Aas, F. E., A. Vik, et al. (2007). "Neisseria gonorrhoeae O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure." Mol Microbiol **65**(3): 607-24.

Neisseria gonorrhoeae expresses an O-linked protein glycosylation pathway that targets PilE, the major pilin subunit protein of the Type IV pilus colonization factor. Efforts to define glycan structure and thus the functions of pilin glycosylation (Pgl) components at the molecular level have been hindered by the lack of sensitive methodologies. Here, we utilized a 'top-down' mass spectrometric approach to characterize glycan status using intact pilin protein from isogenic mutants. These structural data enabled us to directly infer the function of six components required for pilin glycosylation and to define the glycan repertoire of strain N400. Additionally, we found that the N. gonorrhoeae pilin glycan is O-acetylated, and identified an enzyme essential for this unique modification. We also identified the N. gonorrhoeae pilin oligosaccharyltransferase using bioinformatics and confirmed its role in pilin glycosylation by directed mutagenesis. Finally, we examined the effects of expressing the PglA glycosyltransferase from the Campylobacter jejuni N-linked glycosylation system that adds N-acetylgalactosamine onto undecaprenylpyrophosphate-linked bacillosamine. The results indicate that the C. jejuni and N. gonorrhoeae pathways can interact in the synthesis of O-linked di- and trisaccharides, and therefore provide the first experimental evidence that biosynthesis of the N. gonorrhoeae pilin glycan involves a lipid-linked oligosaccharide precursor. Together, these findings underpin more detailed studies of pilin glycosylation biology in both N. gonorrhoeae and N. meningitidis, and demonstrate how components of bacterial O- and N-linked pathways can be combined in novel glycoengineering strategies.

Abu-Qarn, M. and J. Eichler (2006). "Protein N-glycosylation in Archaea: defining Haloferax volcanii genes involved in S-layer glycoprotein glycosylation." Mol Microbiol **61**(2): 511-25.

In this study, characterization of the N-glycosylation process in the haloarchaea Haloferax volcanii was undertaken. Initially, putative Hfx. volcanii homologues of genes involved in eukaryal or bacterial N-glycosylation were identified by bioinformatics. Reverse transcription polymerase chain reaction (RT-PCR) confirmed that the proposed N-glycosylation genes are transcribed, indicative of true proteins being encoded. Where families of related gene sequences were detected, differential transcription of family members under a variety of physiological and environmental conditions was shown. Gene deletions point to certain genes, like alg11, as being essential yet revealed that others, such as the two versions of alg5, are not. Deletion of alg5-A did, however, lead to slower growth and interfered with surface (S)-layer glycoprotein glycosylation, as detected by modified migration on SDS-PAGE and glycostaining approaches. As deletion of stt3, the only component of the oligosaccharide transferase complex detected in Archaea, did not affect cell viability, it appears that N-glycosylation is not essential in Hfx. volcanii. Deletion of stt3 did, nonetheless, hinder both cell growth and S-layer glycoprotein glycosylation. Thus, with genes putatively

involved in Hfx. volcanii protein glycosylation identified and the ability to address the roles played by the encoded polypeptides in modifying a reporter glycoprotein, the steps of the archaeal N-glycosylation pathway can be defined.

Abu-Qarn, M., J. Eichler, et al. (2008). "Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea." Curr Opin Struct Biol **18**(5): 544-50.

Of the many post-translational modifications proteins can undergo, glycosylation is the most prevalent and the most diverse. Today, it is clear that both N-glycosylation and O-glycosylation, once believed to be restricted to eukaryotes, also transpire in Bacteria and Archaea. Indeed, prokaryotic glycoproteins rely on a wider variety of monosaccharide constituents than do those of eukaryotes. In recent years, substantial progress in describing the enzymes involved in bacterial and archaeal glycosylation pathways has been made. It is becoming clear that enhanced knowledge of bacterial glycosylation enzymes may be of therapeutic value, while the demonstrated ability to introduce bacterial glycosylation genes into Escherichia coli represents a major step forward in glyco-engineering. A better understanding of archaeal protein glycosylation provides insight into this post-translational modification across evolution as well as protein processing under extreme conditions. Here, we discuss new structural and biosynthetic findings related to prokaryotic protein glycosylation, until recently a neglected topic.

Abu-Qarn, M., A. Giordano, et al. (2008). "Identification of AglE, a second glycosyltransferase involved in N glycosylation of the Haloferax volcanii S-layer glycoprotein." J Bacteriol **190**(9): 3140-6.

Archaea, like Eukarya and Bacteria, are able to N glycosylate select protein targets. However, in contrast to relatively advanced understanding of the eukaryal N glycosylation process and the information being amassed on the bacterial process, little is known of this posttranslational modification in Archaea. Toward remedying this situation, the present report continues ongoing efforts to identify components involved in the N glycosylation of the Haloferax volcanii S-layer glycoprotein. By combining gene deletion together with mass spectrometry, AglE, originally identified as a homologue of murine Dpm1, was shown to play a role in the addition of the 190-Da sugar subunit of the novel pentasaccharide decorating the S-layer glycoprotein. Topological analysis of an AglE-based chimeric reporter assigns AglE as an integral membrane protein, with its N terminus and putative active site facing the cytoplasm. These finding, therefore, contribute to the developing picture of the N glycosylation pathway in Archaea.

Abu-Qarn, M., S. Yurist-Doutsch, et al. (2007). "Haloferax volcanii AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer." J Mol Biol **374**(5): 1224-36.

In this study, the effects of deleting two genes previously implicated in Haloferax volcanii N-glycosylation on the assembly and attachment of a novel Asn-linked pentasaccharide decorating the H. volcanii S-layer glycoprotein were considered. Mass spectrometry revealed the pentasaccharide to comprise two hexoses, two

hexuronic acids and an additional 190 Da saccharide. The absence of AglD prevented addition of the final hexose to the pentasaccharide, while cells lacking AglB were unable to N-glycosylate the S-layer glycoprotein. In AglD-lacking cells, the S-layer glycoprotein-based surface layer presented both an architecture and protease susceptibility different from the background strain. By contrast, the absence of AglB resulted in enhanced release of the S-layer glycoprotein. H. volcanii cells lacking these N-glycosylation genes, moreover, grew significantly less well at elevated salt levels than did cells of the background strain. Thus, these results offer experimental evidence showing that N-glycosylation endows H. volcanii with an ability to maintain an intact and stable cell envelope in hypersaline surroundings, ensuring survival in this extreme environment.

Arora, S. K., M. Bangera, et al. (2001). "A genomic island in Pseudomonas aeruginosa carries the determinants of flagellin glycosylation." Proc Natl Acad Sci U S A **98**(16): 9342-7.

Protein glycosylation has been long recognized as an important posttranslational modification process in eukaryotic cells. Glycoproteins, predominantly secreted or surface localized, have also been identified in bacteria. We have identified a cluster of 14 genes, encoding the determinants of the flagellin glycosylation machinery in Pseudomonas aeruginosa PAK, which we called the flagellin glycosylation island. Flagellin glycosylation can be detected only in bacteria expressing the a-type flagellin sequence variants, and the survey of 30 P. aeruginosa isolates revealed coinheritance of the a-type flagellin genes with at least one of the flagellin glycosylation island genes. Expression of the b-type flagellin in PAK, an a-type strain carrying the glycosylation island, did not lead to glycosylation of the b-type flagellin of PAO1, suggesting that flagellins expressed by b-type bacteria not only lack the glycosylation island, they cannot serve as substrates for glycosylation. Providing the entire glycosylation island of PAK, including its a-type flagellin in a flagellin mutant of a b-type strain, results in glycosylation of the heterologous flagellin. These results suggest that some or all of the 14 genes on the glycosylation island are the genes that are missing from strain PAO1 to allow glycosylation of an appropriate flagellin. Inactivation of either one of the two flanking genes present on this island abolished flagellin glycosylation. Based on the limited homologies of these gene products with enzymes involved in glycosylation, we propose that the island encodes similar proteins involved in synthesis, activation, or polymerization of sugars that are necessary for flagellin glycosylation.

Arora, S. K., M. C. Wolfgang, et al. (2004). "Sequence polymorphism in the glycosylation island and flagellins of Pseudomonas aeruginosa." J Bacteriol **186**(7): 2115-22.

A genomic island consisting of 14 open reading frames, orfA to orfN was previously identified in Pseudomonas aeruginosa strain PAK and shown to be essential for glycosylation of flagellin. DNA microarray hybridization analysis of a number of P. aeruginosa strains from diverse origins showed that this island is polymorphic. PCR and sequence analysis confirmed that many P. aeruginosa

strains carry an abbreviated version of the island (short island) in which orfD, -E and -H are polymorphic and orfI, -J, -K, -L, and -M are absent. To ascertain whether there was a relationship between the inheritance of the short island and specific flagellin sequence variants, complete or partial nucleotide sequences of flagellin genes from 24 a-type P. aeruginosa strains were determined. Two distinct flagellin subtypes, designated A1 and A2, were apparent. Strains with the complete 14-gene island (long island) were almost exclusively of the A1 type, whereas strains carrying the short island were associated with both A1- and A2-type flagellins. These findings indicate that P. aeruginosa possesses a relatively low number of distinct flagellin types and probably has the capacity to further diversify this antigenic surface protein by glycosylation.

Bartels, K. M., H. Funken, et al. "Glycosylation is required for outer membrane localization of the lectin LecB in Pseudomonas aeruginosa." J Bacteriol **193**(5): 1107-13.

The fucose-/mannose-specific lectin LecB from Pseudomonas aeruginosa is transported to the outer membrane; however, the mechanism used is not known so far. Here, we report that LecB is present in the periplasm of P. aeruginosa in two variants of different sizes. Both were functional and could be purified by their affinity to mannose. The difference in size was shown by a specific enzyme assay to be a result of N glycosylation, and inactivation of the glycosylation sites was shown by site-directed mutagenesis. Furthermore, we demonstrate that this glycosylation is required for the transport of LecB.

Bayley, D. P. and K. F. Jarrell (1999). "Overexpression of Methanococcus voltae flagellin subunits in Escherichia coli and Pseudomonas aeruginosa: a source of archaeal preflagellin." J Bacteriol **181**(14): 4146-53.

Methanococcus voltae is a flagellated member of the Archaea. Four highly similar flagellin genes have previously been cloned and sequenced, and the presence of leader peptides has been demonstrated. While the flagellins of M. voltae are predicted from their gene sequences to be approximately 22 to 25 kDa, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of purified flagella revealed flagellin subunits with apparent molecular masses of 31 and 33 kDa. Here we describe the expression of a M. voltae flagellin in the bacteria Escherichia coli and Pseudomonas aeruginosa. Both of these systems successfully generated a specific expression product with an apparently uncleaved leader peptide migrating at approximately 26.5 kDa. This source of preflagellin was used to detect the presence of preflagellin peptidase activity in the membranes of M. voltae. In addition to the native flagellin, a hybrid flagellin gene containing the sequence encoding the M. voltae FlaB2 mature protein fused to the P. aeruginosa pilin (PilA) leader peptide was constructed and transformed into both wild-type P. aeruginosa and a prepilin peptidase (pilD) mutant of P. aeruginosa. Based on migration in SDS-PAGE, the leader peptide appeared to be cleaved in the wild-type cells. However, the archaeal flagellin could not be detected by immunoblotting when expressed in the pilD mutant, indicating a role of the peptidase in the ultimate stability of the fusion product.

When the +5 position of the mature flagellin portion of the pilin-flagellin fusion was changed from glycine to glutamic acid (as in the P. aeruginosa pilin) and expressed in both wild-type and pilD mutant P. aeruginosa, the product detected by immunoblotting migrated slightly more slowly in the pilD mutant, indicating that the fusion was likely processed by the prepilin peptidase present in the wild type. Potential assembly of the cleaved fusion product by the type IV pilin assembly system in a P. aeruginosa PilA-deficient strain was tested, but no filaments were noted on the cell surface by electron microscopy.

Benz, I. and M. A. Schmidt (2001). "Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin." Mol Microbiol **40**(6): 1403-13.

The diffuse adherence of Escherichia coli strain 2787 (O126:H27) is mediated by the autotransporter adhesin AIDA-I (adhesin-involved-in-diffuse-adherence) encoded by the plasmid-borne aidA gene. AIDA-I exhibits an aberrant mobility in denaturing gel electrophoresis. Deletion of the open reading frame (ORF) A immediately upstream of aidA restores the predicted mobility of AIDA-I, but the adhesin is no longer functional. This indicates that the mature AIDA-I adhesin is post-translationally modified and the modification is essential for adherence function. Labelling with digoxigenin hydrazide shows AIDA-I to be glycosylated. Using carbohydrate composition analysis, AIDA-I contains exclusively heptose residues (ratio heptose:AIDA-I approximately 19:1). The deduced amino acid sequence of the cytoplasmic open reading frame (ORF) A gene product shows homologies to heptosyltransferases. In addition, the modification was completely abolished in an ADP-glycero-manno-heptopyranose mutant. Our results provide direct evidence for glycosylation of the AIDA-I adhesin by heptoses with the ORF A gene product as a specific (mono)heptosyltransferase generating the functional mature AIDA-I adhesin. Consequently, the ORF A gene has been denoted 'aah' (autotransporter-adhesin-heptosyltransferase). Glycosylation by heptoses represents a novel protein modification in eubacteria.

Benz, I. and M. A. Schmidt (2002). "Never say never again: protein glycosylation in pathogenic bacteria." Mol Microbiol **45**(2): 267-76.

