

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres



A novel N-linked flagellar glycan from Methanococcus maripaludis

John Kelly a, Susan M. Logan a, Ken F. Jarrell b, David J. VanDyke b, Evgeny Vinogradov a,*

ARTICLE INFO

Article history:
Received 22 November 2008
Received in revised form 6 January 2009
Accepted 7 January 2009
Available online 17 January 2009

Keywords:
Archaea
Methanococcus maripaludis
Flagellar glycosylation
N-Linked tetrasaccharide
Structure
NMR

ABSTRACT

The archaea *Methanococcus maripaludis* strain Mm900 produces flagella that are glycosylated with an N-linked tetrasaccharide. Mass spectrometric analysis of flagellar tryptic peptides identified a number of tryptic glycopeptides carrying a glycan of mass 1036.4 Da, and fragmentation of the glycan oxonium ion indicated that the glycan was a tetrasaccharide. The glycan was purified, following extensive pronase digestion of flagellar filaments, by size-exclusion and anion-exchange chromatography. NMR spectroscopy revealed that the glycan had the following structure: Sug-4- β -ManNAc3NAmA6Thr-4- β -Glc-NAc3NAcA-3- β -GalNAc-Asn where Sug is a novel monosaccharide unit, (5S)-2-acetamido-2,4-dideoxy-5- θ -methyl- θ -L-erythro-hexos-5-ulo-1,5-pyranose. This oligosaccharide has significant similarity to the oligosaccharide that was found previously in *Methanococcus voltae*.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Flagellar proteins have been shown to be glycosylated in a diverse number of species of the domains Bacteria and Archaea. In bacteria, glycosylation of flagellins is exclusively found to be O-linked, while in archaea the glycan is linked to the flagellin protein backbone via an N-linkage through a typical eukaryotic N-linked sequon (Asn-X-Ser/Thr). In archaea, the flagellar system is distinct from that of bacteria, and a body of evidence now indicates that the archaeal flagellar system much more closely resembles bacterial type IV pilus systems in both architecture and assembly.^{2,3} Structural characterisation of flagellin glycans in archaea is thus far limited to only two species, Methanococcus voltae and Halobacterium salinarum, 4-6 while genetic investigation of the archaeal N-linked glycosylation process has been significantly advanced through mutagenesis and functional studies in M. voltae^{7,8} and also in Haloferax volcanii where a pentasaccharide is N-linked to the S-layer protein. 9-13 The flagellin of *H. salinarum* is modified with N-linked sulfated oligosaccharides of glucose in $(1\rightarrow 4)$ -linkage to hexuronic acids, while extensive structural characterisation of flagella from M. voltae revealed that the four flagellin structural proteins such as FlaA, FlaB1, FlaB2 and FlaB3 were all glycosylated with a trisaccharide composed of β-ManNAc3NAcA6Thr-4-β-GlcNAc3NAcA-3-β-GlcNAc-Asn. In addition, the same trisaccharide was identified on the respective S-layer proteins from these strains indicating a common N-linked glycosylation pathway for both processes.

Studies have recently shown that *M. voltae* can also glycosylate both the flagellin and S layer with a tetrasaccharide composed of the afore-mentioned trisaccharide carrying an additional residue of either 220 or 262 Da mass. ¹⁴ A biological role for these glycans is yet to be determined although in *M. voltae* and *Methanococcus maripaludis*, flagellins that are nonglycosylated or have only truncated glycans attached result in either nonflagellated cells or flagellated cells that are less motile than the wildtype. ^{7,15} In addition, it has been recently demonstrated that a mutant of *M. maripaludis* that contains a deletion of an acetyltransferase gene (MMP0350) required for the biosynthesis of a glycan sugar was not only nonflagellated but also defective in attaching pili to the cell surface. ¹⁵

In this work, we have elucidated the full structure of the flagellar glycan from M. maripaludis and presented the first example of a naturally occurring diglycoside of an aldulose as a component of the tetrasaccharide. In addition, the linkage of this novel tetrasaccharide to Asn is shown to occur via a β -GalNAc residue, which is in contrast to the related flagellar glycan from M. voltae that had been previously shown to occur via a β -GlcNAc residue.

