Glycoconj J (2011) 28:57–66
DOI 10.1007/s10719-011-9324-7

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

John Raedts · Servé W. M. Kengen · John van der Oost

Received: 13 December 2010 / Revised: 14 January 2011 / Accepted: 18 January 2011 / Published online: 24 February 2011

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.

Keywords Glycosaminoglycans · L-iduronic acid · D-glucuronyl C5-epimerase · Lipopolysaccharide · Capsule polysaccharide

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. An overview is given in Table 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 [1, 2]. 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 [3].

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. In the past decades the C5-epimerase in animal heparin biosynthesis has been identified and characterized [4], and has some major limitations to use in large scale biotechnological production processes. Novel D-glucuronyl C5-epimerases that have less restric-

J. Raedts (✉) · S. W. M. Kengen · J. van der Oost
Laboratory of Microbiology, Wageningen University,
Dreijenplein 10,
6703 HB Wageningen, The Netherlands
e-mail: John.Raedts@wur.nl

Table 1 Overview of the most common glycosaminoglycans. GlcA, glucuronic acid; GalNAc, *N*-acetyl-galactosamine; GlcNAc, *N*-acetyl-glucosamine; IdoA, iduronic acid; GlcNS, *N*-sulfate-glucosamine; Gal, galactose; C#, carbon number; ECM, extracellular matrix

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

tions 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 [5]. 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 [6]. 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 [7]. 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 [8, 9], 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 [10]. 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 [11].

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 [12]. 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 hyaluronian 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 [13]. 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 [14].

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 [15]. 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) [16]. 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* [17], however strain KMM 223 remains the only example that has an L-iduronic acid-containing O-antigen.

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 [18, 19]. Many more O-antigen 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* [20]. 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 [21]. Although archaeal flagellins often undergo posttranslational modification [22], 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.

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 ↑ OSO_3^-; 3 ↑ OSO_3^-; 3 ↑ OSO_3^-; X; Thr / Ser; 1/3 of GlcA residues can be replaced by **IdoA**	Wieland *et al.*, 1985

Fig. 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

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 [23]. 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 [24].

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. [25]. 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 [26].

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 [27].

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) [28]. Multiple prokaryotic sequences were identified that have homology to the human 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 [29] 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 2.

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. 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 (Fig. 2). 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 [3] and the prokaryotic candidate epimerases, reveal a role of conserved histidines and tyrosines in catalysis for both functionally distinct C5-epimerases [30, 31]. 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 [32] containing several eukaryotic D-glucuronyl C5-epimerases, as well as a selection of prokaryotic homologs. The multiple sequence alignment [33, 34] is mostly based on the C-terminus of the genes (Fig. 2). 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. 3). 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.

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. 4). Genes in this gene cluster are likely to be involved in strain specific O-antigen production [35], since some of the sequences have homology to the *wbb* operon [36], which is known to be involved in the O-antigen biosynthesis [37]. 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* [38].

A remarkable similarity in genomic organization is seen for some of the putative C5-epimerases. Figure 4 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* [39].

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.

