsible for the C5-proton abstraction. To prevent lyase activity, the abstracted proton is removed from the active site to prevent donation of the proton to the glycosidic oxygen. The proposed general acid responsible for the reprotonation is a histide residue (His-154).

Via which catalytic mechanism the heparan sulfate D-glucuronyl C5-epimerase acts is hard to predict. We propose two alternative mechanisms (Fig.7), based on the mechanisms of the other two hexuronyl C5-epimerases. A β-elimination reaction mechanism (Fig.7A) similar to dermatan sulfate D-glucuronyl C5-epimerase would require a general base (B:), a general acid (B:H) and a residue that facilitates the β-elimination (A:). A mechanism similar to that of the alginate C5-epimerase (Fig.7B) would also require a general base (B:), a general acid (B:H), but additionally needs a stabilizing factor. In lyases having such mechanism this stabilization generally is achieved via ionic interaction with the divalent cation Ca^{2+}. However, we could not detect any calcium ion in the heparan sulfate D-glucuronyl C5-epimerase (**Chapter 4**). This not necessary excluded the possibility of such mechanism, as also the alginate C5-epimerase lacks a calcium ion in the active site, so this stabilizing factor could be analogous to the alginate C5-epimerase.

The catalytic mechanisms of either known hexuronic acid C5-epimerases have a highly conserved tyrosine and one or two histidines involved. Only few histidines and tyrosines are conserved in the family of heparan sulfate D-glucuronyl C5-epimerases (Fig.6). Especially promising are Tyr-168 and Tyr-222, since recent research by Li and co-workers indicated complete loss of activity upon substitution of these residues by alanine (Li *et al*., 2010). Based on our multiple sequence alignment (Fig.6) we only propose a potential role in catalysis for Tyr-168 which is, in contrary to Tyr-222, completely conserved in both animal heparan sulfate D-glucuronyl C5-epimerase (**Chapter 4**) and RED-C5-epimerase (**Chapter 3**).

No studies have been performed yet on the conserved C-terminal residues. Very promising are the C-terminal residues Tyr-561 and Tyr-579, since these residues also are completely conserved in all the prokaryotic sequences (**Chapter 2**). Additionally promising

are residues Tyr-501, His-580 and His-583 which are well-conserved in prokaryotes, and Tyr-320, His-565, Tyr-609 and His-617. The latter four are not completely conserved in all prokaryotic sequences, however, they are conserved between MmC5 and RED-C5-epimerase (**Chapter 3**). Conserved residues potentially involved in heparin binding are in close proximity as well, especially when considering their spacial orientation based on secondary structure predictions (Supp. Fig. 1). All these conserved residues would be a great start for site directed mutagenesis studies.

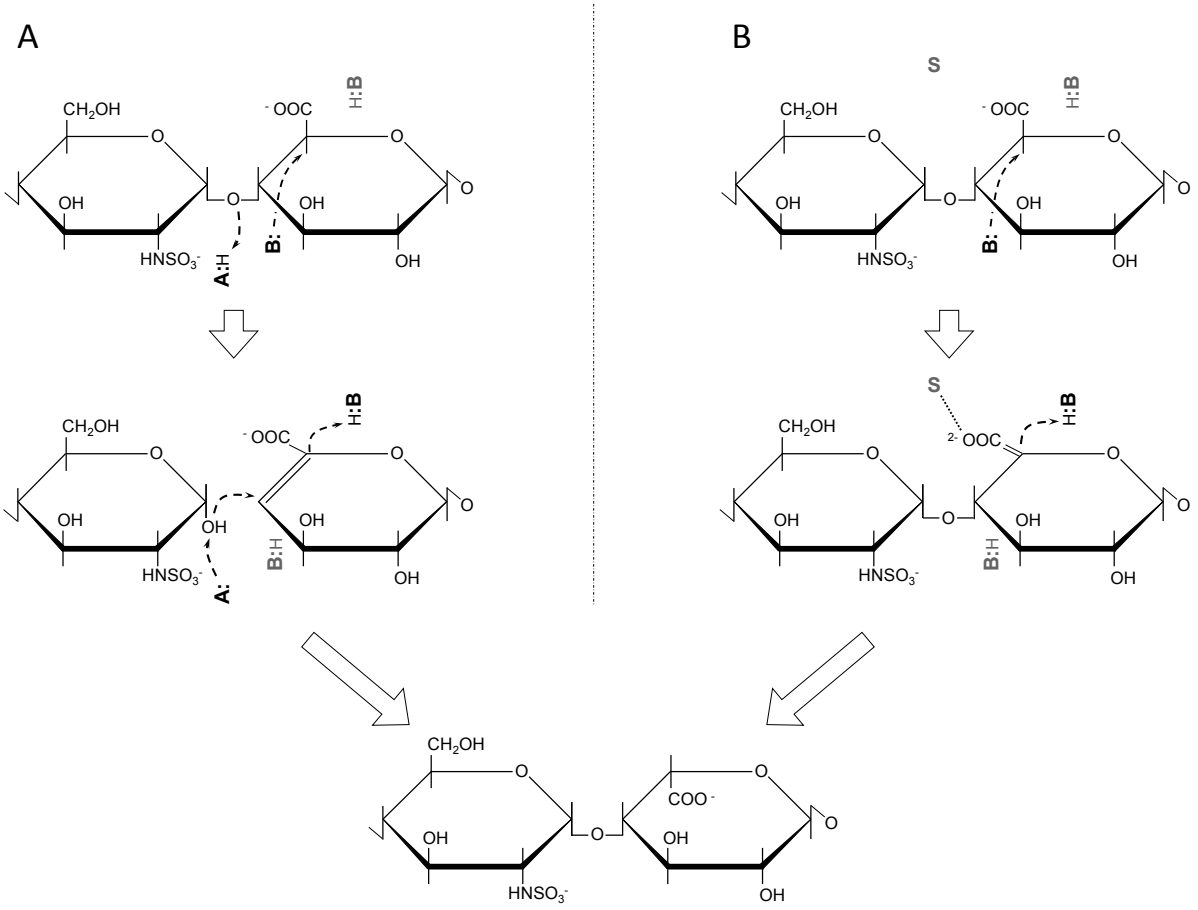


Figure 7 – Possible catalytic mechanisms of the D-glucuronyl C5-epimerase; (A) β-elimination based mechanism. Upon proton abstraction a 4,5-unsaturated hexuronic acid is formed. Reprotonation of C5 results in reformation of the glycosidic linkage. An acidic amino acid "A" facilitates the breakage of the glycosidic bond. (B) Non-β-elimination based mechanism. The removal of the proton leads to the formation of a carbanion, with a delocalized negative charge at the carboxyl-group. This negative charge needs to be stabilized, here indicated by "S". Reprotonation of the carbanion results in the reformation of a hexuronic acid.