2. Results and discussion

2.1. Mass spectrometry analysis of flagellar proteins

Attempts to determine the intact mass of respective flagellin monomers by mass spectrometry (MS) were unsuccessful, presumably due to poor ionisation of the intact proteins from the filaments. This had previously been observed for flagellins from *M. voltae.*⁴ MS analysis was carried out on the tryptic digest of

^a Institute for Biological Sciences, National Research Council, Ottawa, Ontario, Canada K1A 0R6

^b Department of Microbiology and Immunology, Queen's University, Kingston, Ontario, Canada K7L 3N6

^{*} Corresponding author. Tel.: +1 613 990 0832; fax: +1 613 952 9092. E-mail address: evguenii.vinogradov@nrc-cnrc.gc.ca (E. Vinogradov).

flagella purified from M. maripaludis. Tryptic peptides were analysed by nano-LC-MS/MS, and most of the resulting MS/MS spectra could be assigned to tryptic peptides from one of the three flagellar proteins expressed by M. maripaludis, FlaB1, 2 and 3 (Fig. 1). However, a number of spectra could not be easily identified and appeared to be derived from glycopeptides containing the same unusual glycan (See Supplementary Table S1). In some instances, the amino acid sequence of the glycopeptides could be determined from their fragment-ion patterns (Fig. 2). A simple subtraction of the peptide mass from that of the glycopeptide determined the mass of the glycan to be 1036.4 Da. Furthermore, each glycopeptide contains one or more N-linked sequons, suggesting that the glycan is N-linked as was found previously for M. voltae flagellar modifications. 4 Evidence for the N-linkage can be seen in the MS/ MS spectrum of the tryptic glycopeptide, T, 146-166 derived from FlaB1 protein (Fig. 2), where the y-fragment ions that contains the asparagine from the N-linked sequon (y₁₁ and greater) are modified with a HexNAc residue (+203 Da).

Initially, the composition and sequence order of the glycan were difficult to determine as the terminal monosaccharide residue was very labile and did not produce a distinctive oxonium ion in the low-mass region of the glycopeptide MS/MS spectra. In addition, ions that arise from the sequential loss of sugar residues from the precursor ion (neutral loss) or from a charge-reduced precursor ion were not in high abundance in these spectra. Nevertheless, the general structure and composition of the glycan were elucidated due to a distinctive group of oxonium ions that were found in all glycopeptide MS/MS spectra, including the intact glycan ion at m/z 1037.4. These ions are identified in Figure 2.

Using mass spectrometry, we determined that the glycan is a tetramer consisting of a HexNAc residue (203 Da) linked to the asparagine residue on the peptide, followed by a di-*N*-acetyl hexuronic acid (258 Da), a 358 Da residue and capped with a 217 Da residue at its non-reducing end. The 358 and 217 Da sugars were not observed before and were considered novel. Nevertheless, this glycan resembles that found on *M. voltae* flagellin, which is N-linked to the peptide through a GlcNAc residue, while the second sugar is di-*N*-acetyl-p-glucuronic acid.⁴ Unlike the *M. voltae* flagellin proteins where every N-linked sequon from each flagellin protein was found to carry a glycan, in *M. maripaludis*, N²⁶ of the first

N-linked sequon of FlaB1 and FlaB2 was not glycosylated, suggesting that these sites are not accessible in *M. maripaludis* for attachment of a glycan. A total of nine sites of N-linked glycosylation were identified in FlaB1, FlaB2 and FlaB3 (Fig. 1, Supplementary Table 1). The mass of each glycopeptide was correlated to a single tetrasaccharide moiety at each N-linked site.

2.2. Purification of flagellar glycan

The flagellar glycan was purified following extensive pronase digestion of purified flagella using size-exclusion and anion-exchange chromatography. Following desalting, the column fractions were monitored by MS to identify glycan-containing samples. These were subsequently lyophilised and examined by NMR spectroscopy.