In recent years, accumulating evidence for glycosylated bacterial proteins has overthrown an almost dogmatic belief that prokaryotes are not able to synthesize glycoproteins. Now it is widely accepted that eubacteria express glycoproteins. Although, at present, detailed information about glycosylation and structure-function relationships is available for only few eubacterial proteins, the variety of different components and structures observed already indicates that the variations in bacterial glycoproteins seem to exceed the rather limited display found in eukaryotes. Numerous virulence factors of bacterial pathogens have been found to be covalently modified with carbohydrate residues, thereby identifying these factors as true glycoproteins. In several bacterial species, gene clusters suggested to represent a general protein glycosylation system have been identified. In other cases, genes encoding highly specific glycosyltransferases have been found to be directly linked with virulence genes.

These findings raise interesting questions concerning a potential role of glycosylation in pathogenesis. In this review, we will therefore focus on protein glycosylation in Gram-negative bacterial pathogens.

Brimer, C. D. and T. C. Montie (1998). "Cloning and comparison of fliC genes and identification of glycosylation in the flagellin of Pseudomonas aeruginosa a-type strains." J Bacteriol **180**(12): 3209-17.

Pseudomonas aeruginosa a-type strains produce flagellin proteins which vary in molecular weight between strains. To compare the properties of a-type flagellins, the flagellin genes of several Pseudomonas aeruginosa a-type strains, as determined by interaction with specific anti-a monoclonal antibody, were cloned and sequenced. PCR amplification of the a-type flagellin gene fragments from five strains each yielded a 1.02-kb product, indicating that the gene size is not likely to be responsible for the observed molecular weight differences among the a-type strains. The flagellin amino acid sequences of several a-type strains (170,018, 5933, 5939, and PAK) were compared, and that of 170,018 was compared with that of PAO1, a b-type strain. The former comparisons revealed that a-type strains are similar in amino acid sequence, while the latter comparison revealed differences between 170,018 and PAO1. Posttranslational modification was explored for its contribution to the observed differences in molecular weight among the a-type strains. A biotin-hydrazide glycosylation assay was performed on the flagellins of three a-type strains (170,018, 5933, and 5939) and one b-type strain (M2), revealing a positive glycosylation reaction for strains 5933 and 5939 and a negative reaction for 170,018 and M2. Deglycosylation of the flagellin proteins with trifluoromethanesulfonic acid (TFMS) confirmed the glycosylation results. A molecular weight shift was observed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis for the TFMS-treated flagellins of 5933 and 5939. These results indicate that the molecular weight discrepancies observed for the a-type flagellins can be attributed, at least in part, to glycosylation of the protein. Anti-a flagellin monoclonal antibody reacted with the TFMS-treated flagellins, suggesting that the glycosyl groups are not a necessary component of the epitope for the human anti-a monoclonal antibody. Comparisons between a-type sequences and a b-type sequence (PAO1) will aid in delineation of the epitope for this monoclonal antibody.

Brockl, G., M. Behr, et al. (1991). "Analysis and nucleotide sequence of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogens Methanothermus fervidus and Methanothermus sociabilis." Eur J Biochem **199**(1): 147-52.

The genes (slgA) encoding the surface-layer glycoproteins of the hyperthermophilic methanogens Methano-thermus pervidus and Methanothermus sociabilis were cloned and sequenced. The nucleotide sequences of these genes differ at only nine positions, resulting in three amino acid differences. In both organisms, the transcription start site was localized by primer extension analyses. The DNA sequence at this site conforms to the

promotor box B motif for promotors of archaea. 24 nucleotides upstream of the transcription start is an A + T-rich region, which closely resembles the consensus box A motif of promoters of methanogens. Ribosome binding sites are exactly complementary to the 3' end of the 16S rRNA of these methanogens. Both slgA genes encode for a precursor of the mature surface-layer protein containing 593 amino acid residues with a putative N-terminal signal sequence of 22 amino acid residues. The deduced protein sequences contain 20 sequon structures representing possible carbohydrate-binding sites. In comparison with other surface-layer proteins, these obtained from the two hyperthermophilic methanogens contain unusually high amounts of isoleucine, asparagine and cysteine residues. Predicted secondary structures have a high content of beta-sheet structure (44%) and only 7% alpha-helix structures.

Calo, D., L. Kaminski, et al. "Protein glycosylation in Archaea: sweet and extreme." Glycobiology **20**(9): 1065-76.

While each of the three domains of life on Earth possesses unique traits and relies on characteristic biological strategies, some processes are common to Eukarya, Bacteria and Archaea. Once believed to be restricted to Eukarya, it is now clear that Bacteria and Archaea are also capable of performing N-glycosylation. However, in contrast to Bacteria, where this posttranslational modification is still considered a rare event, numerous species of Archaea, isolated from a wide range of environments, have been reported to contain proteins bearing Asn-linked glycan moieties. Analysis of the chemical composition of the Asn-linked polysaccharides decorating archaeal proteins has, moreover, revealed the use of a wider variety of sugar subunits than seen in either eukaryal or bacterial glycoproteins. Still, although first reported some 30 years ago, little had been known of the steps or components involved in the archaeal version of this universal posttranslational modification. Now, with the availability of sufficient numbers of genome sequences and the development of appropriate experimental tools, molecular analysis of archaeal N-glycosylation pathways has become possible. Accordingly using halophilic, methanogenic and thermophilic model species, insight into the biosynthesis and attachment of N-linked glycans decorating archaeal glycoproteins is starting to amass. In this review, current understanding of N-glycosylation in Archaea is described.

Castric, P. (1995). "pilO, a gene required for glycosylation of Pseudomonas aeruginosa 1244 pilin." Microbiology **141 (Pt 5)**: 1247-54.

Nucleotide sequencing of a region downstream from the Pseudomonas aeruginosa 1244 pilin structural gene, pilA, revealed an ORF potentially able to code for a protein of M(r) 50,862. This ORF, called pilO, was flanked by a tRNAthr gene, which was followed by a transcriptional termination sequence. The tRNAthr gene and the termination sequence were nearly identical to sequences found immediately adjacent to the pilA gene of several P. aeruginosa strains. A 2200 base mRNA strand, which contained both the pilO and pilA transcripts, was produced from this region, while a 650 base transcript containing only pilA was present in a 100-fold excess over the longer transcript. Hyperexpression of the

pilA gene in a PilO- strain resulted in normal pilus-specific phage sensitivity and twitching motility. The pilin produced by this strain had a lower apparent M(r) and a more neutral pI compared to that produced by a strain containing a functional pilO gene. This pilin failed to react with a sugar-specific reagent which recognized pilin produced by the strain containing a functional pilO gene.

Castric, P., F. J. Cassels, et al. (2001). "Structural characterization of the Pseudomonas aeruginosa 1244 pilin glycan." J Biol Chem **276**(28): 26479-85.

An antigenic similarity between lipopolysaccharide (LPS) and glycosylated pilin of Pseudomonas aeruginosa 1244 was noted. We purified a glycan-containing molecule from proteolytically digested pili and showed it to be composed of three sugars and serine. This glycan competed with pure pili and LPS for reaction with an LPS-specific monoclonal antibody, which also inhibited twitching motility by P. aeruginosa bearing glycosylated pili. One-dimensional NMR analysis of the glycan indicated the sugars to be 5N beta OHC(4)7NfmPse, Xyl, and FucNAc. The complete proton assignments of these sugars as well as the serine residue were determined by COSY and TOCSY. Electrospray ionization mass spectrometry (MS) determined the mass of this molecule to be 771.5. The ROESY NMR spectrum, tandem MS/MS analysis, and methylation analysis provided information on linkage and the sequence of oligosaccharide components. These data indicated that the molecule had the following structure: alpha-5N beta OHC(4)7NFmPse-(2-->4)beta-Xyl-(1-->3)-beta-FucNAc-(1-->3)-beta-Ser.

Chaban, B., S. M. Logan, et al. (2009). "AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the N-glycan in Methanococcus voltae." J Bacteriol **191**(1): 187-95.

Recent advances in the field of prokaryotic N-glycosylation have established a foundation for the pathways and proteins involved in this important posttranslational protein modification process. To continue the study of the Methanococcus voltae N-glycosylation pathway, characteristics of known eukaryotic, bacterial, and archaeal proteins involved in the N-glycosylation process were examined and used to select candidate M. voltae genes for investigation as potential glycosyl transferase and flippase components. The targeted genes were knocked out via linear gene replacement, and the resulting effects on N-glycan assembly were identified through flagellin and surface (S) layer protein glycosylation defects. This study reports the finding that deletion of two putative M. voltae glycosyl transferase genes, designated aglC (for archaeal glycosylation) and aglK, interfered with proper N-glycosylation. This resulted in flagellin and S-layer proteins with significantly reduced apparent molecular masses, loss of flagellar assembly, and absence of glycan attachment. Given previous knowledge of both the N-glycosylation pathway in M. voltae and the general characteristics of N-glycosylation components, it appears that AglC and AglK are involved in the biosynthesis or transfer of diacetylated glucuronic acid within the glycan structure. In addition, a knockout of the putative flippase candidate gene (Mv891) had no effect on N-glycosylation but did result in the

production of giant cells with diameters three to four times that of wild-type cells.

Chaban, B., S. Voisin, et al. (2006). "Identification of genes involved in the biosynthesis and attachment of Methanococcus voltae N-linked glycans: insight into N-linked glycosylation pathways in Archaea." Mol Microbiol **61**(1): 259-68.

N-linked glycosylation is recognized as an important post-translational modification across all three domains of life. However, the understanding of the genetic pathways for the assembly and attachment of N-linked glycans in eukaryotic and bacterial systems far outweighs the knowledge of comparable processes in Archaea. The recent characterization of a novel trisaccharide [beta-ManpNAcA6Thr-(1-4)-beta-GlcpNAc3NAcA-(1-3)-beta-GlcpNAc]N-linked to asparagine residues in Methanococcus voltae flagellin and S-layer proteins affords new opportunities to investigate N-linked glycosylation pathways in Archaea. In this contribution, the insertional inactivation of several candidate genes within the M. voltae genome and their resulting effects on flagellin and S-layer glycosylation are reported. Two of the candidate genes were shown to have effects on flagellin and S-layer protein molecular mass and N-linked glycan structure. Further examination revealed inactivation of either of these two genes also had effects on flagella assembly. These genes, designated agl (archaeal glycosylation) genes, include a glycosyl transferase (aglA) involved in the attachment of the terminal sugar to the glycan and an STT3 oligosaccharyl transferase homologue (aglB) involved in the transfer of the complete glycan to the flagellin and S-layer proteins. These findings document the first experimental evidence for genes involved in any glycosylation process within the domain Archaea.

Chamot-Rooke, J., B. Rousseau, et al. (2007). "Alternative Neisseria spp. type IV pilin glycosylation with a glyceramido acetamido trideoxyhexose residue." Proc Natl Acad Sci U S A **104**(37): 14783-8.

The importance of protein glycosylation in the interaction of pathogenic bacteria with their host is becoming increasingly clear. Neisseria meningitidis, the etiological agent of cerebrospinal meningitis, crosses cellular barriers after adhering to host cells through type IV pili. Pilin glycosylation genes (pgl) are responsible for the glycosylation of PilE, the major subunit of type IV pili, with the 2,4-diacetamido-2,4,6-trideoxyhexose residue. Nearly half of the clinical isolates, however, display an insertion in the pglBCD operon, which is anticipated to lead to a different, unidentified glycosylation. Here the structure of pilin glycosylation was determined in such a strain by "top-down" MS approaches. MALDI-TOF, nanoelectrospray ionization Fourier transform ion cyclotron resonance, and nanoelectrospray ionization quadrupole TOF MS analysis of purified pili preparations originating from N. meningitidis strains, either wild type or deficient for pilin glycosylation, revealed a glycan mass inconsistent with 2,4-diacetamido-2,4,6-trideoxyhexose or any sugar in the databases. This unusual modification was determined by in-source dissociation of the sugar from the protein followed by tandem MS analysis with collision-induced fragmentation to be a hexose modified with a glyceramido and an acetamido group. We further show

genetically that the nature of the sugar present on the pilin is determined by the carboxyl-terminal region of the pglB gene modified by the insertion in the pglBCD locus. We thus report a previously undiscovered monosaccharide involved in posttranslational modification of type IV pilin subunits by a MS-based approach and determine the molecular basis of its biosynthesis.

Charbonneau, M. E., V. Girard, et al. (2007). "O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence." J Bacteriol **189**(24): 8880-9.

The Escherichia coli adhesin involved in diffuse adherence (AIDA-I) is one of the few glycosylated proteins found in Escherichia coli. Glycosylation is mediated by a specific heptosyltransferase encoded by the aah gene, but little is known about the role of this modification and the mechanism involved. In this study, we identified several peptides of AIDA-I modified by the addition of heptoses by use of mass spectrometry and N-terminal sequencing of proteolytic fragments of AIDA-I. One threonine and 15 serine residues were identified as bearing heptoses, thus demonstrating for the first time that AIDA-I is O-glycosylated. We observed that unglycosylated AIDA-I is expressed in smaller amounts than its glycosylated counterpart and shows extensive signs of degradation upon heat extraction. We also observed that unglycosylated AIDA-I is more sensitive to proteases and induces important extracytoplasmic stress. Lastly, as was previously shown, we noted that glycosylation is required for AIDA-I to mediate adhesion to cultured epithelial cells, but purified mature AIDA-I fused to GST was found to bind in vitro to cells whether or not it was glycosylated. Taken together, our results suggest that glycosylation is required to ensure a normal conformation of AIDA-I and may be only indirectly necessary for its cell-binding function.

Che, F. S., Y. Nakajima, et al. (2000). "Flagellin from an incompatible strain of Pseudomonas avenae induces a resistance response in cultured rice cells." J Biol Chem **275**(41): 32347-56.