Enzyme structure of D-glucuronyl C5-epimerase

To date, structural knowledge of heparan sulfate D-glucuronyl C5-epimerase is still lacking. In **Chapter 4** of this thesis we have discussed secondary structure predictions of the murine heparan sulfate D-glucuronyl C5-epimerase. These secondary structure predictions indicate the C-terminus to be composed of seven α-helices, while the N-terminal domain has no significant homology to any known sequence. Based on predictions with PHYRE Protein Fold Recognition Server (http://www.sbg.bio.ic.ac.uk/phyre/; Kelley and Sternberg, 2009), these helices are likely to be arranged in an (incomplete) alpha/alpha toroid fold. This fold comprises up to seven alpha-hairpins being arranged in closed circular array. A topology diagram of the putative secondary structure organization of the murine heparan sulfate D-glucuronyl C5-epimerase is provided in Figure 8.

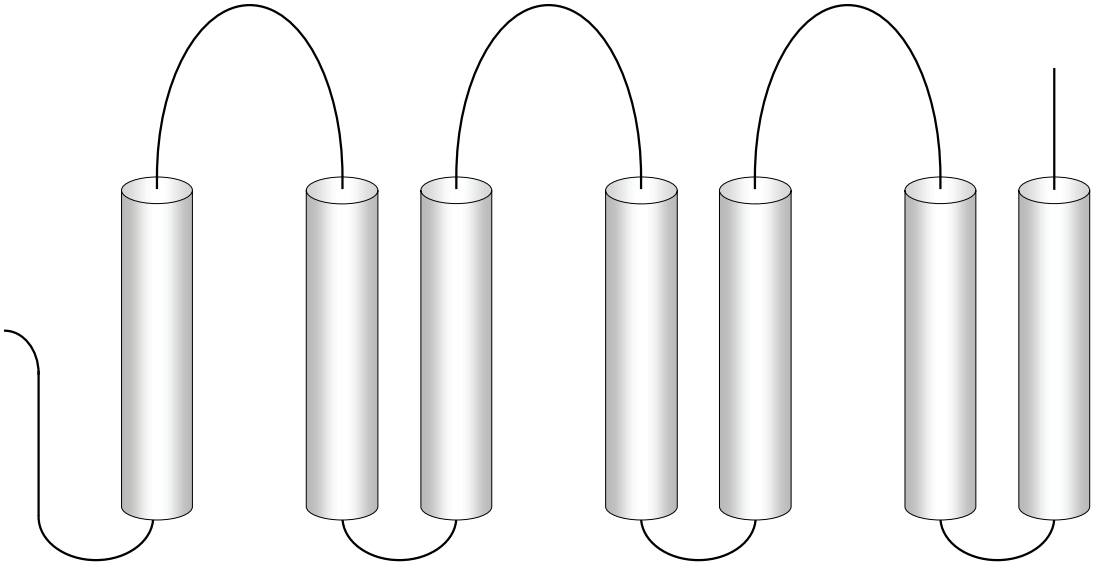


Figure 8 – Topology diagram of the seven predicted α-helices (cylinders) of heparan sulfate D-glucuronyl C5-epimerase carboxy-terminus.

Members that are proven to have an alpha/alpha toroid fold include glycosidases and lyases (e.g. the earlier mentioned polygalacturonic acid lyase (Charnock, 2002) and heparinase II). Also the 13 α-helices of the N-terminal domain of DS-epimerase 1 are predicted to adopt in an incomplete alpha/alpha toroid fold (Pacheco *et al*., 2009). Unfortunately, none of these known alpha/alpha toroid protein structures, nor any other structure in the Protein Data Bank (PDB), has sufficient homology with a heparan sulfate D-glucuronyl C5-epimerase to generate a reliable 3D-model.

Despite the lack of template structures with sufficient homology, we tried several methods of protein threading, scanning the C-terminal amino acid sequence (aa 400-618) of the heparan sulfate D-glucuronyl C5-epimerase against a database of solved structures (Table 2).

Table 2 – Used protein threading methods to model the heparan sulfate D-glucuronyl C5-epimerase C-terminus

Model	Method	Templates	Website (Reference)
Green	Phyre	1gmx	http://www.sbg.bio.ic.ac.uk/~phyre (Kelley and Sternberg, 2009)
Blue	hhPred / modeller	3gt5, 3k7x, 2gz6, 3pmm, 1gxm	http://toolkit.tuebingen.mpg.de/hhpred (Sali *et al*., 1995; Söding *et al*., 2005)
Red	Itasser	2okx, 3k7x, 1gxm, 1r76	http://zhanglab.ccmb.med.umich.edu/I-TASSER (Zhang, 2008; Roy *et al*., 2010)
Purple	Robetta	2zbl	http://robetta.bakerlab.org (Chivian *et al*., 2003; Kim *et al*., 2004)

Used template structures include (pectate) lyases, glycosyl hydrolases and N-acetylglucosamine 2-epimerases (Table 3), all proteins with potential similar molecular mechanisms as a C5-epimerase.

Table 3 – Used templates for protein threading

Template	Description	Organism
1gxm	Pectate lyase	*Cellvibrio cellulosa*
1r75	Pectate lyase	*Azospirillum irakense*
2okx	Rhamnosidase	*Bacillus Sp.*
2zbl	Putative isomerase (similar to N-acetylglucosamine 2-epimerase)	*Salmonella enterica*
2gz6	N-acetylglucosamine 2-epimerase	*Anabaena Sp.*
3gt5	N-acetylglucosamine 2-epimerase	*Xylella fastidiosa*
3k7x	Lin 0763 protein (glycosyl hydrolase fam76)	*Listeria innocua*
3pmm	Possible member GH105 family (rhamnogalacturonyl hydrolase)	*Klebsiella pneumonia*

Interestingly, the generated models all show the predicted alpha/alpha toroid fold of the α-helices with an RMSD < 5, making us confident that the C-terminus of heparan sulfate D-

glucuronyl C5-epimerase can adopt a similar configuration. However, none of these generated models successfully passes quality checks (Z-scores ≤-2), and obviously, large deviation is seen for the coils linking the helices. Even though the exact orientation of the helices could not be determined, we propose a tentative model in Figure 9, indicating the seven predicted α-helices and the predicted alpha/alpha toroid fold.