2.3. NMR structural analysis of pronase-digested glycopeptide

Standard homo- and heteronuclear NMR experiments were used to characterise the structure of the flagellar glycan. The structure is presented in Figure 3, and chemical shifts for each sugar residue and derivative are presented in Table 1. Spectra assignment revealed four monosaccharide spin systems. Residues C and D were identified on the basis of chemical shifts and coupling constants as β-GalNAc and β-GlcNAc3NAcA. Residue B had a β-manno configuration as was determined from coupling constants and an NOE between H-1 and H-3, and H-5. It had amino groups at C-2 and C-3 (C-2 at 53.4 and C-3 at 57.1 ppm). ¹³C chemical shifts for this monosaccharide were close to those of the same sugar in Pseudomonas aeruginosa O3a,c polysaccharide, where β-D-ManNAc3NAmA is substituted by β-D-FucNAc.¹⁶ The amino group at C-3 of 2,3-diaminomannuronic acid was substituted by an amidine residue, as followed from the observation of the HMBC correlation between proton B3 and C-1 of the amidine group, which agreed with the characteristic low field shift of the C-3 signal for amidine substitution (compared with acetylation). Mass spectral data confirmed that the amidine unit was attached to residue B. The linkage between residue B and threonine was determined from HMBC data as described previously.⁴ Threonine was differentiated from allo-threonine based on the position of its C-4 signal (20.6 ppm,

FlaB1 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK ESTEQVASGL LINGITGSVG TSDVKLLAIY LAPNAGSSAI DLAQTKVMLD YNGKSVVLGY GGNQDMSSGN SSVFSNDTGA TATTFQVNIL QDYDDSAVDN AVINKGDAVA LIVDVNASFA GEIPERTAIS GKVQPEFGAP GVISFTTPAS YTTTLVELQ FlaB2 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTYGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT MDVNSTSKVS GEVIPDYGAS GIVEFRAPSV FTETVVTLQ					
ESTEQVASGL LINGITGSVG TSDVKLLAIY LAPNAGSSAI DLAQTKVMLD YNGKSVVLGY GGNQDMSSGN SSVFSNDTGA TATTFQVNIL QDYDDSAVDN AVINKGDAVA LIVDVNASFA GEIPERTAIS GKVQPEFGAP GVISFTTPAS YTTTLVELQ FlaB2 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	FlaB1				
DLAQTKVMLD YNGKSVVLGY GGNQDMSSGN SSVFSNTTGA TATTFQVNIL QDYDDSAVDN AVINKGDAVA LIVDVNASFA GEIPERTAIS GKVQPEFGAP GVISFTTPAS YTTTLVELQ FlaB2 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	ASGIGTLIVF	IAMVLVAAVA	ASVLI NTS GF	LQQKASTTGK	
TATTFQVNIL QDYDDSAVDN AVINKGDAVA LIVDVNASFA GEIPERTAIS GKVQPEFGAP GVISFTTPAS YTTTLVELQ FlaB2 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	ESTEQVASGL	LINGITGSVG	TSDVKLLAIY		
FlaB2 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	DLAQTKVMLD	YNGKSVVLGY	GGNQDMSSG N	SSVFSNDTGA	
FlaB2 ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	TATTFQVNIL	QDYDDSAVDN	AVINKGDAVA	LIVDV NAS FA	
ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVMWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	GEIPERTAIS	GKVQPEFGAP	GVISFTTPAS	YTTTLVELQ	
ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVMWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT					
ASGIGTLIVF IAMVLVAAVA ASVLINTSGF LQQKASTTGK DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVMWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT					
DSTEQVASGL QIMGISGYQA GTANANITKL AIYITPNAGS AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	FlaB2				
AAIDMNQVVL TLSDGTTKTV TKYDTTAYTN LTAGGDLYNT TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	ASGIGTLIVF	IAMVLVAAVA	ASVLI NTS GF	LQQKASTTGK	
TTVNWSKLAD TTEFGIVEIQ DADLSFTSSA PVINKGDIVA IIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	DSTEQVASGL	QIMGISGYQA	GTANA NIT KL	AIYITPNAGS	
TIVSGVSFDT RMEISGTVQP EFGAPGVISF TTPSTFTEKV VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	AAIDMNQVVL	TLSDGTTKTV	TKYDTTAYT N	LT AGGDLY NT	
VSLQ FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	TTV NWS KLAD	TTEFGIVEIQ	DADLSFTSSA	PVINKGDIVA	
FlaB3 AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSDNS TAEATHPTIN YGDKVYITVT	IIVSGVSFDT	RMEISGTVQP	EFGAPGVISF	TTPSTFTEKV	
AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSD NS T AEATHPTIN YGDKVYITVT	VSLQ				
AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSD NS T AEATHPTIN YGDKVYITVT					
AVGIGTLIIF IAMVLVAAIA ASVIINTAGK LQHKASTVGE ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGEI DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSD NS T AEATHPTIN YGDKVYITVT					
ESTEQVASGI QVLKVIGHAD TKTTIDKLAV MVAPNVGGET DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSD NS TAEATHPTIN YGDKVYITVT	FlaB3				
DLSTTILTLG TGDAKYSLVY DSTQHNADVK DDGSDSIFDE TEWGSGSKYG LIVLQDSD NS T AEATHPTIN YGDKVYITVT	AVGIGTLIIF	IAMVLVAAIA	ASVIINTAGK	LQHKASTVGE	
TEWGSGSKYG LIVLQDSD NS T AEATHPTIN YGDKVYITVT	ESTEQVASGI	QVLKVIGHAD	TKTTIDKLAV	MVAPNVGGEI	
2 2 2	DLSTTILTLG	TGDAKYSLVY	DSTQHNADVK	DDGSDSIFDE	
MDV NST SKVS GEVIPDYGAS GIVEFRAPSV FTETVVTLQ	TEWGSGSKYG	LIVLQDSD NS	T AEATHPTIN	YGDKVYITVT	
	MDV nst skvs	GEVIPDYGAS	GIVEFRAPSV	FTETVVTLQ	