The host range of Pseudomonas avenae is wide among monocotyledonous plants, but individual strains can infect only one or a few host species. The resistance response of rice cells to pathogens has been previously shown to be induced by a rice-incompatible strain, N1141, but not by a rice-compatible strain, H8301. To clarify the molecular mechanism of the host specificity in P. avenae, a strain-specific antibody that was raised against N1141 cells and then absorbed with H8301 cells was prepared. When a cell extract of strain N1141 was separated by SDS-polyacrylamide gel electrophoresis and immunostained with the N1141 strain-specific antibody, only a flagellin protein was detected. Purified N1141 flagellin induced the hypersensitive cell death in cultured rice cells within 6 h of treatment, whereas the H8301 flagellin did not. The hypersensitive cell death could be blocked by pretreatment with anti-N1141 flagellin antibody. Furthermore, a flagellin-deficient N1141 strain lost not only the induction ability of hypersensitive cell death but also the expression ability of the EL2 gene, which is thought to be one of the defense-related genes. These results demonstrated that

the resistance response in cultured rice cells is induced by the flagellin existing in the incompatible strain of P. avenae but not in the flagellin of the compatible strain.

Choi, K. J., S. Grass, et al. "The Actinobacillus pleuropneumoniae HMW1C-like glycosyltransferase mediates N-linked glycosylation of the Haemophilus influenzae HMW1 adhesin." PLoS One **5**(12): e15888.

The Haemophilus influenzae HMW1 adhesin is an important virulence exoprotein that is secreted via the two-partner secretion pathway and is glycosylated at multiple asparagine residues in consensus N-linked sequons. Unlike the heavily branched glycans found in eukaryotic N-linked glycoproteins, the modifying glycan structures in HMW1 are mono-hexoses or di-hexoses. Recent work demonstrated that the H. influenzae HMW1C protein is the glycosyltransferase responsible for transferring glucose and galactose to the acceptor sites of HMW1. An Actinobacillus pleuropneumoniae protein designated ApHMW1C shares high-level homology with HMW1C and has been assigned to the GT41 family, which otherwise contains only O-glycosyltransferases. In this study, we demonstrated that ApHMW1C has N-glycosyltransferase activity and is able to transfer glucose and galactose to known asparagine sites in HMW1. In addition, we found that ApHMW1C is able to complement a deficiency of HMW1C and mediate HMW1 glycosylation and adhesive activity in whole bacteria. Initial structure-function studies suggested that ApHMW1C consists of two domains, including a 15-kDa N-terminal domain and a 55-kDa C-terminal domain harboring glycosyltransferase activity. These findings suggest a new subfamily of HMW1C-like glycosyltransferases distinct from other GT41 family O-glycosyltransferases.

Christian, R., G. Schulz, et al. (1986). "Structure of a rhamnan from the surface-layer glycoprotein of Bacillus stearothermophilus strain NRS 2004/3a." Carbohydr Res **150**: 265-72.

The structure of a glycan from the surface-layer glycoprotein of Bacillus stearothermophilus strain NRS 2004/3a has been studied by 1H- and 13C-n.m.r. spectroscopy. The results indicate the glycan to be a polymer of the trisaccharide repeating-unit ----2)-alpha-L-Rhap-(1----2)-alpha-L-Rhap-(1----3)-beta-L-++ +Rhap-(1----.

Cohen-Krausz, S. and S. Trachtenberg (2002). "The structure of the archeabacterial flagellar filament of the extreme halophile Halobacterium salinarum R1M1 and its relation to eubacterial flagellar filaments and type IV pili." J Mol Biol **321**(3): 383-95.

Although the phenomenology and mechanics of swimming are very similar in eubacteria and archaeabacteria (e.g. reversible rotation, helical polymorphism of the filament and formation of bundles), the dynamic flagellar filaments seem completely unrelated in terms of morphogenesis, structure and amino acid composition. Archeabacterial flagellar filaments share important features with type IV pili, which are components of retractable linear motors involved in twitching motility and cell adhesion. The archeabacterial filament is unique in: (1) having a relatively smooth surface and a small diameter of approximately 100A

as compared to approximately 240A of eubacterial filaments and approximately 50A of type IV pili; (2) being glycosylated and sulfated in a pattern similar to the S-layer; (3) being synthesized as pre-flagellin with a signal-peptide cleavable by membrane peptidases upon transport; and (4) having an N terminus highly hydrophobic and homologous with that of the olygomerization domain of pilin. The synthesis of archeabacterial flagellin monomers as pre-flagellin and their post-translational, extracellular glycosylation suggest a different mode of monomer transport and polymerization at the cell-proximal end of the filament, similar to pili rather than to eubacterial flagellar filaments. The polymerization mode and small diameter may indicate the absence of a central channel in the filament. Using low-electron-dose images of cryo-negative-stained filaments, we determined the unique symmetry of the flagellar filament of the extreme halophile Halobacterium salinarum strain R1M1 and calculated a three-dimensional density map to a resolution of 19A. The map is based on layer-lines of order n=0, +10, -7, +3, -4, +6, and -1. The cross-section of the density map has a triskelion shape and is dominated by seven outer densities clustered into three groups, which are connected by lower-density arms to a dense central core surrounded by a lower-density shell. There is no evidence for a central channel. On the basis of the homology with the oligomerization domain of type IV pilin and the density distribution of the filament map, we propose a structure for the central core.

Comer, J. E., M. A. Marshall, et al. (2002). "Identification of the Pseudomonas aeruginosa 1244 pilin glycosylation site." Infect Immun **70**(6): 2837-45.

Previous work (P. Castric, F. J. Cassels, and R. W. Carlson, J. Biol. Chem. 276:26479-26485, 2001) has shown the Pseudomonas aeruginosa 1244 pilin glycan to be covalently bound to a serine residue. N-terminal sequencing of pilin fragments produced from endopeptidase treatment and identified by reaction with a glycan-specific monoclonal antibody indicated that the glycan was present between residue 75 and the pilin carboxy terminus. Further sequencing of these peptides revealed that serine residues 75, 81, 84, 105, 106, and 108 were not modified. Conversion of serine 148, but not serine 118, to alanine by site-directed mutagenesis, resulted in loss of the ability to carry out pilin glycosylation when tested in an in vivo system. These results showed the pilin glycan to be attached to residue 148, the carboxy-terminal amino acid. The carboxy-proximal portion of the pilin disulfide loop, which is adjacent to the pilin glycan, was found to be a major linear B-cell epitope, as determined by peptide epitope mapping analysis. Immunization of mice with pure pili produced antibodies that recognized the pilin glycan. These sera also reacted with P. aeruginosa 1244 lipopolysaccharide as measured by Western blotting and enzyme-linked immunosorbent assay.

Cooper, H. N., S. S. Gurcha, et al. (2002). "Characterization of mycobacterial protein glycosyltransferase activity using synthetic peptide acceptors in a cell-free assay." Glycobiology **12**(7): 427-34.

Synthetic peptides derived from a 45-kDa glycoprotein antigen of Mycobacterium tuberculosis were shown to function as glycosyltransferase acceptors for mannose residues in a mannosyltransferase cell-free assay. The

mannosyltransferase activity was localized within both isolated membranes and a P60 cell wall fraction prepared from the rapidly growing mycobacterial strain, Mycobacterium smegmatis. Incorporation of radiolabel from GDP-[(14)C]mannose was inhibited by the addition of amphomycin, indicating that the glycosyl donor for the peptide acceptors was a member of the mycobacterial polyprenol-P-mannose (PPM) family of activated glycosyl donors. Furthermore, a direct demonstration of transfer from the in situ generated PP[(14)C]Ms was also demonstrated. It was also found that the enzyme activity was sensitive to changes in overall peptide length and amino acid composition. Because glycoproteins are present on the mycobacterial cell surface and are available for interaction with host cells during infection, protein glycosyltransferases may provide novel drug targets. The development of a cell-free mannosyltransferase assay will now facilitate the cloning and biochemical characterisation of the relevant enzymes from M. tuberculosis.

Craig, L., N. Volkmann, et al. (2006). "Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions." Mol Cell **23**(5): 651-62.

Type IV pili (T4P) are long, thin, flexible filaments on bacteria that undergo assembly-disassembly from inner membrane pilin subunits and exhibit astonishing multifunctionality. Neisseria gonorrhoeae (gonococcal or GC) T4P are prototypic virulence factors and immune targets for increasingly antibiotic-resistant human pathogens, yet detailed structures are unavailable for any T4P. Here, we determined a detailed experimental GC-T4P structure by quantitative fitting of a 2.3 A full-length pilin crystal structure into a 12.5 A resolution native GC-T4P reconstruction solved by cryo-electron microscopy (cryo-EM) and iterative helical real space reconstruction. Spiraling three-helix bundles form the filament core, anchor the globular heads, and provide strength and flexibility. Protruding hypervariable loops and posttranslational modifications in the globular head shield conserved functional residues in pronounced grooves, creating a surprisingly corrugated pilus surface. These results clarify T4P multifunctionality and assembly-disassembly while suggesting unified assembly mechanisms for T4P, archaeal flagella, and type II secretion system filaments.

Davis, B. G., R. C. Lloyd, et al. (2000). "Controlled site-selective protein glycosylation for precise glycan structure-catalytic activity relationships." Bioorg Med Chem **8**(7): 1527-35.

Glycoproteins occur naturally as complex mixtures of differently glycosylated forms which are difficult to separate. To explore their individual properties, there is a need for homogeneous sources of carbohydrate-protein conjugates and this has recently prompted us to develop a novel method for the site-selective glycosylation of proteins. The potential of the method was illustrated by site-selective glycosylations of subtilisin Bacillus lentus (SBL) as a model protein. A representative library of mono- and disaccharide MTS reagents were synthesized from their parent carbohydrates and used to modify cysteine mutants of SBL at positions 62 in the S2 site, 156 and 166 in the S1 site and 217 in the S1' site.

These were the first examples of preparations of homogeneous neoglycoproteins in which both the site of glycosylation and structure of the introduced glycan were predetermined. The scope of this versatile method was expanded further through the combined use of peracetylated MTS reagents and careful pH adjustment to introduce glycans containing different numbers of acetate groups. This method provides a highly controlled and versatile route that is virtually unlimited in the scope of the sites and glycans that may be conjugated, and opens up hitherto inaccessible opportunities for the systematic determination of the properties of glycosylated proteins. This potential has been clearly demonstrated by the determination of detailed glycan structure-hydrolytic activity relationships for SBL. The 48 glycosylated CMMs formed display kcat/KM values that range from 1.1-fold higher than WT to 7-fold lower than WT. The anomeric stereochemistry of the glycans introduced modulates changes in kcat/KM upon acetylation. At positions 62 and 217 acetylation enhances the activity of alpha-glycosylated CMMs but decreases that of beta-glycosylated. This trend is reversed at position 166 where, in contrast, acetylation enhances the kcat/KMs of beta-glycosylated CMMs but decreases those of alpha-glycosylated. Consistent with its surface exposed nature changes at position 156 are more modest, but still allow control of activity, particularly through glycosylation with disaccharide lactose.

Davis, L. M., T. Kakuda, et al. (2009). "A Campylobacter jejuni znuA orthologue is essential for growth in low-zinc environments and chick colonization." J Bacteriol **191**(5): 1631-40.

Campylobacter jejuni infection is a leading cause of bacterial gastroenteritis in the United States and is acquired primarily through the ingestion of contaminated poultry products. Here, we describe the C. jejuni orthologue of ZnuA in other gram-negative bacteria. ZnuA (Cj0143c) is the periplasmic component of a putative zinc ABC transport system and is encoded on a zinc-dependent operon with Cj0142c and Cj0141c, which encode the other two likely components of the transport system of C. jejuni. Transcription of these genes is zinc dependent. A mutant lacking Cj0143c is growth deficient in zinc-limiting media, as well as in the chick gastrointestinal tract. The protein is glycosylated at asparagine 28, but this modification is dispensable for zinc-limited growth and chick colonization. Affinity-purified FLAG-tagged Cj0143c binds zinc in vitro. Based on our findings and on its homology to E. coli ZnuA, we conclude that Cj0143c encodes the C. jejuni orthologue of ZnuA.

de Vos, W. M., W. G. Voorhorst, et al. (2001). "Purification, characterization, and molecular modeling of pyrolysin and other extracellular thermostable serine proteases from hyperthermophilic microorganisms." Methods Enzymol **330**: 383-93.

Dell, A., A. Galadari, et al. "Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes." Int J Microbiol **2010**: 148178.

Recent years have witnessed a rapid growth in the number and diversity of prokaryotic proteins shown to carry N- and/or O-glycans, with protein glycosylation now considered as fundamental to the biology of these organisms

as it is in eukaryotic systems. This article overviews the major glycosylation pathways that are known to exist in eukarya, bacteria and archaea. These are (i) oligosaccharyltransferase (OST)-mediated N-glycosylation which is abundant in eukarya and archaea, but is restricted to a limited range of bacteria; (ii) stepwise cytoplasmic N-glycosylation that has so far only been confirmed in the bacterial domain; (iii) OST-mediated O-glycosylation which appears to be characteristic of bacteria; and (iv) stepwise O-glycosylation which is common in eukarya and bacteria. A key aim of the review is to integrate information from the three domains of life in order to highlight commonalities in glycosylation processes. We show how the OST-mediated N- and O-glycosylation pathways share cytoplasmic assembly of lipid-linked oligosaccharides, flipping across the ER/periplasmic/cytoplasmic membranes, and transferring "en bloc" to the protein acceptor. Moreover these hallmarks are mirrored in lipopolysaccharide biosynthesis. Like in eukaryotes, stepwise O-glycosylation occurs on diverse bacterial proteins including flagellins, adhesins, autotransporters and lipoproteins, with O-glycosylation chain extension often coupled with secretory mechanisms.