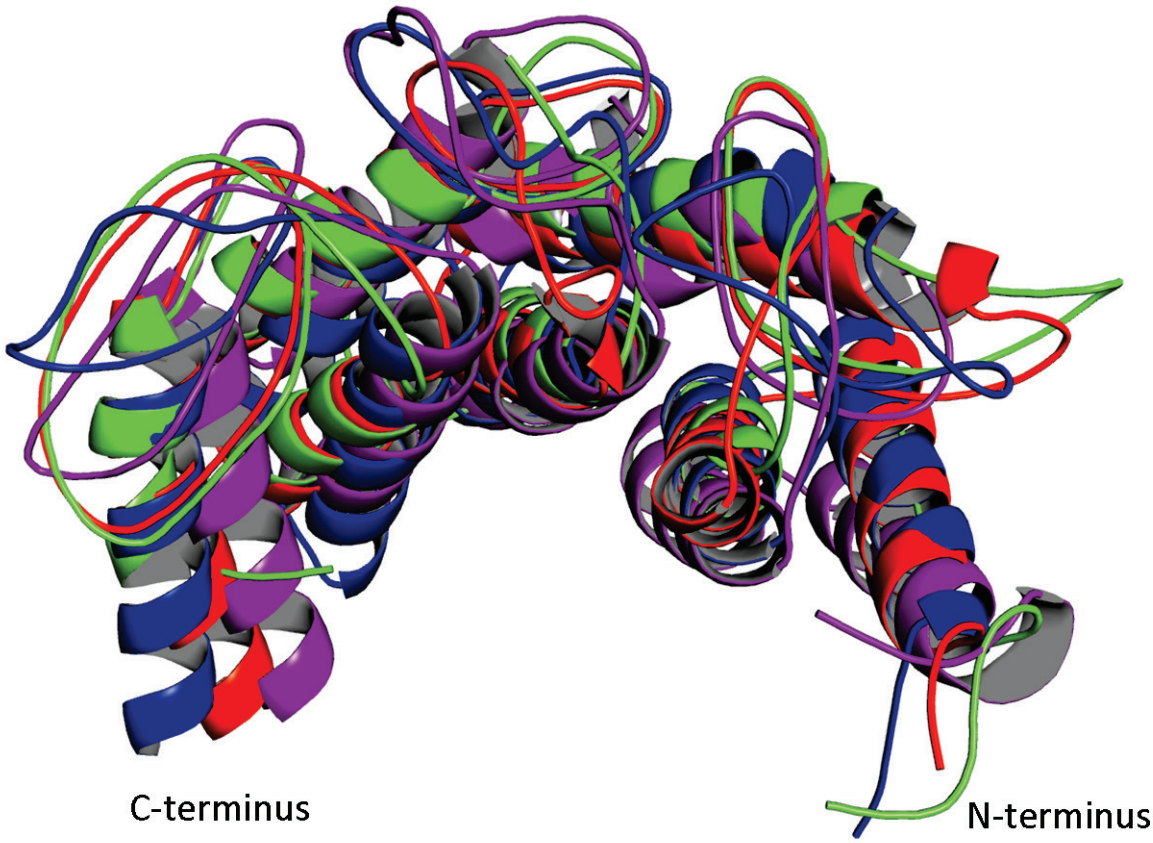


Figure 9 – Structure modelling of the heparan sulfate D-glucuronyl C5-epimerase carboxy-terminus.

However, to get ultimate confirmation of the heparan sulfate D-glucuronyl C5-epimerase structure, a high resolution crystal structure is required. And although we have made many attempts to crystallize the bacterial candidate C5-epimerases RED-C5-epimerase and TTE-C5-epimerase (**Chapter 3**), none of these trials have been successful thus far.

In conclusion

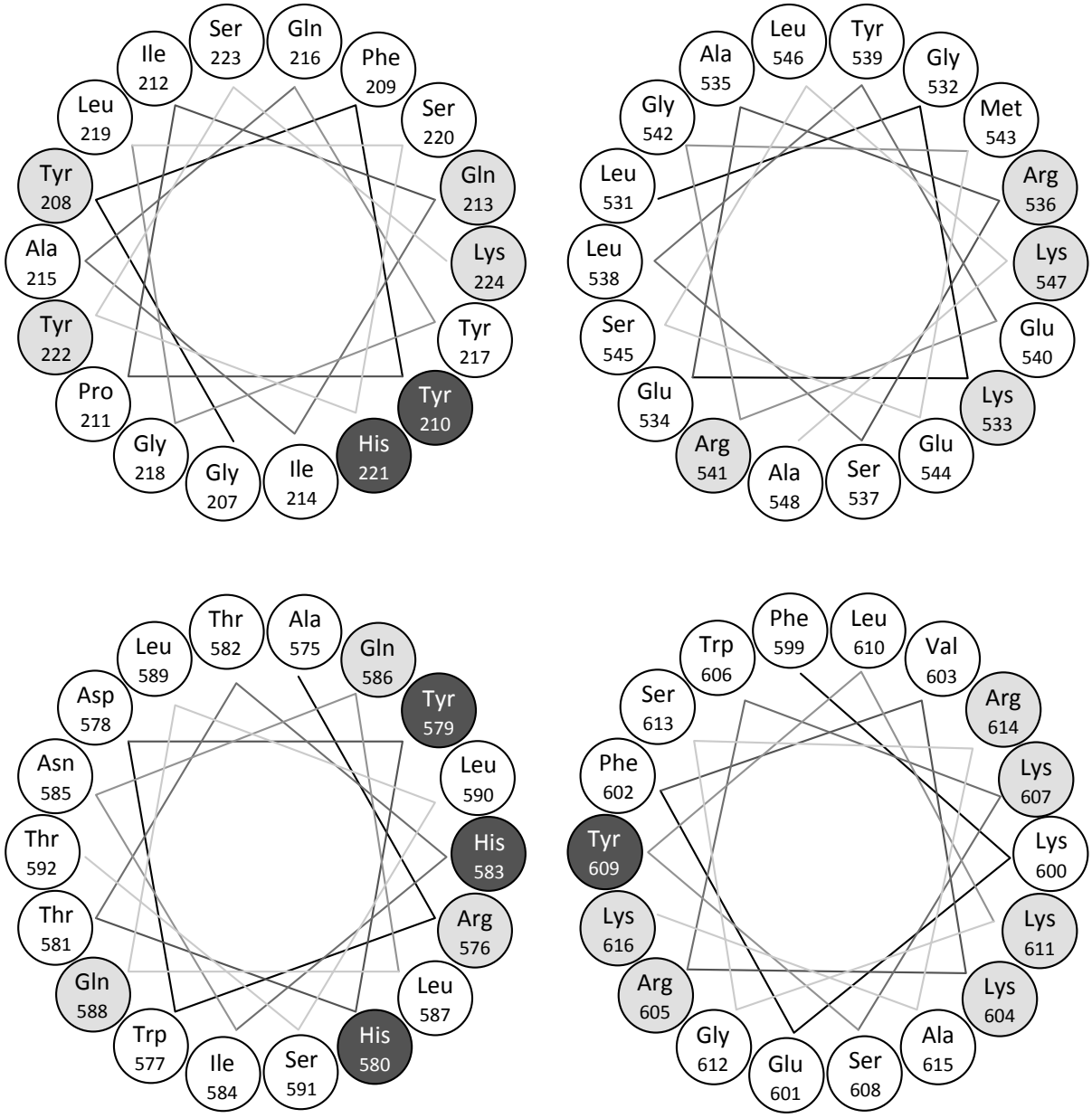
We aimed to develop industrially applicable of the (murine) heparan sulfate D-glucuronyl C5-epimerases, to be used in the chemo-enzymatic synthesis of highly defined heparin variants. The natural enzyme has some serious limitations in production and use, of which some have been overcome upon fusion to a solubilizing protein tag (MBP). Improvement in protein stability is still desirable, however requires an adequate method to assay protein activity.