Figure 1. Assignment map of flagellin proteins from *M. maripaludis*. The predicted amino acid sequence for the mature flagellin proteins, FlaB1, FlaB2 and FlaB3, is presented, and the N-linked sequence are highlighted. Peptides and glycopeptides identified by nanoLC–MS/MS of the flagellar tryptic digest are underlined. Note that all N-linked sites are occupied with the exception of N²⁶ in FlaB1 and FlaB2. Only the unmodified peptide, T¹⁻³⁴, was observed by nanoLC–MS/MS for these two flagellin proteins.

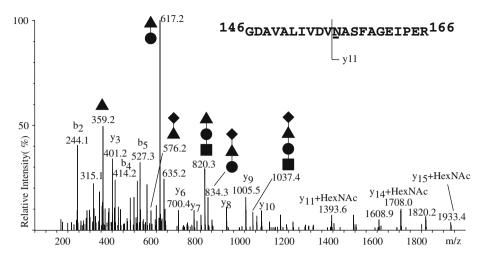


Figure 2. NanoLC-MS/MS analysis of the tryptic glycopeptide, $T^{146-166}$. MS/MS spectrum of the triply protonated glycopeptide ion at m/z 1060.5. The amino acid sequence of this peptide is presented in the inset, and the site of the N-linkage is underlined. The b- and y-fragment ions arising from the fragmentation of the peptide bonds are indicated in the spectrum. Note that the y-fragment ions containing the N¹⁵⁶ residue (y_{11} and greater) are all heavier than that predicted by extra 203 Da, indicating that the glycan is N-linked through a HexNAc residue. The major carbohydrate oxonium ions are identified in this spectrum using symbols to indicate the sugar residues present (■ 203 Da; ● 258 Da; \bullet 358 Da and \bullet 217 Da). The intact glycan oxonium ion is observed at m/z 1037.4 and loss of the 203 Da (■) and 217 Da (\bullet) sugars yields the ions at m/z 834.3 and 820.3, respectively. The intense oxonium ion at m/z 617.2 corresponds to the 258 Da (\bullet) and 358 Da (\bullet) sugar residues. The weaker ion at m/z 576.2 corresponds to the 358 Da (\bullet) and 217 Da sugars (\bullet). The only single residue oxonium ion observed originates from the 358 Da sugar (\bullet). The only single residue oxonium ion observed

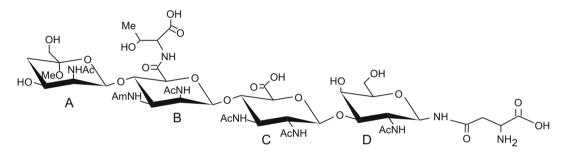


Figure 3. Structure of M. maripaludis flagellar glycan.