DiGiandomenico, A., M. J. Matewish, et al. (2002). "Glycosylation of Pseudomonas aeruginosa 1244 pilin: glycan substrate specificity." Mol Microbiol **46**(2): 519-30.

The structural similarity between the pilin glycan and the O-antigen of Pseudomonas aeruginosa 1244 suggested that they have a common metabolic origin. Mutants of this organism lacking functional wbpM or wbpL genes synthesized no O-antigen and produced only non-glycosylated pilin. Complementation with plasmids containing functional wbpM or wbpL genes fully restored the ability to produce both O-antigen and glycosylated pilin. Expression of a cosmid clone containing the O-antigen biosynthetic gene cluster from P. aeruginosa PA103 (LPS serotype O11) in P. aeruginosa 1244 (LPS serotype O7) resulted in the production of strain 1244 pili that contained both O7 and O11 antigens. The presence of the O11 repeating unit was confirmed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Expression of the O-antigen biosynthesis cluster from Escherichia coli O157:H7 in strain 1244 resulted in the production of pilin that contained both the endogenous Pseudomonas as well as the Escherichia O157 O-antigens. A role for pilO in the glycosylation of pilin in P. aeruginosa is evident as the cloned pilAO operon produced glycosylated strain 1244 pilin in eight heterologous P. aeruginosa strains. Removal of the pilO gene resulted in the production of unmodified strain 1244 pilin. These results show that the pilin glycan of P. aeruginosa 1244 is a product of the O-antigen biosynthetic pathway. In addition, the structural diversity of the O-antigens used by the 1244 pilin glycosylation apparatus indicates that the glycan substrate specificity of this reaction is extremely low.

Dobos, K. M., K. H. Khoo, et al. (1996). "Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of Mycobacterium tuberculosis." J Bacteriol **178**(9): 2498-506.

Chemical evidence for the true glycosylation of mycobacterial proteins was recently provided in the context of the 45-kDa MPT 32 secreted protein of Mycobacterium tuberculosis (K. Dobos, K. Swiderek, K.-H. Khoo, P. J. Brennan, and J. T. Belisle, Infect. Immun. 63:2846-2853, 1995). However, the full extent and nature of glycosylation as well as the location of glycosylated amino acids remained undefined. First, to examine the nature of the covalently attached sugars, the 45-kDa protein was obtained from cells metabolically labeled with D-[U-14C] glucose and subjected to compositional analysis, which revealed mannose as the only covalently bound sugar. Digestion of the protein with the endoproteinase subtilisin and analysis of products by liquid chromatography-electrospray-mass spectrometry on the basis of fragments demonstrating neutral losses of hexose (m/z 162) or pentose (m/z 132) revealed five glycopeptides, S7, S18, S22, S29, and S41 among a total of 50 peptides, all of which produced only m/z 162 fragmentation ion deletions. Fast atom bombardment-mass spectrometry, N-terminal amino acid sequencing, and alpha-mannosidase digestion demonstrated universal O glycosylation of Thr residues with a single alpha-D-Man, mannobiose, or mannotriose unit. Linkages within the mannobiose and mannotriose were all alpha 1-2, as proven by gas chromatography-mass spectrometry of oligosaccharides released by beta-elimination. Total sequences of many of the glycosylated and nonglycosylated peptides combined with published information on the deduced amino acid sequence of the entire 45-kDa protein demonstrated that the sites of glycosylation were located in Pro-rich domains near the N terminus and C terminus of the polypeptide backbone. Specifically, the Thr residues at positions 10 and 18 were substituted with alpha-D-Manp(1-->2)alpha-D-Manp, the Thr residue at position 27 was substituted with a single alpha-D-Manp, and Thr-277 was substituted with either alpha-D-Manp, alpha-D-Manp(1-->2)alpha-D-Manp, or alpha-D-Manp(1--> 2)alpha-D-Manp(1-->2)alpha-D-Manp. This report further corroborates the existence of true prokaryotic glycoproteins, defines the complete structure of a mycobacterial mannoprotein and the first complete structure of a mannosylated mycobacterial protein, and establishes the principles for the study of other mycobacterial glycoproteins.

Dobos, K. M., K. Swiderek, et al. (1995). "Evidence for glycosylation sites on the 45-kilodalton glycoprotein of Mycobacterium tuberculosis." Infect Immun **63**(8): 2846-53.

The occurrence of glycosylated proteins in Mycobacterium tuberculosis has been widely reported. However, unequivocal proof for the presence of true glycosylated amino acids within these proteins has not been demonstrated, and such evidence is essential because of the predominance of soluble lipoglycans and glycolipids in all mycobacterial extracts. We have confirmed the presence of several putative glycoproteins in subcellular fractions of M. tuberculosis by reaction with the lectin concanavalin A. One such product, with a molecular mass of 45 kDa, was purified from the culture filtrate. Compositional analysis demonstrated that the protein was rich in proline and that mannose, galactose, glucose, and arabinose together represented about 4% of the total mass. The 45-kDa glycoprotein was subjected to proteolytic digestion with either the Asp-N

or the Glu-C endopeptidase or subtilisin, peptides were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and glycopeptides were identified by reaction with concanavalin A. Peptides were further separated, and when they were analyzed by liquid chromatography-electrospray mass spectrometry for neutral losses of hexoses (162 mass units), four peptides were identified, indicating that these were glycosylated with hexose residues. One peptide, with an average molecular mass of 1,516 atomic mass units (AMU), exhibited a loss of two hexose units. The N-terminal sequence of the 1,516-AMU glycopeptide was determined to be DPEPAPPVP, which was identical to the sequence of the amino terminus of the mature protein, DPEPAP PVPXTA. Furthermore, analysis of the glycopeptide by secondary ion mass spectrometry demonstrated that the complete sequence of the glycopeptide was DPEPAPPVPTTA. From this, it was determined that the 10th amino acid, threonine, was O-glycosidically linked to a disaccharide composed of two hexose residues, probably mannose. This report establishes that true, O-glycosylated proteins exist in mycobacteria.

Doig, P., N. Kinsella, et al. (1996). "Characterization of a post-translational modification of Campylobacter flagellin: identification of a sero-specific glycosyl moiety." Mol Microbiol **19**(2): 379-87.

The flagellins of Campylobacter spp. differ antigenically. In variants of C. coli strain VC167, two antigenic flagellin types determined by sero-specific antibodies have been described (termed T1 and T2). Post-translational modification has been suggested to be responsible for T1 and T2 epitopes, and, using mild periodate treatment and biotin hydrazide labelling, flagellin from both VC167-T1 and T2 were shown to be glycosylated. Glycosylation was also shown to be present on other Campylobacter flagellins. The ability to label all Campylobacter flagellins examined with the lectin LFA demonstrated the presence of a terminal sialic acid moiety. Furthermore, mild periodate treatment of the flagellins of VC167 eliminated reactivity with T1 and T2 specific antibodies LAH1 and LAH2, respectively, and LFA could also compete with LAH1 and LAH2 antibodies for binding to their respective flagellins. These data implicate terminal sialic acid as part of the LAH strain-specific epitopes. However, using mutants in genes affecting LAH serorecognition of flagellin it was demonstrated that sialic acid alone is not the LAH epitope. Rather, the epitope(s) is complex, probably involving multiple glycosyl and/or amino acid residues.

Eichler, J. (2000). "Novel glycoproteins of the halophilic archaeon Haloferax volcanii." Arch Microbiol **173**(5-6): 445-8.

Archaea possess many eukaryote-like properties, including the ability to glycosylate proteins. Using oligosaccharide staining and lectin binding, this study revealed the existence of several glycosylated Haloferax volcanii membrane proteins, besides the previously reported surface layer (S-layer) glycoprotein. While the presence of glycoproteins in archaeal S-layers and flagella is well-documented, few archaeal glycoproteins that are not part of these structures have been reported. The glycosylated 150, 98, 58 and 54 kDa protein species

detected were neither precursors nor breakdown products of the 190 kDa S-layer glycoprotein. Furthermore, these novel glycoproteins were outwardly oriented and intimately associated with the membrane.

Eichler, J. and M. W. Adams (2005). "Posttranslational protein modification in Archaea." Microbiol Mol Biol Rev **69**(3): 393-425.

One of the first hurdles to be negotiated in the postgenomic era involves the description of the entire protein content of the cell, the proteome. Such efforts are presently complicated by the various posttranslational modifications that proteins can experience, including glycosylation, lipid attachment, phosphorylation, methylation, disulfide bond formation, and proteolytic cleavage. Whereas these and other posttranslational protein modifications have been well characterized in Eucarya and Bacteria, posttranslational modification in Archaea has received far less attention. Although archaeal proteins can undergo posttranslational modifications reminiscent of what their eucaryal and bacterial counterparts experience, examination of archaeal posttranslational modification often reveals aspects not previously observed in the other two domains of life. In some cases, posttranslational modification allows a protein to survive the extreme conditions often encountered by Archaea. The various posttranslational modifications experienced by archaeal proteins, the molecular steps leading to these modifications, and the role played by posttranslational modification in Archaea form the focus of this review.

Espitia, C., R. Espinosa, et al. (1995). "Antigenic and structural similarities between Mycobacterium tuberculosis 50- to 55-kilodalton and Mycobacterium bovis BCG 45- to 47-kilodalton antigens." Infect Immun **63**(2): 580-4.

The relationship between Mycobacterium tuberculosis 50- to 55-kDa protein and Mycobacterium bovis BCG 45- to 47-kDa antigen was examined by using immunological and biochemical criteria. Reciprocal cross-reactivity with a rabbit polyclonal antiserum against the M. bovis BCG protein and with a monoclonal antibody raised against the M. tuberculosis antigen was observed. The epitope recognized by this antibody was apparently present only in proteins of M. tuberculosis and M. bovis BCG among the 11 mycobacterial species tested. The amino-terminal sequences and total amino acid contents of these proteins showed strong similarities. Both antigens are glycoproteins as assessed by binding of concanavalin A, labeling of carbohydrate moieties with biotin-hydrazide, and digestion of carbohydrates with jack bean alpha-D-mannosidase, which produced a reduction of the molecular weights of the proteins and totally eliminated concanavalin A binding. Both M. tuberculosis and M. bovis BCG proteins are secreted, since they were found mainly in the culture medium. Analysis of M. tuberculosis 50- to 55-kDa antigen by two-dimensional gel electrophoresis showed at least seven different components, as previously described for the M. bovis BCG antigen. Solid-phase immunoassays showed that the purified M. tuberculosis 50- to 55-kDa protein was recognized by serum specimens from 70% of individuals with pulmonary tuberculosis from a total of 77 Mexican patients examined.

Espitia, C. and R. Mancilla (1989). "Identification, isolation and partial characterization of Mycobacterium tuberculosis glycoprotein antigens." Clin Exp Immunol **77**(3): 378-83.

In Mycobacterium tuberculosis culture filtrates, three concanavalin A (ConA)-binding bands of 55, 50 and 38 kilodaltons (kD) were identified by labelling blotted proteins with a ConA-peroxidase conjugate. Binding was inhibited by the competitor sugar alpha-methyl mannoside and by reduction with sodium m-periodate. Bands of 55, 50 and 38 kD stained with Coomasie blue were sensitive to digestion with proteases, thus indicating that they are proteins. Glycoproteins were isolated by lectin affinity chromatography or by elution from nitrocellulose membranes. On the isolated form, the 55-50 kD doublet glycoprotein was 65.4% protein and 34.6% sugar. The purified 38 kD molecule was 74.3% protein and 25.7% carbohydrate. By immunoblot, antibodies against mycobacterial glycoproteins were demonstrated in immunized rabbits and in patients with pulmonary tuberculosis, but not in healthy individuals. Treatment with sodium m-periodate abolished binding of rabbit antibodies to the 38 kD glycoprotein. Reactivity of the 55-50 kD doublet glycoprotein was not altered by reduction. By immunoblot with monoclonal antibodies TB71 and TB72, a carbohydrate-dependent and a carbohydrate-independent epitope could be identified on the 38 kD glycoprotein.

Espitia, C., L. Servin-Gonzalez, et al. "New insights into protein O-mannosylation in actinomycetes." Mol Biosyst **6**(5): 775-81.

Glycosylation is a common post-translational modification of surface exposed proteins and lipids present in all kingdoms of life. Information derived from bacterial genome sequencing, together with proteomic and genomic analysis has allowed the identification of the enzymatic glycosylation machinery. Among prokaryotes, O-mannosylation of proteins has been found in the actinomycetes and resembles protein O-mannosylation in fungi and higher eukaryotes. In this review we summarize the main features of the biosynthetic pathway of O-mannosylation in prokaryotes with special emphasis on the actinomycetes, as well as the biological role of the glycosylated target proteins.

Ewing, C. P., E. Andreishcheva, et al. (2009). "Functional characterization of flagellin glycosylation in Campylobacter jejuni 81-176." J Bacteriol **191**(22): 7086-93.

The major flagellin of Campylobacter jejuni strain 81-176, FlaA, has been shown to be glycosylated at 19 serine or threonine sites, and this glycosylation is required for flagellar filament formation. Some enzymatic components of the glycosylation machinery of C. jejuni 81-176 are localized to the poles of the cell in an FlhF-independent manner. Flagellin glycosylation could be detected in flagellar mutants at multiple levels of the regulatory hierarchy, indicating that glycosylation occurs independently of the flagellar regulon. Mutants were constructed in which each of the 19 serine or threonines that are glycosylated in FlaA was converted to an alanine. Eleven of the 19 mutants displayed no observable phenotype, but the remaining 8 mutants had two distinct phenotypes. Five mutants (mutations S417A, S436A, S440A, S457A, and T481A) were fully

motile but defective in autoagglutination (AAG). Three other mutants (mutations S425A, S454A, and S460A) were reduced in motility and synthesized truncated flagellar filaments. The data implicate certain glycans in mediating filament-filament interactions resulting in AAG and other glycans appear to be critical for structural subunit-subunit interactions within the filament.