We have tested a range of analytical methods to monitor D-glucuronyl C5-epimerase activity. Although not every method was equally successful, eventually we were able to show activity with both HPLC-PAD and the tritium labeled substrate method. However, for further protein optimization by means of directed evolution experiments, a medium to high-throughput method is indispensable. Potential assays could be based on heparinase I in combination of 2-OST, or on an antibody against L-iduronic acid (e.g. HS4E4). A biosensor sensing L-iduronic acid also has great potential, but its development is expected to be an extensive and laborious project.

The lack of structural knowledge of the heparan sulfate D-glucuronyl C5-epimerase, makes it hard to predict via what mechanism the catalysis proceeds, let alone to predict mutations to improve protein characteristics. However, based on sequence analysis and secondary structure predictions we were able to identify different domains, of which the C-terminus is most conserved. This domain is likely to consist of seven α-helices that is predicted to adopt an (incomplete) alpha/alpha toroid fold, a common fold in glycosyl hydrolases, lyases and the N-acetylglucosamine 2-epimerase. Conserved residues that might be involved in substrate binding and catalysis have been indicated. All this adds to the steady increase in knowledge on this key enzyme in heparin biosynthesis and brings us closer to the understanding and integration of this enzyme in novel heparin production methods.

Acknowledgements

We would like to thank Elly Versteeg and Toin van Kuppevelt for their help with the HS4E4 antibody tests, Henk Schols and Willy van den Berg for their help in initiating the HPLC-PAD analysis, Jin-Ping Li for her help with the tritium labeled substrate assays and useful discussions, and Michel Eppink for useful discussions on structural modelling.



Supplemantary Figure 1 – Helical wheel diagrams of the α-helices (murine heparan sulfate D-glucurnonyl C5-epimerase) with conserved amino acid residues putatively involved in heparin binding (grey shading) or residues that could be involved in catalysis (black shading).

References

Bae J, Desai UR, Pervin A, Caldwell EE, Weiler JM, Linhardt RJ. (1998). Interaction of heparin with synthetic antithrombin III peptide analogues. *Biochemical Journal* **301**, 121-129.

Bitter T, Muir HM. (1962). A modified uronic acid carbazole reaction. *Analytical Biochemistry* **4**, 330-334.

Blumenkrantz N, Asboe-Hansen G. (1973). New method for quantitative determination of uronic acids. *Analytical Biochemistry* 54(2), 484-489.

Boersma YL, Dröge MJ, Quax WJ. (2007). Selection strategies for improved biocatalysts. *The FEBS Journal* **274**(9), 2181-2195.

Cadwell RC, Joyce GF. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods and Applications* **2**(1), 28-33.

Campbell P, Feingold DS, Jensen JW, Malmström A, Rodén L. (1983). New assay for uronosyl 5-epimerases. *Analytical Biochemistry* **131**(1), 146-152.

Capila I, Linhardt RJ. (2002). Heparin-protein interactions. *Angewandte Chemie (International ed. in English)* **41**(3), 391-412.

Cardin AD, Weintraub HJ. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**(1), 21-32.

Cesaretti M, Luppi E, Maccari F, Volpi N. (2003). A 96-well assay for uronic acid carbazole reaction. *Carbohydrate Polymers* **54**(1), 59-61.

Charnock SJ, Brown IE, Turkenburg JP, Black GW, Davies GJ. (2002). Convergent evolution sheds light on the anti-beta-elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proceedings of the National Academy of Sciences of the United States of America* **99**(19), 12067-12072.

Chivian D, Kim DE, Malmström L, Bradley P, Robertson T, Murphy P, Strauss CE, Bonneau R, Rohl CA, Baker D. (2003). Automated prediction of CASP-5 structures using the Robetta server. *Proteins* **53**(6), 524-533.

Conrad HE. (1993). Dissection of heparin – past and future. *Pure and Applied Chemistry* **65**(4), 787-791.

Dennissen MA, Jenniskens GJ, Pieffers M, Versteeg EM, Petitou M, Veerkamp JH, van Kuppevelt TH. (2002). Large, tissue-regulated domain diversity of heparan sulfates demonstrated by phage display antibodies. *The Journal of Biological Chemistry* **277**(13), 10982-10986.

Desai UR, Wang HM, Linhardt RJ. (1993). Substrate specificity of the heparin lyases from *Flavobacterium heparinum*. *Archives of Biochemistry and Biophysics* **306**(2), 461-468.

Dische Z. (1946). A new specific color reaction of hexuronic acids. *The Journal of Biological Chemistry* **167**(1), 189-198.

Filisetti-Cozzi TM, Carpita NC. (1991). Measurement of uronic acids without interference from neutral sugars. *Analytical Biochemistry* **197**(1), 157-162.

Guntas G, Mansell TJ, Kim JR, Ostermeier M. (2005). Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proceedings of the National Academy of Sciences of*

the United States of America **102**(32), 11224-11229.

Hileman RE, Fromm JR, Weiler JM, Linhardt RJ. (1998). Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays* **20**(2), 156-167.

Hodson N, Griffiths G, Cook N, Pourhossein M, Gottfridson E, Lind T, Lidholt K, Roberts IS. (2000). Identification that KfiA, a protein essential for the biosynthesis of the *Escherichia coli* K5 capsular polysaccharide, is an alpha -UDP-GlcNAc glycosyltransferase: the formation of a membrane-assosiated K5 biosynthetic complex requires KfiA, KfiB and KfiC. *The Journal of Biological Chemistry* **275**(35), 27311-27315.