Table 1NMR chemical shifts of flagellar glycan^a

Unit		1	2	3	4	5/Me	6a/b
β-GalNAc D	H C	5.03 79.6	4.06 50.6	3.84 81.5	4.19 68.9	3.73 77.6	3.73 60.4
β-GlcNAc3NAcA C	H C	4.68 103.5	3.87 54.5	3.95 54.5	3.78 79.2	3.81 77.7	176.4
β-ManNAc3NAcA B	H C	4.74 100.1	4.45 51.6	4.13 57.1	3.87 74.7	4.18 77.5	176.9
A	H C	4.78 J _{1,2} 1.8 97.0	4.28 J _{2,3} 4 53.4	4.07 J _{3,4ax} 12 65.6	1.43/2.00 J _{3.4eq} 5 34.7	J _{4ax,4eq} 13 101.2	3.38/3.69 J _{6a,6b} 11.8 62.4
A OMe	H C					3.26 50.0	
Thr	H C	176.9	4.18 61.4	4.27 70.0	1.24 20.6		
Asn	H C	174.3	3.97 52.1	2.91/2.91 36.0			

^a Amidine signal (C-1, H-2/C-2): 168.0, 2.26/19.9; *N*-acetates (not assigned to particular locations): 176.1, 1.96/23.3; 176.1, 1.96/23.3; 175.9, 2.00.23.3; 177.3, 2.07/23.3; 177.5, 2.14/23.3 ppm.

expected for *allo*-threonine is 17.3 ppm). The absolute configuration of the threonine residue has not been determined.

Residue A was identified as 2-acetamido-2,4-dideoxyhexose (two H-4 protons, ¹³C NMR C-2 signal at 34.7 ppm, C-4 signal at

53.4 ppm). No H-5 proton was present. HMBC correlations from H-4 and H-6 identified the C-5 resonance at 101.2 ppm. HMBC and NOE correlations showed that an OMe group was attached to C-5 (Figs. 4 and 5). Thus, this monosaccharide was the 5-methyl-

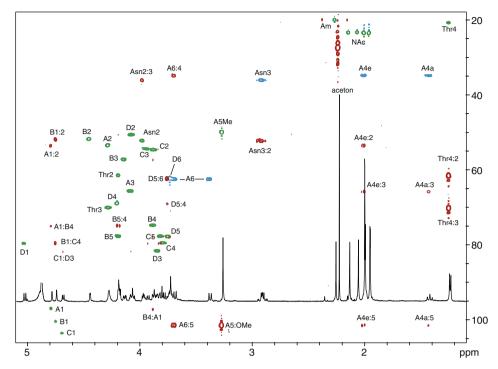


Figure 4. Overlap of ¹H, HSQC (green, cyan) and HMBC (red) spectra of the *M. maripaludis* flagellar oligosaccharide.

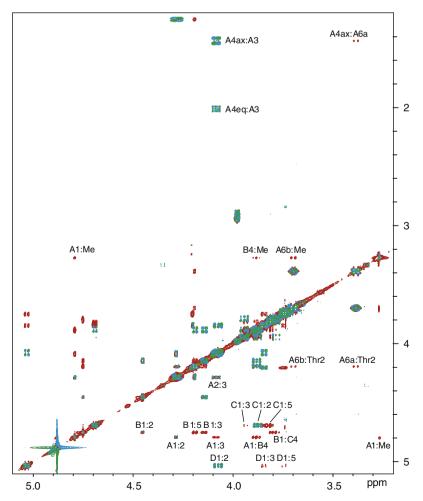


Figure 5. Overlap of COSY (green, cyan) and NOESY (red) spectra of the M. maripaludis flagellar oligosaccharide.