Table 2 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

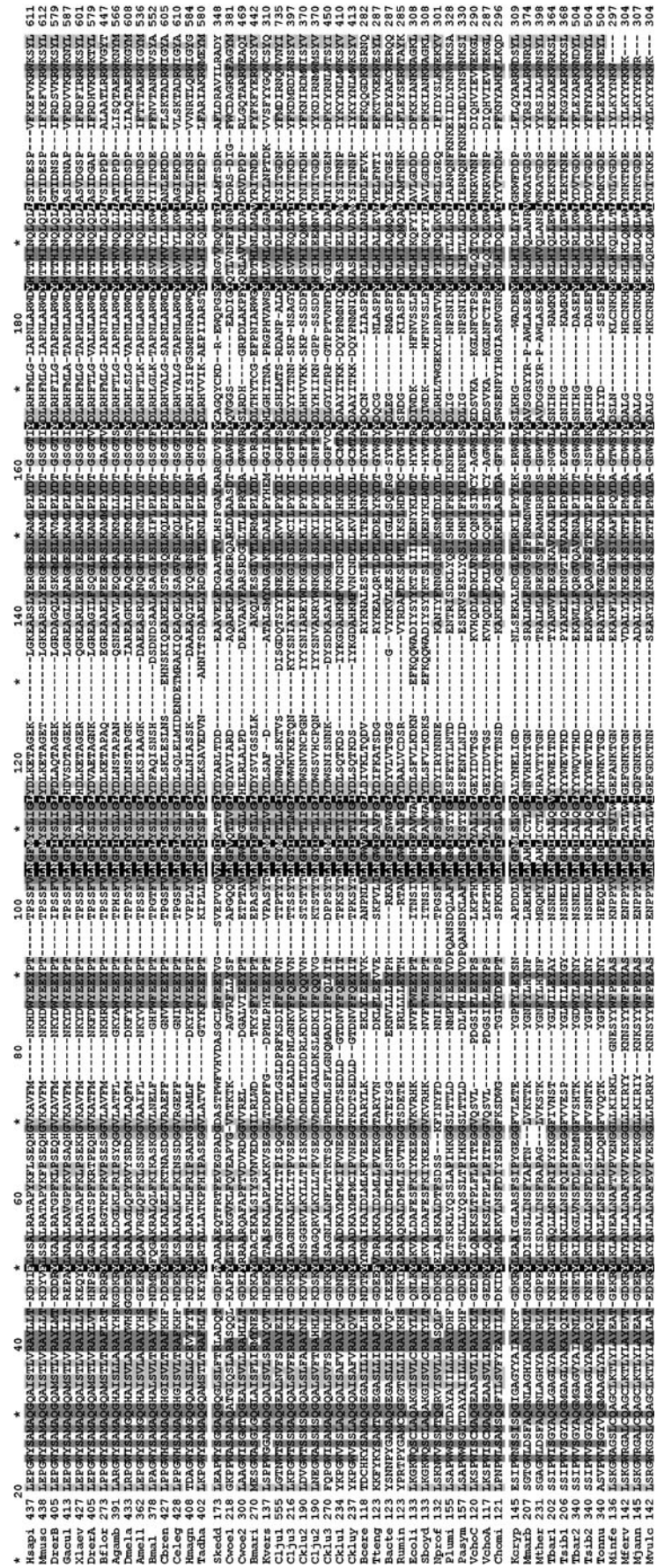

Fig. 2 Multiple sequence alignment of C-terminal part of the candidate D-glucuronyl C5-epimerases. The positions of the first and last residues of the aligned region of the corresponding candidate C5-epimerase are indicated for each sequence. Names are abbreviated and in the same order as in Fig. 3