Hoffman P, Linker A, Meyer K. (1956). Uronic acid of chondroitin sulfate B *Science* **124**(3234), 1252.

Imanari T, Toida T, Koshiishi I, Toyoda H. (1996). High-performance liquid chromatographic analysis of glycosaminoglycan-derived oligosaccharides. *Journal of Chromatography A.* **720**(1-2), 275-293.

Jerga A, Stanley MD, Tipton PA. (2006). Chemical mechanism and specificity of the C5-mannuronan epimerase reaction. *Biochemistry* **45**(30), 9138-9144.

Jing W, DeAngelis PL. (2003). Analysis of the two active sites of the hyaluronan synthase and the chondroitin synthase of *Pasteurella multocida*. *Glycobiology* **13**(10), 661-671.

Kapust RB, Waugh DS. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**(8), 1668–1674.

Kelley LA, Sternberg MJE. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* **4**, 363 – 371.

Kim DE, Chivian D, Baker D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **32**(2), W526-531.

Kurup S, Wijnhoven TJ, Jenniskens GJ, Kimata K, Habuchi H, Li JP, Lindahl U, van Kuppevelt TH, Spillmann D. (2007). Characterization of anti-heparan sulfate phage display antibodies AO4B08 and HS4E4. *The Journal of Biological Chemistry* **282**(29), 21032-21042.

Kusche M, Hannesson HH, Lindahl U. (1991). Biosynthesis of heparin. Use of *Escherichia coli* K5 capsular polysaccharide as a model substrate in enzymic polymer-modification reactions. *Biochemical Journal* **275**, 151-158.

Li K, Bethea HN, Liu J. (2010). Using engineered 2-O-sulfotransferase to determine the activity of heparan sulfate C5-epimerase and its mutants. *The Journal of Biological Chemistry* **285**, 11106-11113.

Machielsen R, Looger LL, Raedts J, Dijkhuizen S, Hummel W, Hennemann HG, Daussmann T, van der Oost J. (2009). Cofactor engineering of *Lactobacillus brevis* alcohol dehydrogenase by computational design. *Engineering in Life Sciences* **9**(1), 38–44,

Pacheco B, Maccarana M, Goodlett DR, Malmström A, Malmström L. (2008). Identification of the active site of DS-epimerase 1 and requirement of N-glycosylation for enzyme function. *The Journal of Biological Chemistry* **284**(3), 1741-1747.

Petit C, Rigg GP, Pazzani C, Smith A, Sieberth V, Stevens M, Boulnois G, Jann K, Roberts IS. (1995).

Region 2 of the *Escherichia coli* K5 capsule gene cluster encoding proteins for the biosynthesis of the K5 polysaccharide. *Molecular Microbiology* **17**(4), 611-620.

Pinhal MAS, Smith B, Olson S, Aikawa J, Kimata K, Esko JD. (2001). Enzyme interactions in heparan sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **98** (23), 12984-12989.

Radhakrishnamurthy B, Berenson GS. (1963). Effect of temperature and time of heating on the carbazole reaction of uronic acids and acid mucopolysaccharides. *Analytical Chemistry.* **35**(9), 1316-1318.

Roy A, Kucukural A, Zhang Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**(4), 725-738.

Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvåg H, Holtan S, Aachmann FL, Valla S Dijkstra BW. (2008) Structural and mutational characterization of the catalytic A-module of the Mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii. The Journal of Biological Chemistry* **283**, 23819-23828.

Sasisekharan R, Venkataraman, G. (2000). Heparin and heparan sulfate: biosynthesis, structure and function. *Current Opinion in Chemical Biology* **4**, 626-631.

Shaya D, Tocilj A, Li Y, Myette J, Venkataraman G, Sasisekharan R, Cygler M. (2006). Crystal structure of heparinase II from *Pedobacter heparinus* and its complex with a disaccharide product. *The Journal of Biological Chemistry* **281**(22), 15225-15531.

Shively JE, Conrad HE. (1976). Formation of anhydrosugars in the chemical depolymerization of heparin. *Biochemistry* **15**(18), 3932-3942.

Sieberth V, Rigg GP, Roberts IS, Jann K. (1995). Expression and characterization of UDPGlc dehydrogenase (KfiD), which is encoded in the type-specific region 2 of the *Escherichia coli* K5 capsule genes. *Journal of Bacteriology* **177**(15), 4562-4565.

Sobel M, Soler DF, Kermode JC, Harris RB. (1992). Localization and characterization of a heparin binding domain peptide of human von Willebrand factor. *The Journal of Biological Chemistry* **267**(13), 8857-8862.

Söding J, Biegert A, and Lupas AN. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33**, W244--W248 (Web Server issue).

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modelling by MODELLER. *Proteins* **23**, 318-326.

Stemmer WP. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**(6488), 389-391.

Sugiura N, Baba Y, Kawaguchi Y, Iwatani T, Suzuki K, Kusakabe T, Yamagishi K, Kimata K, Kakuta Y, Watanabe H. (2010). Glucuronyltransferase activity of KfiC from *Escherichia coli* strain K5 requires association of KfiA: KfiC and KfiA are essential enzymes for production of K5 polysaccharide, N-acetylheparosan. *The Journal of Biological Chemistry* **285**(3), 1597-1606.

Van den Hoogen BM, van Weeren PR, Lopes-Cardozo M, van Golde LM, Barneveld A, van de Lest CH. A microtiter plate assay for the determination of uronic acids. *Analytical Biochemistry* **257**(2), 107-111.

Waldo GS. (2003). Genetic screens and directed evolution for protein solubility. *Current Opinion in*

Chemical Biology **7**, 33-38.

Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.

Nederlandse Samenvatting

Samenvatting

Het onderzoek beschreven in dit proefschrift is erop gericht om op biotechnologische wijze glycosaminoglycanen te kunnen produceren. Er is onderzoek gedaan naar een tweetal enzymen, beide betrokken in de biosynthese van twee industrieel relevante glycosaminoglycanen, genaamd hyaluronzuur en heparine.