glycoside of a hexos-5-ulose. Vicinal coupling constants (Table 1) between ring protons corresponded to the α -erythro configuration of the C-1–C-3 fragment. Strong NOE correlations between protons A1:A3, A1:Me, A6b:Me, and A4ax:A6a indicated that the OMe group occupied the axial position in the 4C_1 conformation, and the whole configuration of this 2-acetamido-2,4-dideoxy-5-O-methyl-hexos-5-ulopyranose was isomorphic to the 4-deoxy- β -D-manno (D-configuration is discussed below).

The monosaccharide sequence as shown in Figure 3 followed from the NOE between protons A1:B4, B1:C4,5, C1:D3, A6a,b:Thr2; B4:Me and HMBC (proton:carbon, A1:B4, B1:C4, C1:D3, D1:Asn C-1) data.

The absolute configuration of the D-GalNAc was determined by GC-MS of acetylated 2-butyl glycosides. 17 The D-configuration of the residues B and C was deduced on the basis of glycosylation effects in the ¹³C NMR spectra. ¹⁸ Thus, for the residue of β-Glc-NAc3NAcA, the position of its C-1 signal (103.5 ppm) was closer to that expected for the same absolute configuration as D-GalNAc (104.7 ppm) than for different configurations (99.6 ppm). The C-3 and C-4 signals of GalNAc (81.5 and 68.9 ppm) were also closer to the DD combination (expected value 81.2 and 69.2 ppm) than to the LD combination (expected value 79.2 and 66.3 ppm). For the β-ManNAc3NAcA, its C-1 signal (100.1 ppm) is closer to the expected value for the DD-pair (100.5 ppm) than to that of LD-pair (101.4 ppm); C-3 signal of the GlcNAc3NAcA also fits slightly better for DD-pair. The same method was used for the determination of the configuration of residue A: A comparison of the position of the C-4 signal of ManNAc3NAmA6Thr (74.7 ppm) with that of the ManNAc3NAmA substituted with α -D-FucNAc (71.4 ppm) and β-D-FucNAc (75.1 ppm) from different P. aeruginosa polysaccharides¹⁶ led to the conclusion that the monosaccharide A was (5S)-2-acetamido-2,4-dideoxy-5-0-methyl- α -L-erythro-hexos-5-ulo-1,5-pyranose, as shown in Figure 3.

3. Discussion

NMR spectroscopy and mass spectrometry have been utilised to deduce the structure of the flagellar glycan from M. maripaludis. In comparison to the structure elucidated for the flagellar glycan from the related organism M. voltae, 4,14 the glycan presented on the flagellins of this organism appears to be novel in structure. M voltae has been shown to glycosylate its flagellin with either the trisaccharide β-ManNAc3NAcA6Thr-4-β-GlcNAc3NAcA-3-β-GlcNAc-Asn or this trisaccharide carrying an additional mass of either 220 or 262 Da.¹⁴ In contrast, we demonstrate in this study that the flagellar glycan from M. maripaludis is composed of a novel tetrasaccharide that is attached in an N-linkage via β -GalNAc. The significance of this linkage to the flagellin proteins via β -GalNAc rather than β-GlcNAc is not known, but may be a reflection of the specificity of the respective glycosyltransferases from each strain. In M. voltae AglH (Mv1751), it has been shown via complementation in yeast to function as the β-GlcNAc transferase.8

In addition to containing β -GalNAc and β -GlcNAc3NAcA, the M. maripaludis flagellar glycan also contains a 3-acetamidino derivative of 2,3-diamino-2,3-dideoxymannuronic acid amidated with a threonine amino group (β -ManNAc3NAmA6Thr), as well as a unique terminal sugar, 2-acetamido-2,4-dideoxy-5- θ -methyl-hexos-5-ulo-1,5-pyranose. The presence of this sugar to our knowledge is the first example of a naturally occurring digly-coside of an aldulose. As the genome sequence of θ -maripaludis is available, and genetic techniques exist that allow targeted studies of individual genes, it will now be possible to identify the biosynthetic enzymes responsible for the production of this novel structure. θ -19,20 The biological role of the tetrasaccharide, and more significantly the novel terminal sugar, is currently unknown but it

is clearly of relevance in the emerging field of prokaryotic glycobiology. The structural elucidation of the glycan monosaccharides may now facilitate the identification of corresponding biosynthetic enzymes.