Fethiere, J., B. Eggimann, et al. (1999). "Crystal structure of chondroitin AC lyase, a representative of a family of glycosaminoglycan degrading enzymes." J Mol Biol **288**(4): 635-47.

Glycosaminoglycans (GAGs), highly sulfated polymers built of hexosamine-uronic acid disaccharide units, are major components of the extracellular matrix, mostly in the form of proteoglycans. They interact with a large array of proteins, in particular of the blood coagulation cascade. Degradation of GAGs in mammalian systems occurs by the action of GAG hydrolases. Bacteria express a large number of GAG-degrading lyases that break the hexosamine-uronic acid bond to create an unsaturated sugar ring. Flavobacterium heparinum produces at least five GAG lyases of different specificity. Chondroitin AC lyase (chondroitinase AC, 75 kDa) is highly active toward chondroitin 4-sulfate and chondroitin-6 sulfate. Its crystal structure has been determined to 1.9 A resolution. The enzyme is composed of two domains. The N-terminal domain of approximately 300 residues contains mostly alpha-helices which form a doubly-layered horseshoe (a subset of the (alpha/alpha)6 toroidal topology). The approximately 370 residues long C-terminal domain is made of beta-strands arranged in a four layered beta-sheet sandwich, with the first two sheets having nine strands each. This fold is novel and has no counterpart in full among known structures. The sequence of chondroitinase AC shows low level of homology to several hyaluronate lyases, which likely share its fold. The shape of the molecule, distribution of electrostatic potential, the pattern of conservation of the amino acids and the results of mutagenesis of hyaluronate lyases, indicate that the enzymatic activity resides primarily within the N-terminal domain. The most likely candidate for the catalytic base is His225. Other residues involved in catalysis and/or substrate binding are Arg288, Arg292, Lys298 and Lys299.

Fifis, T., C. Costopoulos, et al. (1991). "Purification and characterization of major antigens from a Mycobacterium bovis culture filtrate." Infect Immun **59**(3): 800-7.

Ten major antigens from Mycobacterium bovis culture filtrate of 39, 32, 30, 25, 24, 22 (a and b forms), 19, 15, and 12 kDa have been purified and characterized by classical physicochemical methods. With monoclonal antibodies and/or N-terminal amino acid sequencing data, it was found that the antigens of 32, 30, 24, 22 (a), 19, and 12 kDa are related to M. bovis or M. tuberculosis antigens P32, MPB59, MPB64, MPB70, 19 kDa, and 12 kDa, respectively. The 39-, 25-, 22 (b)-, and 19-kDa antigens showed concanavalin A-binding properties and were positive in a glycan detection test, suggesting that they are glycoproteins. The 25- and 22 (b)-kDa proteins were found to be glycosylated forms of MPB70.

Fletcher, C. M., M. J. Coyne, et al. "Theoretical and experimental characterization of the

scope of protein O-glycosylation in Bacteroides fragilis." J Biol Chem **286**(5): 3219-26.

Among bacterial species demonstrated to have protein O-glycosylation systems, that of Bacteroides fragilis and related species is unique in that extracytoplasmic proteins are glycosylated at serine or threonine residues within the specific three-amino acid motif D(S/T)(A/I/L/M/T/V). This feature allows for computational analysis of the proteome to identify candidate glycoproteins. With the criteria of a signal peptidase I or II cleavage site or a predicted transmembrane-spanning region and the presence of at least one glycosylation motif, we identified 1021 candidate glycoproteins of B. fragilis. In addition to the eight glycoproteins identified previously, we confirmed that another 12 candidate glycoproteins are in fact glycosylated. These included four glycoproteins that are predicted to localize to the inner membrane, a compartment not previously shown to include glycosylated proteins. In addition, we show that four proteins involved in cell division and chromosomal segregation, two of which are encoded by candidate essential genes, are glycosylated. To date, we have not identified any extracytoplasmic proteins containing a glycosylation motif that are not glycosylated. Therefore, based on the list of 1021 candidate glycoproteins, it is likely that hundreds of proteins, comprising more than half of the extracytoplasmic proteins of B. fragilis, are glycosylated. Site-directed mutagenesis of several glycoproteins demonstrated that all are glycosylated at the identified glycosylation motif. By engineering glycosylation motifs into a naturally unglycosylated protein, we are able to bring about site-specific glycosylation at the engineered sites, suggesting that this glycosylation system may have applications for glycoengineering.

Fletcher, C. M., M. J. Coyne, et al. (2009). "A general O-glycosylation system important to the physiology of a major human intestinal symbiont." Cell **137**(2): 321-31.

The Bacteroides are a numerically dominant genus of the human intestinal microbiota. These organisms harbor a rare bacterial pathway for incorporation of exogenous fucose into capsular polysaccharides and glycoproteins. The infrequency of glycoprotein synthesis by bacteria prompted a more detailed analysis of this process. Here, we demonstrate that Bacteroides fragilis has a general O-glycosylation system. The proteins targeted for glycosylation include those predicted to be involved in protein folding, protein-protein interactions, peptide degradation as well as surface lipoproteins. Protein glycosylation is central to the physiology of B. fragilis and is necessary for the organism to competitively colonize the mammalian intestine. We provide evidence that general O-glycosylation systems are conserved among intestinal Bacteroides species and likely contribute to the predominance of Bacteroides in the human intestine.

Forest, K. T., S. A. Dunham, et al. (1999). "Crystallographic structure reveals phosphorylated pilin from Neisseria: phosphoserine sites modify type IV pilus surface chemistry and fibre morphology." Mol Microbiol **31**(3): 743-52.

Understanding the structural biology of type IV pili, fibres responsible for the virulent attachment and motility of numerous bacterial pathogens, requires a

detailed understanding of the three-dimensional structure and chemistry of the constituent pilin subunit. X-ray crystallographic refinement of Neisseria gonorrhoeae pilin against diffraction data to 2.6 A resolution, coupled with mass spectrometry of peptide fragments, reveals phosphoserine at residue 68. Phosphoserine is exposed on the surface of the modelled type IV pilus at the interface of neighbouring pilin molecules. The site-specific mutation of serine 68 to alanine showed that the loss of the phosphorylation alters the morphology of fibres examined by electron microscopy without a notable effect on adhesion, transformation, piliation or twitching motility. The structural and chemical characterization of protein phosphoserine in type IV pilin subunits is an important indication that this modification, key to numerous regulatory aspects of eukaryotic cell biology, exists in the virulence factor proteins of bacterial pathogens. These O-linked phosphate modifications, unusual in prokaryotes, thus merit study for possible roles in pilus biogenesis and modulation of pilin chemistry for optimal in vivo function.

Gerl, L., R. Deutzmann, et al. (1989). "Halobacterial flagellins are encoded by a multigene family. Identification of all five gene products." FEBS Lett **244**(1): 137-40.

Flagellins of Halobacterium halobium are encoded in five different but homologous genes. Flagellins isolated from purified flagella were digested and the resulting peptides sequenced. The amino acid sequence data obtained prove that all five gene products are expressed and integrated into the flagellar bundle.

Gerl, L. and M. Sumper (1988). "Halobacterial flagellins are encoded by a multigene family. Characterization of five flagellin genes." J Biol Chem **263**(26): 13246-51.

Purified flagellar filaments of Halobacterium halobium contain three different protein species based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. These proteins were designated as flagellins Fla I, Fla II, and Fla III and were characterized as sulfated glycoproteins with N-glycosidically linked oligosaccharides of the type GlcA-(1----4)-GlcA-(1----4)-GlcA-(1----4)-Glc. All halobacterial flagellin polypeptides are immunologically cross-reactive. A gene fragment of one flagellin was isolated in an expression vector using antibody probes. Using this gene fragment as probe, we identified, subcloned, and determined the nucleotide sequences of five different but highly homologous flagellin genes. Two flagellin (flg) genes are arranged tandemly at one locus (flg A1 and -2), and the other three in a tandem arrangement at a different locus (flg B1, -2, and -3), Two flg mRNAs were detected, one from the A genes and the other from the B genes. Based on immunological analysis, the products of the flg A1 and A2 are Fla II and Fla I, respectively.

Glover, K. J., E. Weerapana, et al. (2005). "Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni." Chem Biol **12**(12): 1311-5.

The gram-negative bacterium Campylobacter jejuni has a general N-linked glycosylation pathway encoded by the pgl gene cluster. One of the proteins in this cluster, PglB, is thought to be the oligosaccharyl transferase due to its

significant homology to Stt3p, a subunit of the yeast oligosaccharyl transferase complex. PglB has been shown to be involved in catalyzing the transfer of an undecaprenyl-linked heptasaccharide to the asparagine side chain of proteins at the Asn-X-Ser/Thr motif. Using a synthetic disaccharide glycan donor (GalNAc-alpha1,3-bacillosamine-pyrophosphate-undecaprenyl) and a peptide acceptor substrate (KDFNVSKA), we can observe the oligosaccharyl transferase activity of PglB in vitro. Furthermore, the preparation of additional undecaprenyl-linked glycan variants reveals the ability of PglB to transfer a wide variety of saccharides. With the demonstration of PglB activity in vitro, fundamental questions surrounding the mechanism of N-linked glycosylation can now be addressed.

Godavarti, R. and R. Sasisekharan (1996). "A comparative analysis of the primary sequences and characteristics of heparinases I, II, and III from Flavobacterium heparinum." Biochem Biophys Res Commun **229**(3): 770-7.

Heparinases I, II and III from F. heparinum cleave heparin-like molecules, with a high degree of substrate specificity, at the glucosamine-uronate linkage by elimination, leaving an unsaturated C4-C5 bond in the uronic acid. The primary sequences of these enzymes have been reported earlier. In this study we perform a comparative analysis of the properties and primary sequences of heparinase I, II and III. Alignment of the primary sequences revealed little sequence homology (15% residue identity in a LALIGN alignment) at both DNA and amino acid levels. There are three basic clusters in heparinase II satisfying the heparin binding consensus sequence with one of the sequences sharing homology with a consensus sequence in the heparin binding site of heparinase I and two basic clusters in heparinase III. Similar to heparinase I, there are two putative 'EF-hand' calcium coordinating motifs in heparinase III, while heparinase II does not contain any such motifs. Recombinant heparinases II and III's degradation of the substrate and the subsequent separation of the oligosaccharide products by POROS anion exchange chromatography were identical to those obtained from native heparinases II and III from F. heparinum.

Goon, S., J. F. Kelly, et al. (2003). "Pseudaminic acid, the major modification on Campylobacter flagellin, is synthesized via the Cj1293 gene." Mol Microbiol **50**(2): 659-71.

Flagellins from Campylobacter jejuni 81-176 and Campylobacter coli VC167 are heavily glycosylated. The major modifications on both flagellins are pseudaminic acid (Pse5Ac7Ac), a nine carbon sugar that is similar to sialic acid, and an acetamidino-substituted analogue of pseudaminic acid (PseAm). Previous data have indicated that PseAm is synthesized via Pse5Ac7Ac in C. jejuni 81-176, but that the two sugars are synthesized using independent pathways in C. coli VC167. The Cj1293 gene of C. jejuni encodes a putative UDP-GlcNAc C6-dehydratase/C4-reductase that is similar to a protein required for glycosylation of Caulobacter crescentus flagellin. The Cj1293 gene is expressed either under the control of a sigma 54 promoter that overlaps the coding region of Cj1292 or as a polycistronic message under the control of a sigma 70 promoter upstream of

Cj1292. A mutant in gene Cj1293 in C. jejuni 81-176 was non-motile and non-flagellated and accumulated unglycosylated flagellin intracellularly. This mutant was complemented in trans with the homologous C. jejuni gene, as well as the Helicobacter pylori homologue, HP0840, which has been shown to encode a protein with UDP-GlcNAc C6-dehydratase/C4-reductase activity. Mutation of Cj1293 in C. coli VC167 resulted in a fully motile strain that synthesized a flagella filament composed of flagellin in which Pse5Ac7Ac was replaced by PseAm. The filament from the C. coli Cj1293 mutant displayed increased solubility in SDS compared with the wild-type filament. A double mutant in C. coli VC167, defective in both Cj1293 and ptmD, encoding part of the independent PseAm pathway, was also non-motile and non-flagellated and accumulated unglycosylated flagellin intracellularly. Collectively, the data indicate that Cj1293 is essential for Pse5Ac7Ac biosynthesis from UDP-GlcNAc, and that glycosylation is required for flagella biogenesis in campylobacters.

Grass, S., A. Z. Buscher, et al. (2003). "The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis." Mol Microbiol **48**(3): 737-51.

Non-typeable Haemophilus influenzae is a common respiratory pathogen and an important cause of morbidity in humans. The non-typeable H. influenzae HMW1 and HMW2 adhesins are related proteins that mediate attachment to human epithelial cells, an essential step in the pathogenesis of disease. Secretion of these adhesins requires accessory proteins called HMW1B/HMW2B and HMW1C/HMW2C. In the present study, we investigated the specific function of HMW1C. Examination of mutant constructs demonstrated that HMW1C influences both the size and the secretion of HMW1. Co-immunoprecipitation and yeast two-hybrid assays revealed that HMW1C interacts with HMW1 and forms a complex in the cytoplasm. Additional experiments and homology analysis established that HMW1C is required for glycosylation of HMW1 and may have glycotransferase activity. The glycan structure contains galactose, glucose and mannose and appears to be generated in part by phosphoglucomutase, an enzyme important for lipooligosaccharide biosynthesis. In the absence of glycosylation, HMW1 is partially degraded and is efficiently released from the surface of the organism, resulting in reduced adherence. Based on these results, we conclude that glycosylation is a prerequisite for HMW1 stability. In addition, glycosylation appears to be essential for optimal HMW1 tethering to the bacterial surface, which in turn is required for HMW1-mediated adherence, thus revealing a novel mechanism by which glycosylation influences cell-cell interactions.