Hoofdstuk 1 geeft een algemene inleiding in een aantal van de meest voorkomende glycosaminoglycanen. Dit zijn biologisch actieve polysacchariden, bestaande uit repeterende disacchariden van een hexosamine gekoppeld aan hexuronzuur. De focus van dit hoofdstuk ligt met name op de biologische relevantie en de biosynthese van hyaluronzuur, heparine en heparine-achtige glycosaminoglycanen. Een gemeenschappelijk aspect binnen deze groep is dat ze allen zijn opgebouwd uit glucosamine en glucuronzuur/iduronzuur. Er wordt ingegaan op de relevantie van L-iduronzuur en de enzymatische omzetting (oftewel C5-epimerisatie) vanuit D-glucuronzuur. Deze omzetting wordt gekatalyseerd door het enzym D-glucuronyl C5-epimerase. Daarnaast wordt ingegaan op het feit dat heparan sulfaat D-glucuronyl C5-epimerase een essentieel enzym is voor de biotechnologische productie van heparine.

Heparan sulfaat D-glucuronyl C5-epimerases, zoals die *in vivo* voorkomen in mens en dier, blijken ongeschikt te zijn voor gebruik in een industrieel proces. Een potentieel alternatief voor deze eukaryote epimerases zijn D-glucuronyl C5-epimerases uit micro-organismen. **Hoofdstuk 2** geeft een overzicht van mogelijke prokaryote D-glucuronyl C5-epimerases in prokaryoten. Er wordt een overzicht gegeven van micro-organismen waaruit glycosaminoglycanen en/of andere L-iduronzuur bevattende polysacchariden zijn geïsoleerd. De aanwezigheid van deze polysacchariden suggereert namelijk dat *in vivo* C5-epimerase activiteit aanwezig kan zijn. Het zoeken naar potentiële D-glucuronyl C5-epimerases kandidaten is gedaan met behulp van *in silico* screening methoden, waarbij het humane heparan sulfate D-glucuronyl C5-epimerases als query gebruikt is. Afsluitend wordt gespeculeerd en gediscussieerd over de mogelijke biologische implicaties van de gevonden kandidaat C5-epimerases.

In **hoofdstuk 3** zijn vervolgens vijf van deze beschreven prokaryote kandidaat C5-epimerases geselecteerd en heterologisch tot expressie gebracht in *Escherichia coli* BL21(DE3). Alle vijf kandidaat C5-epimerases zijn succesvol tot expressie gebracht, waarbij een aantal bovendien voldoende zuiver verkregen zijn om kristallisatie-experimenten mee te

initiëren. *In vitro* omzetting van L-iduronzuur naar D-glucuronzuur is bevestigd voor de kandidaat C5-epimerase uit de bacterie *Bermanella marisrubri* sp. Red65, hetgeen het eerste experimentele bewijs is van D-glucuronyl C5-epimerase activiteit in een micro-organisme.

Als een alternatieve aanpak, is er tevens getracht de muizen heparan sulfaat D-glucuronyl C5-epimerase te optimaliseren voor toepassingen in biotechnologische processen. In **hoofdstuk 4** staat beschreven hoe twee verschillende microbiële expressiesystemen zijn getest voor de heterologe productie van de muizen heparan sulfaat D-glucuronyl C5-epimerase, namelijk in *Pichia pastoris* en *Escherichia coli*. Het gebruik van de gist *P. pastoris* heeft geen succesvolle enzymproductie opgeleverd. Daarentegen resulteerde genexpressie in de bacterie *E. coli* wel in productie van grote hoeveelheden enzym. Het product bleek echter onoplosbaar en onbruikbaar te zijn. De oplosbaarheid van deze C5-epimerase is verbeterd door het enzym N-terminaal te fuseren met het "maltose binding protein" (MBP). Truncaties van de N-terminus zorgen voor een verdere toename van de oplosbaarheid. Deze N-terminale truncaties resulteren echter in verlies van de activiteit, ondanks dat de N-terminus van dit enzym slecht geconserveerd is. Significante activiteit is wel aangetoond bij het fusie-eiwit waar enkel het signaalpeptide van de C5-epimerase N-terminus getrunceerd is. Dit bevestigd dat functioneel muizen heparan sulfaat D-glucuronyl C5-epimerase kan worden geproduceerd via heterologe genexpressie in *E. coli.*

Net als de heparan sulfaat D-glucuronyl C5-epimerase, geeft ook het tweede enzym waaraan we gewerkt hebben oplosbaarheid problemen wanneer het geproduceerd wordt in *E. coli*. **Hoofdstuk 5** beschrijft een vergelijkingsstudie voor de *Pasteurella multosida* hyaluronan synthase (PmHAS) productie in *E. coli*. PmHAS is in andere studies al succesvol tot expressie gebracht. Er bestaan echter nog limitaties voor optimale enzym productie, zuivering en stabiliteit. Derhalve worden twee verschillende PmHAS fusie-eiwitten vergeleken, ter verhelpen van deze bestaande limitaties. Er is aangetoond dat de PmHAS productie aanzienlijk verhoogd kan worden door een N-terminale fusie met "maltose binding protein", zonder dat dit een negatief effect geeft op de specifieke activiteit van het enzym.

In **hoofdstuk 6** wordt een voorbeeld van enzym optimalisatie gegeven, door in plaats van het screenen van een grote mutantenbank, een zogenaamde "smart library design" te gebruiken. Hiervoor is een beperkte mutantenbank gemaakt van het *Pyrococcus furiosus* phosphoglucose isomerase (PfPGI). Mutaties zijn gebaseerd op voorspellingen, welke

gemaakt zijn met Comulator, een bio-informatica methode om gecorreleerde aminozuurresiduen te identificeren. De gecreëerde mutantenbank bevat PfPGI mutanten met aminozuur variaties voor het sterk gecorreleerde aminozuur paar Pro132 en Tyr133. Gebruikmakende van enzymkinetiek metingen is aangetoond dat voor deze mutanten een correlatie bestaat tussen de aanwezigheid van een specifiek animozuur paar, en verhoogde phosphoglucose isomerase activiteit. We hypothetiseren dat deze verandering in activiteit het gevolg zijn van subtiele veranderingen in de eiwitstructuur, met als gevolg veranderingen in de metaal-coördinatie in het PfPGI. Om meer inzicht te krijgen in de exacte veranderingen zijn er eiwitkristallisaties opgezet. Meerdere verkregen eiwitkristallen worden momenteel uitgewerkt en onderzocht.

Afsluitend worden alle eerdere hoofdstukken samengevat en bediscussieerd in **hoofdstuk 7.** Hierbij worden drie bottlenecks aangeven die opgelost dienen te worden, voordat een succesvolle implementatie van deze enzymen in een industrieel proces kan worden bewerkstelligd. Naast de behaalde resultaten worden additionele wegen besproken om de genoemde enzymen verder te optimaliseren. Bovendien worden verscheidene geteste assay methoden, geschikt om de epimerisatie van D-glucuronzuur aan te tonen, bediscussieerd om hun bruikbaarheid. Ook wordt er gespeculeerd over welke aminozuur residuen mogelijk een cruciale rol hebben in de D-glucuronyl C5-epimerase en worden de mogelijkheden binnen eiwitstructuur / vouwing besproken.