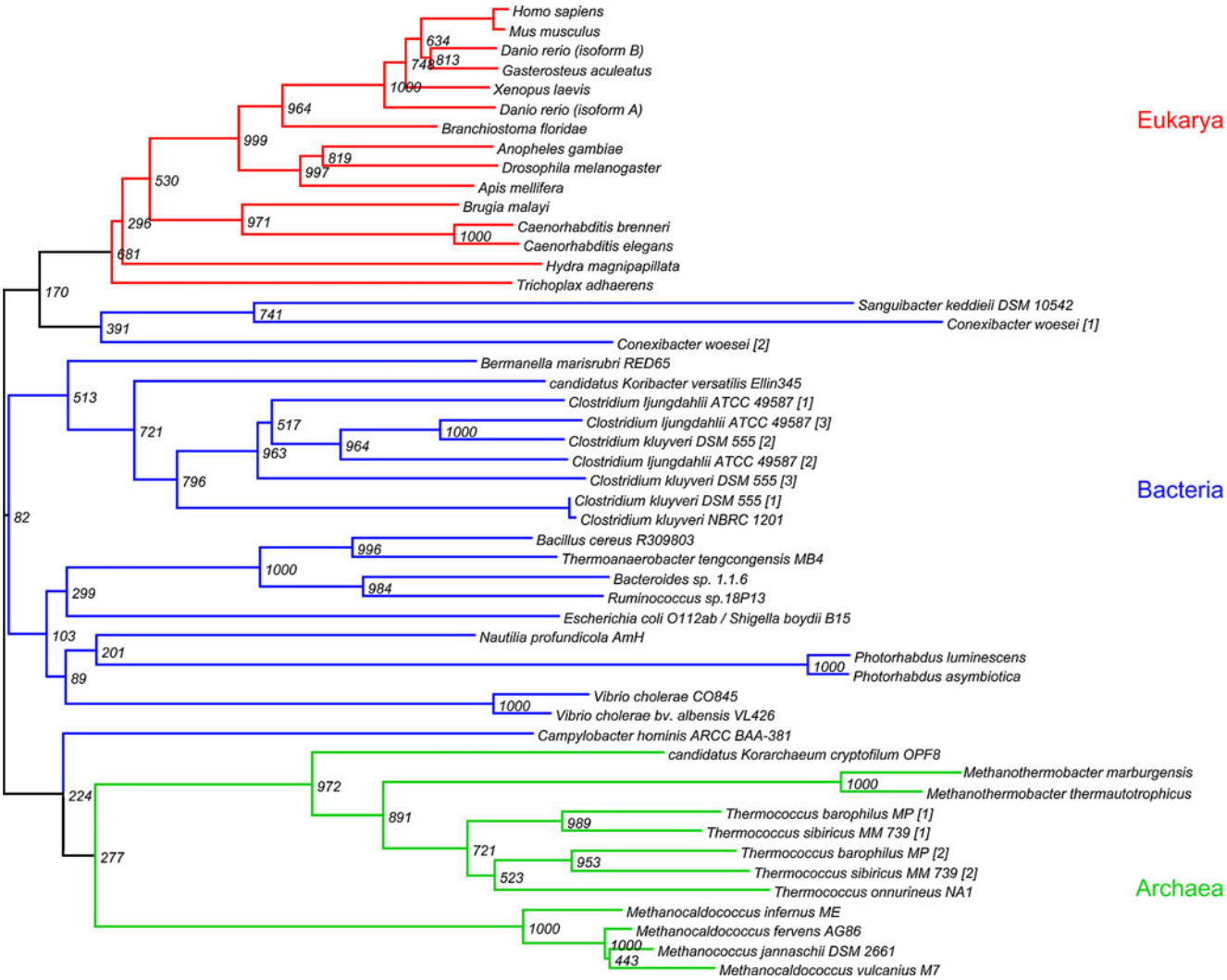

Fig. 3 Phylogenetic analysis candidate C5-epimerases. The coloring of branches is domain specific; eukarya in red, bacteria in blue and archaea in green. The used bootstrap value is 1000

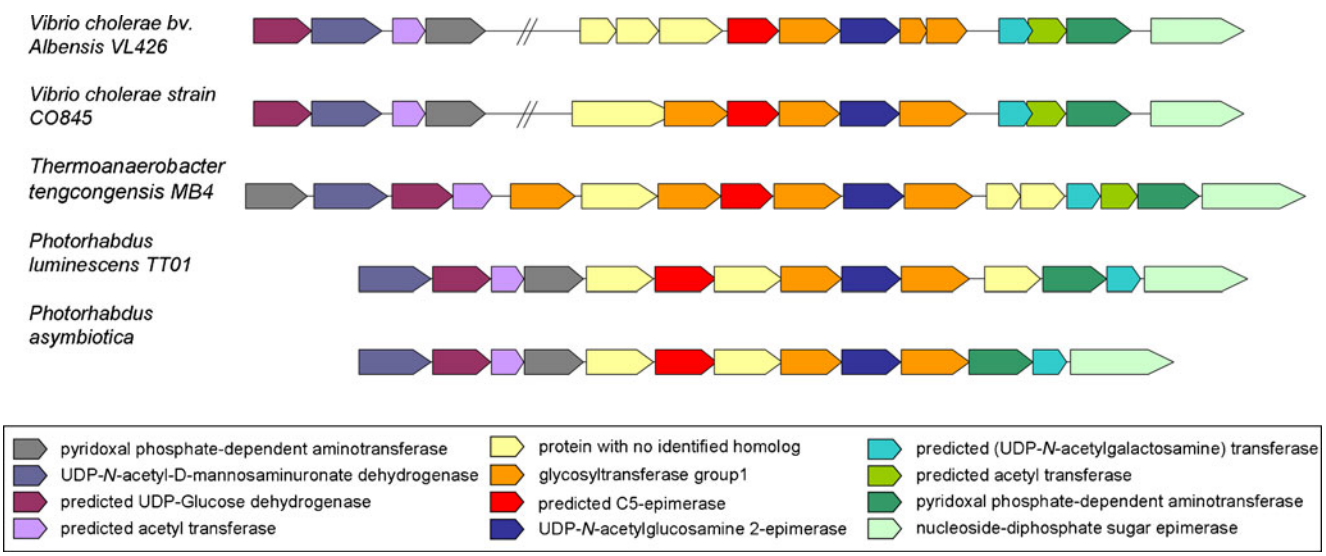

Fig. 4 Neighborhood analysis candidate C5-epimerases

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).