Grass, S., C. F. Lichti, et al. "The Haemophilus influenzae HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin." PLoS Pathog **6**(5): e1000919.

The Haemophilus influenzae HMW1 adhesin is a high-molecular weight protein that is secreted by the bacterial two-partner secretion pathway and mediates adherence to respiratory epithelium, an essential early step in the pathogenesis of H. influenzae disease. In recent work, we discovered that HMW1 is a

glycoprotein and undergoes N-linked glycosylation at multiple asparagine residues with simple hexose units rather than N-acetylated hexose units, revealing an unusual N-glycosidic linkage and suggesting a new glycosyltransferase activity. Glycosylation protects HMW1 against premature degradation during the process of secretion and facilitates HMW1 tethering to the bacterial surface, a prerequisite for HMW1-mediated adherence. In the current study, we establish that the enzyme responsible for glycosylation of HMW1 is a protein called HMW1C, which is encoded by the hmw1 gene cluster and shares homology with a group of bacterial proteins that are generally associated with two-partner secretion systems. In addition, we demonstrate that HMW1C is capable of transferring glucose and galactose to HMW1 and is also able to generate hexose-hexose bonds. Our results define a new family of bacterial glycosyltransferases.

Graycar, T., M. Knapp, et al. (1999). "Engineered Bacillus lentus subtilisins having altered flexibility." J Mol Biol **292**(1): 97-109.
The three-dimensional structures of engineered variants of Bacillus lentus subtilisin having increased enzymatic activity, K27R/N87S/V104Y/N123S/T274A (RSYSA) and N76D/N87S/S103A/V104I (DSAI), were determined by X-ray crystallography. In addition to identifying changes in atomic position we report a method that identifies protein segments having altered flexibility. The method utilizes a statistical analysis of variance to delineate main-chain temperature factors that represent significant departures from the overall variance between equivalent regions seen throughout the structure. This method reveals changes in main-chain mobility in both variants. Residues 125-127 have increased mobility in the RSYSA variant while residues 100-104 have decreased mobility in the DSAI variant. These segments are located at the substrate-binding site and changes in their mobility are believed to relate to the observed changes in proteolytic activity. The effect of altered crystal lattice contacts on segment flexibility becomes apparent when identical variants, determined in two crystal forms, are compared with the native enzyme.

Grogan, D. W. (1989). "Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains." J Bacteriol **171**(12): 6710-9.
Though amenable to routine manipulation and a popular subject of molecular genetic and biochemical studies on archaebacteria, the genus Sulfolobus has remained poorly described in phenotypic terms. To delineate their physiological capabilities and diversity, five laboratory strains, including type strains of the described species Sulfolobus acidocaldarius and S. solfataricus, were compared with respect to a variety of growth and biochemical parameters, including component profile of the surface-layer cell wall, inhibitors of growth, growth rate as a function of temperature and pH, and compounds used as sole sources of carbon or nitrogen. Motility and photoregulated production of an orange pigment were detected in all five strains tested. The results provide new criteria for distinguishing Sulfolobus strains as well as potential tools for the physiological and genetic manipulation of these extreme thermophiles.

Gross, J., S. Grass, et al. (2008). "The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification." J Biol Chem **283**(38): 26010-5.

The Haemophilus influenzae HMW1 adhesin mediates adherence to respiratory epithelial cells, a critical early step in the pathogenesis of H. influenzae disease. In recent work, we demonstrated that HMW1 undergoes glycosylation. In addition, we observed that glycosylation of HMW1 is essential for HMW1 tethering to the bacterial surface, a prerequisite for HMW1-mediated adherence to host epithelium. In this study, we examined HMW1 proteolytic fragments by mass spectrometry, achieved 89% amino acid sequence coverage, and identified 31 novel modification sites. All of the modified sites were asparagine residues, in all but one case in the conventional consensus sequence of N-linked glycans, viz. NX(S/T). Liquid chromatography-tandem mass spectrometry analysis using a hybrid linear quadrupole ion trap Fourier transform ion cyclotron mass spectrometer, accurate mass measurements, and deuterium exchange studies established that the modifying glycan structures were mono- or dihexoses rather than the N-acetylated chitobiosyl core that is characteristic of N-glycosylation. This unusual carbohydrate modification suggests that HMW1 glycosylation requires a glycosyltransferase with a novel activity.

Guerry, P., P. Doig, et al. (1996). "Identification and characterization of genes required for post-translational modification of Campylobacter coli VC167 flagellin." Mol Microbiol **19**(2): 369-78.

Two genes have been identified in Campylobacter coli VC167 which are required for the biosynthesis of post-translational modifications on flagellin proteins. The ptmA gene encodes a protein of predicted M(r) 28,486 which shows significant homology to a family of alcohol dehydrogenases from a variety of bacteria. The ptmB gene encodes a protein of predicted M(r) 26,598 with significant homology to CMP-N-acetylneuraminic acid synthetase enzymes involved in sialic acid capsular biosynthesis in Neisseria meninigitidis and Escherichia coli K1. Site-specific mutation of either ptmA or ptmB caused loss of reactivity with antisera specific to the post-translational modifications and a change in the isoelectric focusing fingerprints relative to the parent strains. Mutation of ptmB, but not of ptmA, caused a change in apparent M(r) of the flagellin subunit in SDS-PAGE gels. The ptmA and ptmB genes are present in other strains of Campylobacter. In a rabbit model the ptmA mutant showed a reduced ability to elicit protection against subsequent challenge with heterologous strains of the same Lior serotype compared to the parental wild-type strain. This suggests that the surface-exposed post-translational modifications may play a significant role in the protective immune response.

Hahn, M., T. Keitel, et al. (1995). "Crystal and molecular structure at 0.16-nm resolution of the hybrid Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-M)." Eur J Biochem **232**(3): 849-58.

H(A16-M) is a hybrid endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase from

Bacillus. Its crystal structure was refined using synchrotron X-ray diffraction data up to a maximal resolution of 0.16 nm. The R value of the resulting model is 14.3% against 21,032 reflections > 2 sigma. 93% of the amino acid residues are in the most favorable regions of the Ramachandran diagram, and geometrical parameters are in accordance with other proteins solved at high resolution. As shown earlier [Keitel, T., Simon, O., Borriss, R. & Heinemann, U. (1993) Proc. Natl Acad. Sci. USA 90, 5287-5291], the protein folds into a compact jellyroll-type beta-sheet structure. A systematic analysis of the secondary structure reveals the presence of two major antiparallel beta-sheets and a three-stranded minor mixed sheet. Amino acid residues involved in catalysis and substrate binding are located inside a deep channel spanning the surface of the protein. To investigate the stereochemical cause of the observed specificity of endo-1,3-1,4-beta-D-glucan 4-glucanohydrolases towards beta-1,4 glycosyl bonds adjacent to beta-1,3 bonds, the high-resolution crystal structure has been used to model an enzyme-substrate complex. It is proposed that productive substrate binding to the subsites p1, p2 and p3 of H(A16-M) requires a beta-1,3 linkage between glucose units bound to p1 and p2.

Hahn, M., K. Piotukh, et al. (1994). "Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis." Proc Natl Acad Sci U S A **91**(22): 10417-21.

A jellyroll beta-sandwich protein, the Bacillus beta-glucanase H(A16-M), is used to probe the role of N-terminal peptide regions in protein folding in vivo. A gene encoding H(A16-M) is rearranged to place residues 1-58 of the protein behind a signal peptide and residues 59-214. The rearranged gene is expressed in Escherichia coli. The resultant circularly permuted protein, cpA16M-59, is secreted into the periplasm, correctly processed, and folded into a stable and active enzyme. Crystal structure analysis at 2.0-A resolution, R = 15.3%, shows cpA16M-59 to have a three-dimensional structure nearly identical with that of the parent beta-glucanase. An analogous experiment based on the wild-type Bacillus macerans beta-glucanase, giving rise to the circularly permuted variant cpMAC-57, yields the same results. Folding of these proteins, therefore, is not a vectorial process depending on the conformation adopted by their native N-terminal oligopeptides after ribosomal synthesis and translocation through the cytoplasmic membrane.

Hegge, F. T., P. G. Hitchen, et al. (2004). "Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of Neisseria gonorrhoeae type IV pili." Proc Natl Acad Sci U S A **101**(29): 10798-803.

Several major bacterial pathogens and related commensal species colonizing the human mucosa express phosphocholine (PC) at their cell surfaces. PC appears to impact host-microbe biology by serving as a ligand for both C-reactive protein and the receptor for platelet-activating factor. Type IV pili of Neisseria gonorrhoeae (Ng) and Neisseria meningitidis, filamentous protein structures critical to the colonization of their human hosts, are known to react variably with monoclonal antibodies recognizing a PC epitope. However, the structural basis

for this reactivity has remained elusive. To address this matter, we exploited the finding that the PilE pilin subunit in Ng mutants lacking the PilV protein acquired the PC epitope independent of changes in pilin primary structure. Specifically, we show by using mass spectrometry that PilE derived from the pilV background is composed of a mixture of subunits bearing O-linked forms of either phosphoethanolamine (PE) or PC at the same residue, whereas the wild-type background carries only PE at that same site. Therefore, PilV can influence pilin structure and antigenicity by modulating the incorporation of these alternative modifications. The disaccharide covalently linked to Ng pilin was also characterized because it is present on the same peptides bearing the PE and PC modifications and, contrary to previous reports, was found to be linked by means of 2,4-diacetamido-2,4,6-trideoxyhexose. Taken together, these findings provide new insights into Ng type IV pilus structure and antigenicity and resolve long-standing issues regarding the nature of both the PC epitope and the pilin glycan.

Herrmann, J. L., R. Delahay, et al. (2000). "Analysis of post-translational modification of mycobacterial proteins using a cassette expression system." FEBS Lett **473**(3): 358-62.

A recombinant expression system was developed to analyse sequence determinants involved in O-glycosylation of proteins in mycobacteria. By expressing peptide sequences corresponding to known glycosylation sites within a chimeric lipoprotein construct, amino acids flanking modified threonine residues were found to have an important influence on glycosylation. The expression system was used to screen mycobacterial sequences selected using a neural network (NetOglyc) trained on eukaryotic O-glycoproteins. Evidence of glycosylation was obtained for eight of 11 proteins tested. The results suggest that sites involved in O-glycosylation of mycobacterial and eukaryotic proteins share similar structural features.

Hettmann, T., C. L. Schmidt, et al. (1998). "Cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius. A novel highly glycosylated, membrane-bound b-type hemoprotein." J Biol Chem **273**(20): 12032-40.

In this study we re-examined the inducible cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius (DSM 639), formerly thought to be a component of a terminal oxidase (Becker, M., and Schafer, G. (1991) FEBS Lett. 291, 331-335). An improved purification method increased the yield of the protein and allowed more detailed investigations. Its molecular mass and heme content have been found to be 64,210 Da and 1 mol of heme/mol of protein, respectively. It is only detectable in cells grown at low oxygen tensions. The composition of the growth medium also exerts significant influence on the cytochrome b558/566 content of S. acidocaldarius membranes. The cytochrome exhibits an extremely high redox potential of +400 mV and shows no CO reactivity; a ligation other than a His/His-coordination of axial ligands appears likely. It turned out to be highly glycosylated (more than 20% of its molecular mass are sugar residues) and is probably exposed to the outer surface of the plasma membrane. The sugar moiety consists of several O-glycosidically linked mannoses and at least one N-glycosidically linked hexasaccharide comprising two glucoses, two mannoses,

and two N-acetyl-glucosamines. The gene of the cytochrome (cbsA) has been sequenced, revealing an interesting predicted secondary structure with two putative alpha-helical membrane anchors flanking the majority of a mainly beta-pleated sheet structure containing unusually high amounts of serine and threonine. A second gene (cbsB) was found to be cotranscribed. The latter displays extreme hydrophobicity and is thought to form a functional unit with cytochrome b558/566 in vivo, although it did not copurify with the latter. Sequence comparisons show no similarity to any entry in data banks indicating that this cytochrome is indeed a novel kind of b-type hemoprotein. A cytochrome c analogous function in the pseudoperiplasmic space of S. acidocaldarius is discussed.

Hirai, H., R. Takai, et al. "Glycosylation regulates the specific induction of rice immune responses by Acidovorax avenae flagellin." J Biol Chem.

Plants have a sensitive system that detects various pathogen-derived molecules to protect against infection. Flagellin, a main component of the bacterial flagellum, from the rice avirulent N1141 strain of gram-negative phytopathogenic bacterium, Acidovorax avenae, induces plant immune responses including H(2)O(2) generation, while flagellin from the rice virulent K1 strain of A. avenae does not induce these immune responses. To clarify the molecular mechanism that leads to these differing responses between the K1 and N1141 flagellins, recombinant K1 and N1141 flagellins were generated using an Escherichia coli expression system. When cultured rice cells were treated with recombinant K1 or N1141 flagellin, both flagellins equally induced H(2)O(2) generation, suggesting that post-translational modifications of the flagellins are involved in the specific induction of immune responses. Mass spectrometry analyses using glycosyltransferase-deficient mutants showed that 1,600 Da and 2,150 Da glycans were present on the flagellins from N1141 and K1, respectively. A deglycosylated K1 flagellin induced immune responses in the same manner as N1141 flagellin. Site-directed mutagenesis revealed that glycans were attached to four amino acid residues ((178)Ser, (183)Ser, (212)Ser and (351)Thr) in K1 flagellin. Among mutant K1 flagellins in which each glycan-attached amino acid residue was changed to alanine, (178)Ser/Ala and (183)Ser/Ala K1 flagellin induced a strong immune response in cultured rice cells, indicating that the glycans at (178)Ser and (183)Ser in K1 flagellin prevent epitope recognition in rice.