Acknowledgements

Acknowledgements

Na al die tijd promotieonderzoek kan ik dan nu eindelijk mijn proefschrift presenteren! Hoewel *onderzoek naar heparine niet bloedstollend zou moeten zijn*, was het zeker wel een spannend avontuur! Dankbaar ben ik dan ook allen die op eniger wijze een bijdrage hebben geleverd aan het completeren van dit proefschrift.

Graag begin ik met het bedanken van mijn promotor John van der Oost. Niet alleen omdat je mij de kans hebt gegeven om een promotieonderzoek te verrichten binnen jouw groep, maar ook voor gegeven vertrouwen, steun, inspiratie en adviezen. John het was (en is) een waar genoegen om werkzaam te zijn binnen jouw groep.

Daarnaast wil ik graag mijn copromotor Servé Kengen bedanken. Servé geweldig dat je altijd bereid was (en bent) om mee te denken waar mogelijk, of een luisterend oor te bieden waar nodig. Tevens kijk ik er naar uit om de komende tijd onze samenwerking voort te zetten, ditmaal echter op het gebied van *Clostridium acetobutylicum*.

Dank gaat ook uit naar Willem de Vos, die keer op keer een sterk staaltje (time) management tentoonspreidde; het deelnemen aan het dagelijks bestuur was zowel leerzaam als inspirerend. Bovendien ben ik hem dankbaar voor zijn kritische vragen en opmerkingen tijdens de tweewekelijkse PhD-meeting.

Uiteraard ook de nodige dank voor mijn oud-kamergenoten. Marcel, ik kende je natuurlijk al van onze avon(d/t)uren tijdens onze studietijd; geweldig dat er de nodige avonturen bij zijn gekomen in de tijd bij Micro. En niet te vergeten de geweldige lunches die we hadden op Micro! Mark, ook jou ken ik natuurlijk al van voor mijn AIO-tijd; je bereidheid tot participatie in menig sociale activiteit is fantastisch. Ik vind het erg leuk dat we nog wat langer collega's zijn! Marco(oooo) heel erg bedankt voor het smakelijk lachen om mijn grappen, ik heb ook erg gelachen om die grappen van jou (mits ik het verstond uiteraard). De volgende keer als ik in Eibergen ben, rijd ik zeker door naar *Hoakseberge*. Ook wil ik jullie alle drie bedanken dat jullie willen optreden als mijn paranimfen en fotograaf. Laten we binnenkort maar weer eens afspreken om aan het brouwen te gaan?!

Natuurlijk ook Bram, Suus, Tom & Elleke bedankt voor de leuke sfeer op onze kamer! Ook Bart, Chris, Vincent, Mark M en Katrin bedankt voor jullie gezelligheid. Of het binnen of

buiten Micro was, vermakelijk was het altijd! Laten we de traditie aanhouden; op naar de volgende activiteit! Farrakh thx a lot for all the nice conversations, the spicy-ricy and fun!

Onvergetelijk waren beide studiereizen naar de Verenigde Staten. The organization of the 2009 Boston-Washington PhD-trip was legendary, thx to my fellow members of the organisation. Marcel, Mattie, Thomas and Faab, I truly enjoyed the cooperation.

Another big thank you is for the three students who helped me out in various stages of my thesis. Abhi, Sjoerd and Jasper, you all worked really dedicated and it was a pleasure to supervise all of you. Thx for all your contributions to my thesis, and I wish you guys all the best in your future careers.

Wim bedankt voor je hulp tijdens al die keren dat ik weer op zoek was naar technische ondersteuning (waarvan de helft was wegens de Dionex?!). Willie bedankt dat je me wegwijs hebt gemaakt met de HPLC-PAD. Sjon niet alleen bedankt voor al je technische ondersteuning, maar ook erg bedankt voor je inzet op het PGI project. Jouw bijdrage aan hoofdstuk zes getuigd daarvan. Ook Henk-Jan, Peter en Jasper bedankt voor jullie bijdrages aan dit hoofdstuk. Daarnaast wil ik ook graag al die andere vaste krachten van Microbiologie bedanken die meerdere keren voor me klaarstonden, in het bijzonder Ans, Nees, Anja, Francis, Renee, Jannie, Hans, Ton en Philippe bedankt! To all others of BacGen (Stan, Wouter, Nicolas, Pierpaolo, Rie, Edze, Pawel, Ratnesh, Harmen, Jasper A&W, Hao, Marke, Colin, Melvin, Teunke, Biniam, Daan, Tessa, Sebastiaan, Ruud, Amos, Raymond, Slavtscho), MicFys, MolEco and SSB: thank you for your help, suggestions and nice discussions.

Ook wil ik graag de “coworkers” bedanken binnen B-Basic project 3.5 (& 3.4) voor hun bijdrages binnen de projectmeetings. Gerrit Eggink, Jan Springer, Emil Wolbert, Floor Kooy, Anaïs Chavaroche, Michel Eppink, Gijs van Dedem, Aukje Zimmerman, Jan Willem Simons, Marenne Welten, Vincent van Gemert, Herman Overkleeft en Hendrik Gold, het was me een genoegen. Jan, Emil en Anaïs jullie ook nog eens erg bedankt voor jullie technische hulp toen ik op A&F wat experimenten deed.

Jin-ping Li thanks for your help on the D-glucuronyl C5-epimerase activity measurement. Also thanks to Magnus. Henk Schols bedankt voor je hulp met de HPLC-PAD methode. Peter Nobels bedankt voor je hulp met de ICP-AES. Pat Baker thank you for your help on PfPGI, Mark Deller thanks for all the efforts on ‘Red65’.

Uiteraard was er ook altijd de welkome afleidingen buiten het lab waarbij er altijd voldoende vrienden zijn om op terug te vallen. Woest & Liesbeth, GJ & Celia, Anja, Broodje, Peter bedankt voor al die gezelligheid en de nodige smakelijke etentjes. Janus, Henk en andere, de jaarlijkse skitripjes zijn altijd erg *verfrissend* geweest, ik kijk alweer uit naar het volgende avontuur in de sneeuw.