References

1. Lindahl, U., Bäckström, G., Malmström, A., Fransson, L.A.: Biosynthesis of L-iduronic acid in heparin: Epimerization of D-glucuronic acid on the polymer level. Biochem. Biophys. Res. Commun. **46**(2), 985–991 (1972)
2. Höök, M., Lindahl, U., Bäckström, G., Malmström, A., Fransson, L.A.: Biosynthesis of heparin. III. Formation of L-iduronic acid residues. J. Biol. Chem. **249**, 3908–3915 (1974)
3. Valla, S., Li, J., Ertesvåg, H., Barbeyron, T., Lindahl, U.: Hexuronyl C5-epimerases in alginate and glycosaminoglycan biosynthesis. Biochimie **83**(8), 819–830 (2001)
4. Lindahl, U.: 'Heparin'— from anticoagulant drug into the new biology. Glycoconj. J. **17**(7–9), 597–605 (2000)
5. Ørskov, I., Ørskov, F., Jann, B., Jann, K.: Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41**(3), 667–710 (1977)
6. Rodriguez, M., Jann, B., Jann, K.: Structure and serological characteristics of the capsular K4 antigen of *Escherichia coli* O5: K4: H4, a fructose-containing polysaccharide with a chondroitin backbone. Eur. J. Biochem. **177**, 117–124 (1988)
7. Vann, W.F., Schmidt, M.A., Jann, B., Jann, K.: The structure of the capsular polysaccharide (K5 antigen) of urinary-tranct-infective *Escherichia coli* O10:K5:H4. A Polymer Similar to Desulfo-Heparin. Eur. J. Biochem. **116**, 359–364 (1981)
8. Kusche, M., Hannesson, H.H., Lindahl, U.: Biosynthesis of heparin. Use of *Escherichia coli* K5 capsular polysaccharide as a model substrate in enzymic polymer-modification reactions. Biochem. J. **275**, 151–158 (1991)
9. Hagner-McWhirter, A., Hannesson, H.H., Campbell, P., Westley, J., Rodén, L., Lindahl, U., Li, J.: 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 (2000)
10. Rimler, R.B.: Presumptive identification of *Pasteurella multocida* serogroups A, D and F by capsule depolymerisation with mucopolysaccharidases. Vet. Rec. **134**, 191–192 (1994)
11. DeAngelis, P.L., Gunay, N.S., Toida, T., Mao, W., Linhardt, R.J.: Identification of the capsular polysaccharides of Type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively. Carbohydr. Res. **337**(17), 1547–1552 (2002)
12. Rosner, H., Grimmecke, H.D., Knirel, Y.A., Shashkov, A.S.: Hyaluronic acid and a (1→4)-β-D-xylan, extracellular polysaccharides of *Pasteurella multocida* (Carter type A) strain 880. Carbohydr. Res. **223**, 329–333 (1992)
13. Darby, G.K., Jones, A.S., Kennedy, J.F., Walker, R.T.: Isolation and analysis of the nucleic acids and polysaccharides from *Clostridium welchii*. J. Bacteriol. **103**(1), 159–165 (1970)
14. Lee, L., Cherniak, R.: Identification of iduronic acid as a constituent of the "type-specific" polysaccharide of *Clostridium perfringens* Hobbs 10. Carbohydr. Res. **33**, 387–390 (1974)