4. Experimental

4.1. Growth of $\it M.$ maripaludis and extraction of flagellar filaments

M. maripaludis strain Mm900 is a Δhpt mutant of *M. maripaludis* strain S2 (formerly LL) that was developed for markerless mutagenesis. ²⁰ It was grown at 30 °C in Balch medium III under an atmosphere of CO_2 – H_2 (20:80). ²⁰ Crude flagellar filament preparations were obtained by detergent extraction of cell membranes followed by banding in a KBr gradient. ²¹

4.2. MS analysis of flagellin tryptic peptides

For MS characterisation of flagellar glycan, flagellar filaments ($\sim\!200~\mu g)$ were digested overnight with trypsin (Promega, Madison WI) at a ratio of 30:1 (protein:enzyme, v/v) in 50 mM ammonium bicarbonate at 37 °C. The protein digest was analysed by nano-LC–MS/MS using a QTOF Ultima hybrid quadrupole time-of-flight mass spectrometer coupled to a NanoAquity UPLC system. The digest was separated on a 100 $\mu m \times 100$ mm, 1.7 μm BEH130 C_{18} column (Waters) using a standard gradient of acetonitrile and 0.2% formic acid (400 nL/min.). MS/MS spectra were acquired on doubly, triply and quadruply charged ions. The resulting MS/MS spectra were submitted to the Mascot® search engine (Matrix Science Ltd., London, UK) and searched against the NCBInr database. Unidentified spectra were then examined manually for the presence of post-translational modifications.

4.3. Purification of flagellar glycan

To obtain sufficient amounts of pure glycan for structural analysis, we digested 2.5 mg of flagellar filaments with pronase (50 μ g) in 0.1 M Tris HCl, 2 mM CaCl₂, pH 8.0. After incubation at 37 °C for 24 h, a second aliquot of pronase was added (30 μ g), and digestion was allowed to continue for 24 h. The lyophilised digest was resuspended in distilled H₂O and applied to a Bio-Gel P4 size-exclusion column. Fractions were collected and examined by infusion-MS/MS on a LTQ linear ion-trap mass spectrometer (ThermoFisher Scientific), and those containing glycan were pooled and applied to a Bio-Gel P2 column. The eluate was collected in 1-mL fractions, concentrated and again examined by MS to identify fractions containing glycan. Fractions were pooled and applied to an anion-exchange Hitrap Q column, and glycan was eluted with a linear gradient of NaCl (0–1 M, 1 h). Desalting was performed on Sephadex G15 prior to analysis by NMR spectroscopy.

4.4. NMR spectroscopic analysis

NMR experiments were carried out with a Varian INOVA 600 MHz (1 H) spectrometer with a Varian Z gradient probe at 25 $^{\circ}$ C in 3-mm tube with acetone internal reference (2.225 ppm for 1 H and 31.5 ppm for 13 C) using the standard pulse sequences DQCOSY, TOCSY (mixing time 120 ms), NOESY (mixing time 200 ms), HSQC and HMBC (100 ms long-range transfer delay).

Acknowledgements

The authors thank D. Watson (NRC-IBS) for helpful assistance with flagellar glycan purification and L. Tessier and W. Ding

(NRC-IBS) for assistance with the mass spectrometric analysis. This work was funded in part by a Discovery Grant from NSERC (to KFJ).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carres.2009.01.006.