Vanzelfsprekend ben ik mijn ouders dankbaar voor hun bijdrage. Zonder jullie vertrouwen en jullie morele & financiële steun zou ik hier nu niet gestaan hebben. Ik ben erg blij deze heugelijke gebeurtenis samen met jullie mee te mogen maken. Gerard bedankt voor je getoonde interesse en de gezellige etentjes in Wageningen. Ook wil ik de rest van de (schoon)familie bedanken voor al die getoonde interesse. Ik realiseer me wel degelijk dat niet altijd even makkelijk te begrijpen zal zijn geweest waar ik mee bezig was.

Tenslotte wil ik graag mijn dank en liefde uitspreken voor Eline. Je was er vanaf de start van mijn promotietraject, en hebt me vervolgens de gehele weg bijgestaan. Ik ben dan ook heel erg blij dat naast deze promotie, ook onze relatie binnenkort een promotie zal gaan maken en ik ga dan ook met het volste ver*trouwen* en met heel erg veel enthousiasme *onze* toekomst tegemoet.

Curriculum Vitae

Jozef Gerardus Johannes (John) Raedts was born on a very cold winter night on the 12th of the 12th 1981 in Evertsoord, a very small town in the south of the Netherlands. After graduating in 2000 for his atheneum diploma at the Dendron College in Horst, he started studying Biotechnology at Wageningen University. He did his minor thesis at the Laboratory of Microbiology (Wageningen University) at the group of Prof. dr. John van der Oost performing directed evolution on an alcohol dehydrogenase from *Lactobacillus brevis.* His major thesis he did at the Laboratory of Molecular Biology (Wageningen University) with Prof. dr. Ton Bisselink where he was involved in the analysis of genes involved in root nodule formation pathway in the legume *Medicago truncatula*. He also was employed as student assistant at the same group for some months. He concluded his study with an internship at the Schmid-Ott Lab at the University of Chicago (Illinois, United States of America) involved in the construction of a BAC library to study early embryonic gene expression in the hoverfly *Episyrphus balteatus*. After obtaining his Master degree in March 2006, he started his PhD at the Laboratory of Microbiology April 2006 under supervision of Prof. dr. John van der Oost and Dr. Servé Kengen. His research was focused on the characterization and optimization of glycosaminoglycan acting enzymes. The results are presented in this thesis. Currently he is working as a postdoc at the Laboratory of Microbiology where he is involved in the setup of a metabolomics approach for the solvent producing bacterium *Clostridium acetobutylicum*.

List of publications

Smit P, Raedts J, Portyanko V, Debellé F, Gough C, Bisseling T, Geurts R. (2005). NSP1 of the GRAS protein family is essential for rhizobial nod factor-induced transcription. *Science* **308**(5729), 1789-1791.

Machielsen R, Looger LL, Raedts J, Dijkhuizen S, Hummel W, Henneman HG, Daussmann T, van der Oost J. (2009). Cofactor engineering of *Lactobacillus brevis* alcohol dehydrogenase by computational design. *Engineering in Life Sciences* **9**(1), 38-44.

Raedts J, Kengen SWM, van der Oost J. (2011). Occurrence of L-iduronic acid and putative D-glucuronyl C5-epimerases in prokaryotes. *Glycoconjugate Journal* **28**(2), 57-66.

Raedts J, Li JP, Lundgren M, Kengen SWM, van der Oost J. Evidence for D-glucuronyl C5-epimerase activity in the bacterium *Bermanella marisrubri* sp. RED65. *Manuscript in preparation.*

Raedts J, Creutzburg S, Li JP, Lundgren M, Kengen SWM, van der Oost J. Microbial production of the Murine heparan sulfate D-glucuronyl C5-epimerase. *Manuscript in preparation.*

Raedts J, Sloothaak J, Nag A, Kengen SWM, van der Oost J. Improved production of *Pasteurella multocida* hyaluronan synthase in *Escherichia coli*. *Manuscript in preparation.*

Raedts J, Feras Almourfi, Hendriks S, Joosten HJ, Schaap P, Kengen SWM, Baker P, van der Oost J. Correlated mutation analysis as a tool for smart library design to improve *Pyrococcus furiosus* phosphoglucose isomerase activity. *Manuscript in preparation.*

Overview of completed training activities

Discipline specific activities

Courses:

- Clinic Protein design using Cameleon software (Wageningen) 2007
- Course polysaccharides as food colloids and biomaterials (VLAG, Wageningen) 2007
- Summer Course Glycosciences (VLAG, Wageningen) 2008

Meetings:

- CW / NWO Protein Research (Lunteren) *-poster presentation-* 2006
- CW / NWO Protein Research (Lunteren) 2007
- 2nd B-Basic Symposium (Nunspeet) *-poster presentation-* 2007
- 3rd International symposium on pectins and pectinases (Wageningen) 2008
- 3rd B-Basic Symposium (Nunspeet) *-poster presentation-* 2008
- 12th Nederlands Biotechnologie Congress (Ede) *-poster presentation-* 2008
- CW / NWO Protein Research (Velthoven) *-poster presentation-* 2008
- 8th Carbohydrate Bioengineering meeting (Ischia, Italy) *-poster presentation-* 2009
- 4th B-Basic Symposium (Rotterdam) *-poster presentation-* 2009
- 13th Nederlands Biotechnologie Congress (Ede) *-poster presentation-* 2010

General courses

- PhD introduction week (VLAG, Bilthoven) 2007
- Course Career Perspectives (WGS, Wageningen) 2009

Optionals

- Preparing PhD research proposal 2006
- Weekly Bacterial Genetics Group Meetings (Laboratory of Microbiology) 2006-2010
- Biweekly PhD/Postdoc meetings (Laboratory of Microbiology) 2006-2010
- Project Meetings (Wageningen/Oss) *-oral presentations-* 2006-2010
- Experimental evolution discussion group (WUR) *-oral presentations-* 2006-2008
- Journal club Microbiology (Laboratory of Microbiology) 2008
- PhD Study trip to California (USA) *-oral & poster presentations-* 2006
- Organization of PhD study trip to Boston/Washington 2008-2009
- PhD Study trip to Boston/Washington (USA) *-oral & poster presentations-* 2009
- 1 week stay Uppsala Biomedicinska Centrum (Uppsala universitet, Sweden) 2010

The research described in this thesis was performed at the Laboratory of Microbiology, Wageningen University, The Netherlands. This project was carried out in close collaboration with MSD Oss (formally Organon N.V.) and financially supported by the Netherlands Ministry of Economic Affairs and the B-Basic partner organizations (www.b-basic.nl) through B-Basic, a public–private NWO-ACTS program (Advanced Chemical Technologies for Sustainability).

Cover design by John Raedts

Cover image: Heparin, from isolation towards biotechnological production

Printed by GVO drukkers & vormgevers B.V. / Ponsen & Looijen, Ede