15. Stack, R.J., Plattner, R.D., Cote, G.L.: Identification of L-iduronic acid as a constituent of the major extracellular polysaccharide produced by *Butyrivibrio fibrisolvens* strain X6C61. FEMS Microbiol. Lett. **51**, 1–5 (1988)
16. Hanniffy, O.M., Shashkov, A.S., Senchenkova, S.N., Tomshich, S.V., Komandrova, N.A., Romanenko, L.A., Knirel, Y.A., Savage, A.V.: Structure of a highly acidic O-specific polysaccharide of lipopolysaccharide of *Pseudoalteromonas haloplanktis* KMM 223 (44-1) containing L-iduronic acid and D-QuiNHb4NHb. Carbohydr. Res. **307**, 291–298 (1998)
17. Nazarenkoa, E.L., Komandrovaa, N.A., Gorshkovaa, R.P., Tomshicha, S.V., Zubkova, V.A., Kilcoyneb, M., Savage, A.V.: Structures of polysaccharides and oligosaccharides of some Gram-negative marine *Proteobacteria*. Carbohydr. Res. **338**(23), 2449–2457 (2003)
18. Perepelov, A.V., Liu, B., Senchenkova, S.N., Shashkov, A.S., Feng, L., Knirel, Y.A., Wang, L.: Structure of the O-polysaccharide of *Escherichia coli* O112ab containing L-iduronic acid. Carbohydr. Res. **343**(3), 571–575 (2008)
19. Liu, B., Knirel, Y.A., Feng, L., Perepelov, A.V., Senchenkova, S. N., Wang, Q., Reeves, P.R., Wang, L.: Structure and genetics of Shigella O antigens. FEMS Microbiol. Rev. **32**, 627–653 (2008)
20. Wieland, F., Lechner, J., Sumper, M.: Iduronic acid: constituent of sulphated dolichyl phosphate oligosaccharides in halobacteria. FEBS Lett. **195**, 77–81 (1986)
21. Wieland, F., Paul, G., Sumper, M.: Halobacterial flagellins are sulfated glycoproteins. J. Biol. Chem. **260**(28), 15180–15185 (1985)
22. Jarrell, K.F., Bayley, D.P., Kostyukova, A.S.: The archaeal flagellum: a unique motility structure. J. Bacteriol. **178**(17), 5057–5064 (1996)
23. Ozaki, N., Sakuda, S., Nagasawa, H.: A novel highly acidic polysaccharide with inhibitory activity on calcification from the calcified scale "coccolith" of a coccolithophorid alga, *Pleurochrysis haptonemofera*. Biochem. Biophys. Res. Commun. **357**, 1172–1176 (2007)
24. Lahaye, M., Robic, A.: Structure and functional properties of ulvan, a polysaccharide from green seaweeds. Biomacromolecules **8**, 1765–1774 (2007)
25. Miyasaki, T., Tsuchihashi, H., Yamada, H., Yodomae, T.: Detection and identification of L-iduronic acid in the glycuronan "protuberic acid" from *Koboyasia nipponica*. Carbohydr. Res. **77**, 281–284 (1979)
26. Tsuchihashi, H., Yadomae, T., Miyazaki, T.: Structure of L-idurono-D-glucuronans from the fungi *Aseroe araclmoidea* and *Pseudocolas fusiformis*. Carbohydr. Res. **108**(1), 123–128 (1983)
27. Carroll, A.R., Duffy, S., Avery, V.M.: Citronamides A and B, tetrapeptides from the Australian sponge *Citronia astra*. J. Nat. Prod. **72**, 764–768 (2009)
28. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402 (1997)
29. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., Eddy, S.R.: The Pfam protein families database. Nucleic Acids Res. **32**, D138–D141 (2004)
30. Rozeboom, H.J., Bjerkan, T.M., Kalk, K.H., Ertesvåg, H., Holtan, S., Aachmann, F.L., Valla, S., Dijkstra, B.W.: Structural and mutational characterization of the catalytic A-module of the Mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii*. J. Biol. Chem. **283**, 23819–23828 (2008)
31. Pacheco, B., Maccarana, M., Goodlett, D.R., Malmström, A., Malmström, L.: Identification of the active site of DS-epimerase 1 and requirement of N-glycosylation for enzyme function. J. Biol. Chem. **284**(3), 1741–1747 (2008)
32. Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., Rupp, R.: Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinform. **8**, 460 (2007)
33. Notredame, C., Higgins, D.G., Heringa, J.: T-Coffee: A novel method for multiple sequence alignments. J. Mol. Biol. **302**, 205–217 (2000)
34. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G.: The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**, 4876–4882 (1997)
35. Whitfield, C.: Biosynthesis of lipopolysaccharide O antigens. Trends Microbiol. **3**(5), 178–185 (1995)
36. Derzell, S., Turlin, E., Duchaud, E., Pages, S., Kunst, F., Givaudan, A., Danchin, A.: The PhoP-PhoQ two-component regulatory system of *Photorhabdus luminescens* is essential for virulence in insects. J. Bacteriol. **186**(5), 1270–1279 (2004)
37. Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C.R., Rick, P.D.: Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. **4**(12), 495–503 (1996)
38. Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J.A., Bush, C.A., Morris Jr., J.G., Stine, O.C.: The capsule polysaccharide structure and biogenesis for non-O1 *Vibrio cholerae* NRT36S: genes are embedded in the LPS region. BMC Microbiol. **7**, 20 (2007)
39. 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.: A complete sequence of the *T. tengcongensis* genome. Genome Res. **12**, 689–700 (2002)