References

- 1. Logan, S. M. Microbiology 2006, 152, 1249-1262.
- Ng, S. Y.; Zolghadr, B.; Driessen, A. J.; Albers, S. V.; Jarrell, K. F. J. Bacteriol. 2008, 190, 6039–6047.
- 3. Trachtenberg, S.; Cohen-Krausz, S. J. Mol. Microbiol. Biotechnol. 2006, 11, 208–220.
- Voisin, S.; Houliston, R. S.; Kelly, J.; Brisson, J. R.; Watson, D.; Bardy, S. L.; Jarrell, K. F.; Logan, S. M. *J. Biol. Chem.* 2005, 280, 16586–16593.
- 5. Sumper, M. Biochim. Biophys. Acta 1987, 906, 69-79.
- 6. Wieland, F.; Paul, G.; Sumper, M. J. Biol. Chem. 1985, 260, 15180-15185.
- Chaban, B.; Voisin, S.; Kelly, J.; Logan, S. M.; Jarrell, K. F. Mol. Microbiol. 2006, 61, 259–268.
- 8. Shams-Eldin, H.; Chaban, B.; Niehus, S.; Schwarz, R. T.; Jarrell, K. F. J. Bacteriol. **2008**, *190*, 2217–2220.
- Yurist-Doutsch, S.; Chaban, B.; VanDyke, D. J.; Jarrell, K. F.; Eichler, J. Mol. Microbiol. 2008, 68, 1079–1084.

- 10. Abu-Qarn, M.; Eichler, J. Mol. Microbiol. 2006, 61, 511-525.
- Abu-Qarn, M.; Giordano, A.; Battaglia, F.; Trauner, A.; Hitchen, P. G.; Morris, H. R.; Dell, A.; Eichler, J. J. Bacteriol. 2008, 190, 3140–3146.
- Abu-Qarn, M.; Yurist-Doutsch, S.; Giordano, A.; Trauner, A.; Morris, H. R.; Hitchen, P.; Medalia, O.; Dell, A.; Eichler, J. J. Mol. Biol. 2007, 374, 1224–1236.
- Yurist-Doutsch, S.; Abu-Qarn, M.; Battaglia, F.; Morris, H. R.; Hitchen, P. G.; Dell, A.; Eichler, J. Mol. Microbiol. 2008, 69, 1234–1245.
- 14. Chaban, B.; Logan, S. M.; Kelly, J. F.; Jarrell, K. F. J. Bacteriol. 2009, 191, 187–195.
- VanDyke, D. J.; Wu, J.; Ng, S. Y.; Kanbe, M.; Chaban, B.; Aizawa, S.; Jarrell, K. F. J. Bacteriol. 2008, 190, 5300–5307.
- Knirel, Y. A.; Paramonov, N. A.; Vinogradov, E. V.; Shashkov, A. S.; Dmitriev, B. A.; Kochetkov, N. K.; Kholodkova, E. V.; Stanislavsky, E. S. Eur. J. Biochem. 1987, 167. 549–561.
- Gerwig, G. J.; Kamerling, J. P.; Vliegenthart, J. F. Carbohydr. Res. 1978, 62, 349–357
- Lipkind, G. M.; Shashkov, A. S.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K. Carbohydr. Res. 1988, 175, 59–75.
- Hendrickson, E. L.; Kaul, R.; Zhou, Y.; Bovee, D.; Chapman, P.; Chung, J.; Conway, d. M.; Dodsworth, J. A.; Gillett, W.; Graham, D. E.; Hackett, M.; Haydock, A. K.; Kang, A.; Land, M. L.; Levy, R.; Lie, T. J.; Major, T. A.; Moore, B. C.; Porat, I.; Palmeiri, A.; Rouse, G.; Saenphimmachak, C.; Soll, D.; Van Dien, S.; Wang, T.; Whitman, W. B.; Xia, Q.; Zhang, Y.; Larimer, F. W.; Olson, M. V.; Leigh, J. A. J. Bacteriol. 2004, 186, 6956–6969.
- 20. Moore, B. C.; Leigh, J. A. J. Bacteriol. 2005, 187, 972-979.
- Bardy, S. L.; Mori, T.; Komoriya, K.; Aizawa, S.; Jarrell, K. F. J. Bacteriol. 2002, 184, 5223–5233.