Horn, C., A. Namane, et al. (1999). "Decreased capacity of recombinant 45/47-kDa molecules (Apa) of Mycobacterium tuberculosis to stimulate T lymphocyte responses related to changes in their mannosylation pattern." J Biol Chem **274**(45): 32023-30.

The Apa molecules secreted by Mycobacterium tuberculosis, Mycobacterium bovis, or BCG have been identified as major immunodominant antigens. Mass spectrometry analysis indicated similar mannosylation, a complete pattern from 1 up to 9 hexose residues/mole of protein, of the native species from the 3 reference strains. The recombinant antigen expressed in M. smegmatis revealed a different mannosylation pattern: species containing 7 to 9 sugar residues/mole

of protein were in the highest proportion, whereas species bearing a low number of sugar residues were almost absent. The 45/47-kDa recombinant antigen expressed in E. coli was devoid of sugar residues. The proteins purified from M. tuberculosis, M. bovis, or BCG have a high capacity to elicit in vivo potent delayed-type hypersensitivity (DTH) reactions and to stimulate in vitro sensitized T lymphocytes of guinea pigs immunized with living BCG. The recombinant Apa expressed in Mycobacterium smegmatis was 4-fold less potent in vivo in the DTH assay and 10-fold less active in vitro to stimulate sensitized T lymphocytes than the native proteins. The recombinant protein expressed in Escherichia coli was nearly unable to elicit DTH reactions in vivo or to stimulate T lymphocytes in vitro. Thus the observed biological effects were related to the extent of glycosylation of the antigen.

Horzempa, J., C. R. Dean, et al. (2006). "Pseudomonas aeruginosa 1244 pilin glycosylation: glycan substrate recognition." J Bacteriol **188**(12): 4244-52.

The pilin of Pseudomonas aeruginosa 1244 is glycosylated with an oligosaccharide that is structurally identical to the O-antigen repeating unit of this organism. Concordantly, the metabolic source of the pilin glycan is the O-antigen biosynthetic pathway. The present study was conducted to investigate glycan substrate recognition in the 1244 pilin glycosylation reaction. Comparative structural analysis of O subunits that had been previously shown to be compatible with the 1244 glycosylation machinery revealed similarities among sugars at the presumed reducing termini of these oligosaccharides. We therefore hypothesized that the glycosylation substrate was within the sugar at the reducing end of the glycan precursor. Since much is known of PA103 O-antigen genetics and because the sugars at the reducing termini of the O7 (strain 1244) and O11 (strain PA103) are identical (beta-N-acetyl fucosamine), we utilized PA103 and strains that express lipopolysaccharide (LPS) with a truncated O-antigen subunit to test our hypothesis. LPS from a strain mutated in the wbjE gene produced an incomplete O subunit, consisting only of the monosaccharide at the reducing end (beta-d-N-acetyl fucosamine), indicating that this moiety contained substrate recognition elements for WaaL. Expression of pilAO(1244) in PA103 wbjE::aacC1, followed by Western blotting of extracts of these cells, indicated that pilin produced has been modified by the addition of material consistent with a single N-acetyl fucosamine. This was confirmed by analyzing endopeptidase-treated pilin by mass spectrometry. These data suggest that the pilin glycosylation substrate recognition features lie within the reducing-end moiety of the O repeat and that structures of the remaining sugars are irrelevant.

Huang, W., L. Boju, et al. (2001). "Active site of chondroitin AC lyase revealed by the structure of enzyme-oligosaccharide complexes and mutagenesis." Biochemistry **40**(8): 2359-72.

The crystal structures of Flavobacterium heparinium chondroitin AC lyase (chondroitinase AC; EC 4.2.2.5) bound to dermatan sulfate hexasaccharide (DS(hexa)), tetrasaccharide (DS(tetra)), and hyaluronic acid tetrasaccharide (HA(tetra)) have been refined at 2.0, 2.0, and 2.1 A resolution, respectively. The

structure of the Tyr234Phe mutant of AC lyase bound to a chondroitin sulfate tetrasaccharide (CS(tetra)) has also been determined to 2.3 A resolution. For each of these complexes, four (DS(hexa) and CS(tetra)) or two (DS(tetra) and HA(tetra)) ordered sugars are visible in electron density maps. The lyase AC DS(hexa) and CS(tetra) complexes reveal binding at four subsites, -2, -1, +1, and +2, within a narrow and shallow protein channel. We suggest that subsites -2 and -1 together represent the substrate recognition area, +1 is the catalytic subsite and +1 and +2 together represent the product release area. The putative catalytic site is located between the substrate recognition area and the product release area, carrying out catalysis at the +1 subsite. Four residues near the catalytic site, His225, Tyr234, Arg288, and Glu371 together form a catalytic tetrad. The mutations His225Ala, Tyr234Phe, Arg288Ala, and Arg292Ala, revealed residual activity for only the Arg292Ala mutant. Structural data indicate that Arg292 is primarily involved in recognition of the N-acetyl and sulfate moieties of galactosamine, but does not participate directly in catalysis. Candidates for the general base, removing the proton attached to C-5 of the glucuronic acid at the +1 subsite, are Tyr234, which could be transiently deprotonated during catalysis, or His225. Tyrosine 234 is a candidate to protonate the leaving group. Arginine 288 likely contributes to charge neutralization and stabilization of the enolate anion intermediate during catalysis.

Huang, W., V. V. Lunin, et al. (2003). "Crystal structure of Proteus vulgaris chondroitin sulfate ABC lyase I at 1.9A resolution." J Mol Biol **328**(3): 623-34.

Chondroitin Sulfate ABC lyase I from Proteus vulgaris is an endolytic, broad-specificity glycosaminoglycan lyase, which degrades chondroitin, chondroitin-4-sulfate, dermatan sulfate, chondroitin-6-sulfate, and hyaluronan by beta-elimination of 1,4-hexosaminidic bond to unsaturated disaccharides and tetrasaccharides. Its structure revealed three domains. The N-terminal domain has a fold similar to that of carbohydrate-binding domains of xylanases and some lectins, the middle and C-terminal domains are similar to the structures of the two-domain chondroitin lyase AC and bacterial hyaluronidases. Although the middle domain shows a very low level of sequence identity with the catalytic domains of chondroitinase AC and hyaluronidase, the residues implicated in catalysis of the latter enzymes are present in chondroitinase ABC I. The substrate-binding site in chondroitinase ABC I is in a wide-open cleft, consistent with the endolytic action pattern of this enzyme. The tryptophan residues crucial for substrate binding in chondroitinase AC and hyaluronidases are lacking in chondroitinase ABC I. The structure of chondroitinase ABC I provides a framework for probing specific functions of active-site residues for understanding the remarkably broad specificity of this enzyme and perhaps engineering a desired specificity. The electron density map showed clearly that the deposited DNA sequence for residues 495-530 of chondroitin ABC lyase I, the segment containing two putative active-site residues, contains a frame-shift error resulting in an incorrectly translated amino acid sequence.

Huang, W., A. Matte, et al. (1999). "Crystal structure of chondroitinase B from

Flavobacterium heparinum and its complex with a disaccharide product at 1.7 A resolution." J Mol Biol **294**(5): 1257-69.

Glycosaminoglycans (GAGs) are a family of acidic heteropolysaccharides, including such molecules as chondroitin sulfate, dermatan sulfate, heparin and keratan sulfate. Cleavage of the O-glycosidic bond within GAGs can be accomplished by hydrolases as well as lyases, yielding disaccharide and oligosaccharide products. We have determined the crystal structure of chondroitinase B, a glycosaminoglycan lyase from Flavobacterium heparinum, as well as its complex with a dermatan sulfate disaccharide product, both at 1.7 A resolution. Chondroitinase B adopts the right-handed parallel beta-helix fold, found originally in pectate lyase and subsequently in several polysaccharide lyases and hydrolases. Sequence homology between chondroitinase B and a mannuronate lyase from Pseudomonas sp. suggests this protein also adopts the beta-helix fold. Binding of the disaccharide product occurs within a positively charged cleft formed by loops extending from the surface of the beta-helix. Amino acid residues responsible for recognition of the disaccharide, as well as potential catalytic residues, have been identified. Two arginine residues, Arg318 and Arg364, are found to interact with the sulfate group attached to O-4 of N-acetylgalactosamine. Cleavage of dermatan sulfate likely occurs at the reducing end of the disaccharide, with Glu333 possibly acting as the general base.

Igura, M. and D. Kohda "Selective control of oligosaccharide transfer efficiency for the N-glycosylation sequon by a point mutation in oligosaccharyltransferase." J Biol Chem **286**(15): 13255-60.

Asn-linked glycosylation is the most ubiquitous posttranslational protein modification in eukaryotes and archaea, and in some eubacteria. Oligosaccharyltransferase (OST) catalyzes the transfer of preassembled oligosaccharides on lipid carriers onto asparagine residues in polypeptide chains. Inefficient oligosaccharide transfer results in glycoprotein heterogeneity, which is particularly bothersome in pharmaceutical glycoprotein production. Amino acid variation at the X position of the Asn-X-Ser/Thr sequon is known to modulate the glycosylation efficiency. The best amino acid at X is valine, for an archaeal Pyrococcus furiosus OST. We performed a systematic alanine mutagenesis study of the archaeal OST to identify the essential and dispensable amino acid residues in the three catalytic motifs. We then investigated the effects of the dispensable mutations on the amino acid preference in the N-glycosylation sequon. One residue position was found to selectively affect the amino acid preference at the X position. This residue is located within the recently identified DXXKXXX(M/I) motif, suggesting the involvement of this motif in N-glycosylation sequon recognition. In applications, mutations at this position may facilitate the design of OST variants adapted to particular N-glycosylation sites to reduce the heterogeneity of glycan occupancy. In fact, a mutation at this position led to 9-fold higher activity relative to the wild-type enzyme, toward a peptide containing arginine at X in place of valine. This mutational approach is potentially applicable to eukaryotic and eubacterial OSTs for the production of homogenous glycoproteins in engineered mammalian and Escherichia coli cells.

Igura, M., N. Maita, et al. (2008). "Structure-guided identification of a new catalytic motif of oligosaccharyltransferase." EMBO J **27**(1): 234-43.

Asn-glycosylation is widespread not only in eukaryotes but also in archaea and some eubacteria. Oligosaccharyltransferase (OST) catalyzes the co-translational transfer of an oligosaccharide from a lipid donor to an asparagine residue in nascent polypeptide chains. Here, we report that a thermophilic archaeon, Pyrococcus furiosus OST is composed of the STT3 protein alone, and catalyzes the transfer of a heptasaccharide, containing one hexouronate and two pentose residues, onto peptides in an Asn-X-Thr/Ser-motif-dependent manner. We also determined the 2.7-A resolution crystal structure of the C-terminal soluble domain of Pyrococcus STT3. The structure-based multiple sequence alignment revealed a new motif, DxxK, which is adjacent to the well-conserved WWDYG motif in the tertiary structure. The mutagenesis of the DK motif residues in yeast STT3 revealed the essential role of the motif in the catalytic activity. The function of this motif may be related to the binding of the pyrophosphate group of lipid-linked oligosaccharide donors through a transiently bound cation. Our structure provides the first structural insights into the formation of the oligosaccharide-asparagine bond.

Jarrell, K. F., G. M. Jones, et al. "Biosynthesis and role of N-linked glycosylation in cell surface structures of archaea with a focus on flagella and s layers." Int J Microbiol **2010**: 470138.

The genetics and biochemistry of the N-linked glycosylation system of Archaea have been investigated over the past 5 years using flagellins and S layers as reporter proteins in the model organisms, Methanococcus voltae, Methanococcus maripaludis, and Haloferax volcanii. Structures of archaeal N-linked glycans have indicated a variety of linking sugars as well as unique sugar components. In M. voltae, M. maripaludis, and H. volcanii, a number of archaeal glycosylation genes (agl) have been identified by deletion and complementation studies. These include many of the glycosyltransferases and the oligosaccharyltransferase needed to assemble the glycans as well as some of the genes encoding enzymes required for the biosynthesis of the sugars themselves. The N-linked glycosylation system is not essential for any of M. voltae, M. maripaludis, or H. volcanii, as demonstrated by the successful isolation of mutants carrying deletions in the oligosaccharyltransferase gene aglB (a homologue of the eukaryotic Stt3 subunit of the oligosaccharyltransferase complex). However, mutations that affect the glycan structure have serious effects on both flagellation and S layer function.

Jennings, M. P., F. E. Jen, et al. "Neisseria gonorrhoeae pilin glycan contributes to CR3 activation during challenge of primary cervical epithelial cells." Cell Microbiol **13**(6): 885-96.

Expression of type IV pili by Neisseria gonorrhoeae plays a critical role in mediating adherence to human epithelial cells. Gonococcal pilin is modified with an O-linked glycan, which may be present as a di- or monosaccharide because

of phase variation of select pilin glycosylation genes. It is accepted that bacterial proteins may be glycosylated; less clear is how the protein glycan may mediate virulence. Using primary, human, cervical epithelial (i.e. pex) cells, we now provide evidence to indicate that the pilin glycan mediates productive cervical infection. In this regard, pilin glycan-deficient mutant gonococci exhibited an early hyper-adhesive phenotype but were attenuated in their ability to invade pex cells. Our data further indicate that the pilin glycan was required for gonococci to bind to the I-domain region of complement receptor 3, which is naturally expressed by pex cells. Comparative, quantitative, infection assays revealed that mutant gonococci lacking the pilin glycan did not bind to the I-domain when it is in a closed, low-affinity conformation and cannot induce an active conformation to complement receptor 3 during pex cell challenge. To our knowledge, these are the first data to directly demonstrate how a protein-associated bacterial glycan may contribute to pathogenesis.

Josenhans, C., L. Vossebein, et al. (2002). "The neuA/flmD gene cluster of Helicobacter pylori is involved in flagellar biosynthesis and flagellin glycosylation." FEMS Microbiol Lett **210**(2): 165-72.

Helicobacter pylori possesses a gene (HP0326/JHP309) homologous to neuA of other bacteria, encoding a cytidyl monophosphate-N-acetylneuraminic acid synthetase-homologous enzyme in its N-terminal portion. We analysed the function of this gene, which is controlled by a flagellar class 2 sigma(54) promoter, in flagellar biosynthesis. HP0326/JHP309 actually represents a bicistronic operon consisting of a neuA and a flmD-like putative glycosyl transferase gene. An isogenic flmD mutant synthesized basal bodies but no filaments, was non-motile, and expressed severely reduced amounts of a FlaA flagellin of reduced molecular mass. FlaA flagellin was found to be glycosylated in its exported form within the flagellar filament, but not inside the cytoplasm. Glycosylated FlaA was not detectable in the flmD mutant. Together with other genes in the H. pylori genome, a proposed function of the neuA/flmD gene products could be to provide a pathway for glycosylation of flagellin and other extracytoplasmic molecules during type III secretion.

Kahlig, H., D. Kolarich, et al. (2005). "N-acetylmuramic acid as capping element of alpha-D-fucose-containing S-layer glycoprotein glycans from Geobacillus tepidamans GS5-97T." J Biol Chem **280**(21): 20292-9.

Geobacillus tepidamans GS5-97(T) is a novel Gram-positive, moderately thermophilic bacterial species that is covered by a glycosylated surface layer (S-layer) protein. The isolated and purified S-layer glycoprotein SgtA was ultrastructurally and chemically investigated and showed several novel properties. By SDS-PAGE, SgtA was separated into four distinct bands in an apparent molecular mass range of 106-166 kDa. The three high molecular mass bands gave a positive periodic acid-Schiff staining reaction, whereas the 106-kDa band was nonglycosylated. Glycosylation of SgtA was investigated by means of chemical analyses, 600-MHz nuclear magnetic resonance spectroscopy, and electrospray ionization quadrupole time-of-fight mass

spectrometry. Glycopeptides obtained after Pronase digestion revealed the glycan structure [-->2)-alpha-L-Rhap-(1-->3)-alpha-D-Fucp-(1-->](n=approximately 20), with D-fucopyranose having never been identified before as a constituent of S-layer glycans. The rhamnose residue at the nonreducing end of the terminal repeating unit of the glycan chain was di-substituted. For the first time, (R)-N-acetylmuramic acid, the key component of prokaryotic peptidoglycan, was found in an alpha-linkage to carbon 3 of the terminal rhamnose residue, serving as capping motif of an S-layer glycan. In addition, that rhamnose was substituted at position 2 with a beta-N-acetylglucosamine residue. The S-layer glycan chains were bound via the trisaccharide core -->2)-alpha-L-Rhap-(1-->3)-alpha-L-Rhap-(1-->3)-alpha-L-Rhap-(1--> to carbon 3 of beta-D-galactose, which was attached in O-glycosidic linkage to serine and threonine residues of SgtA of G. tepidamans GS5-97(T).

Kakuda, T. and V. J. DiRita (2006). "Cj1496c encodes a Campylobacter jejuni glycoprotein that influences invasion of human epithelial cells and colonization of the chick gastrointestinal tract." Infect Immun **74**(8): 4715-23.

Campylobacter jejuni has an N-linked protein glycosylation pathway that is required for efficient cell invasion and chick gastrointestinal colonization by the microbe. In this study, we constructed insertion mutants of 22 putative glycoprotein genes and examined the ability of each to invade the human intestinal epithelial cell line INT-407. Among the mutants tested, one carrying an insertion in Cj1496c was defective for invasion into INT-407 cells; this defect was also observed in an in-frame deletion mutant of Cj1496c (delta Cj1496c). The delta Cj1496c mutant C. jejuni also showed a reduced ability to colonize chick ceca. Site-specific mutagenesis combined with Western blot analysis suggested that the Cj1496c protein is glycosylated at N73 and N169. However, the delta Cj1496c mutant expressing a nonglycosylated form of Cj1496c exhibited levels of invasion and colonization equivalent to those of the parent strain, suggesting that glycans are not directly involved in the function of Cj1496c.

Kalmokoff, M. L., S. F. Koval, et al. (1992). "Relatedness of the flagellins from methanogens." Arch Microbiol **157**(6): 481-7.

Purified flagellar filaments isolated from six methanogens were composed of multiple flagellins. Two flagellins were present in Methanococcus deltae (Mr = 34,000 and 32,000), Methanoculleus marisnigri (Mr = 31,000 and 25,500) and Methanococcus jannaschii (Mr = 31,000 and 27,500), three in Methanothermus fervidus (Mr = 34,000, 25,000 and 24,000) and four or more in both Methanococcus vannielii and Methanococcus maripaludis (Mr ranging from 27,500 to 32,000). The flagellins of M. fervidus and M. deltae reacted positively with glycoprotein-specific stains. The flagellins of M. deltae, M. maripaludis and M. vannielii were closely related to those of M. voltae based on cross-reactivity with antisera raised against M. voltae flagellins and homology with flagellin-specific oligonucleotide probes to the N-terminus and leader peptide of M. voltae flagellins. Similarities appear to exist among the flagellins of M. fervidus, M. marisnigri and Halobacterium halobium based on cross-reactivity with antisera

produced against the flagella of Methanospirillum hungatei JF1. The N-termini of the flagellins from the mesophilic Methanococcus spp. and M. marisnigri show homology with the N-termini of other archaebacterial flagellins. These N-termini may undergo a modification involving removal of a leader peptide.

Kaminski, L., M. Abu-Qarn, et al. "AglJ adds the first sugar of the N-linked pentasaccharide decorating the Haloferax volcanii S-layer glycoprotein." J Bacteriol **192**(21): 5572-9.

Like the Eukarya and Bacteria, the Archaea also perform N glycosylation. Using the haloarchaeon Haloferax volcanii as a model system, a series of Agl proteins involved in the archaeal version of this posttranslational modification has been identified. In the present study, the participation of HVO_1517 in N glycosylation was considered, given its homology to a known component of the eukaryal N-glycosylation pathway and because of the genomic proximity of HVO_1517 to agl genes encoding known elements of the H. volcanii N-glycosylation process. By combining the deletion of HVO_1517 with mass spectrometric analysis of both dolichol phosphate monosaccharide-charged carriers and the S-layer glycoprotein, evidence was obtained showing the participation of HVO_1517, renamed AglJ, in adding the first hexose of the N-linked pentasaccharide decorating this reporter glycoprotein. The deletion of aglJ, however, did not fully prevent the attachment of a hexose residue to the S-layer glycoprotein. Moreover, in the absence of AglJ, the level of only one of the three monosaccharide-charged dolichol phosphate carriers detected in the cell was reduced. Nonetheless, in cells lacking AglJ, no further sugar subunits were added to the remaining monosaccharide-charged dolichol phosphate carriers or to the monosaccharide-modified S-layer glycoprotein, pointing to the importance of the sugar added through the actions of AglJ for proper N glycosylation. Finally, while aglJ can be deleted, H. volcanii surface layer integrity is compromised in the absence of the encoded protein.

Kaminski, L. and J. Eichler "Identification of residues important for the activity of Haloferax volcanii AglD, a component of the archaeal N-glycosylation pathway." Archaea **2010**: 315108.

In Haloferax volcanii, AglD adds the final hexose to the N-linked pentasaccharide decorating the S-layer glycoprotein. Not knowing the natural substrate of the glycosyltransferase, together with the challenge of designing assays compatible with hypersalinity, has frustrated efforts at biochemical characterization of AglD activity. To circumvent these obstacles, an in vivo assay designed to identify amino acid residues important for AglD activity is described. In the assay, restoration of AglD function in an Hfx. volcanii aglD deletion strain transformed to express plasmid-encoded versions of AglD, generated through site-directed mutagenesis at positions encoding residues conserved in archaeal homologues of AglD, is reflected in the behavior of a readily detectable reporter of N-glycosylation. As such Asp110 and Asp112 were designated as elements of the DXD motif of AglD, a motif that interacts with metal cations associated with nucleotide-activated sugar donors, while Asp201 was predicted to be the

catalytic base of the enzyme.

Karcher, U., H. Schroder, et al. (1993). "Primary structure of the heterosaccharide of the surface glycoprotein of Methanothermus fervidus." J Biol Chem **268**(36): 26821-6.

The outer surface of the cells of the hyperthermophile Methanothermus fervidus is covered by crystalline glycoprotein subunits (S-layer). From the purified S-layer glycoprotein, a heterosaccharide was isolated. The heterosaccharide consists of D-3-O-methylmannose, D-mannose, and D-N-acetylgalactosamine in a molar ratio of 2:3:1 corresponding to a relative molecular mass of 1061.83 Da. 3-O-methylmannose could be partly replaced by 3-O-methylglucose. The primary structure of the glycan was revealed by methylation analysis, by plasma desorption mass spectrometry, and by high field NMR spectroscopy. The purified heterosaccharide is linked via N-acetylgalactosamine to an asparagine residue of the peptide moiety. The following structure is proposed for the heterosaccharide: alpha-D-3-O-MetManp-(1-->6)-alpha-D-3-O-MetManp-((1-->2)-alp ha-D-Manp)3-(1-->4) - D-GalNAc.

Keitel, T., M. Meldgaard, et al. (1994). "Cation binding to a Bacillus (1,3-1,4)-beta-glucanase. Geometry, affinity and effect on protein stability." Eur J Biochem **222**(1): 203-14.

The hybrid Bacillus (1,3-1,4)-beta-glucanase H(A16-M), consisting of 16 N-terminal amino acids derived from the mature form of the B. amyloliquefaciens enzyme and of 198 C-proximal amino acids from the B. macerans enzyme, binds a calcium ion at a site at its molecular surface remote from the active center [T. Keitel, O. Simon, R. Borriss & U. Heinemann (1993) Proc. Natl Acad. Sci. USA 90, 5287-5291]. X-ray diffraction analysis at 0.22-nm resolution of crystals grown in the absence of calcium and in the presence of EDTA shows this site to be occupied by a sodium ion. Whereas the calcium ion has six oxygen atoms in its coordination sphere, two of which are from water molecules, sodium is fivefold coordinated with a fifth ligand belonging to a symmetry-related protein molecule in the crystal lattice. The affinity of H(A16-M) for calcium over sodium has been determined calorimetrically. Calcium binding stabilizes the native three-dimensional structure of the protein as shown by guanidinium chloride unfolding and thermal inactivation experiments. The enhanced enzymic activity of Bacillus beta-glucanases at elevated temperatures in the presence of calcium ions is attributed to a general stabilizing effect by the cation.

Kelly, J., S. M. Logan, et al. (2009). "A novel N-linked flagellar glycan from Methanococcus maripaludis." Carbohydr Res **344**(5): 648-53.

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 1036.4Da, 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-beta-ManNAc3NAmA6Thr-4-beta-GlcNAc3NAcA-3-beta-GalNAc-Asn where Sug is a novel monosaccharide unit, (5S)-2-acetamido-2,4-dideoxy-5-O-methyl-alpha-l-erythro-hexos-5-ulo-1,5-py ranose. This oligosaccharide has significant similarity to the oligosaccharide that was found previously in Methanococcus voltae.

Kneidinger, B., M. Graninger, et al. (2001). "Biosynthesis of nucleotide-activated D-glycero-D-manno-heptose." J Biol Chem **276**(24): 20935-44.

The glycan chain repeats of the S-layer glycoprotein of Aneurinibacillus thermoaerophilus DSM 10155 contain d-glycero-d-manno-heptose, which has also been described as constituent of lipopolysaccharide cores of Gram-negative bacteria. The four genes required for biosynthesis of the nucleotide-activated form GDP-d-glycero-d-manno-heptose were cloned, sequenced, and overexpressed in Escherichia coli, and the corresponding enzymes GmhA, GmhB, GmhC, and GmhD were purified to homogeneity. The isomerase GmhA catalyzed the conversion of d-sedoheptulose 7-phosphate to d-glycero-d-manno-heptose 7-phosphate, and the phosphokinase GmhB added a phosphate group to form d-glycero-d-manno-heptose 1,7-bisphosphate. The phosphatase GmhC removed the phosphate in the C-7 position, and the intermediate d-glycero-alpha-d-manno-heptose 1-phosphate was eventually activated with GTP by the pyrophosphorylase GmhD to yield the final product GDP-d-glycero-alpha-d-manno-heptose. The intermediate and end products were analyzed by high performance liquid chromatography. Nuclear magnetic resonance spectroscopy was used to confirm the structure of these substances. This is the first report of the biosynthesis of GDP-d-glycero-alpha-d-manno-heptose in Gram-positive organisms. In addition, we propose a pathway for biosynthesis of the nucleotide-activated form of l-glycero-d-manno-heptose.

Knudsen, S. K., A. Stensballe, et al. (2008). "Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation." Biochem J **412**(3): 563-77.

Autotransporters constitute the biggest group of secreted proteins in Gram-negative bacteria and contain a membrane-bound beta-domain and a passenger domain secreted to the extracellular environment via an unusually long N-terminal sequence. Several passenger domains are known to be glycosylated by cytosolic glycosyl transferases, promoting bacterial attachment to mammalian cells. In the present study we describe the effect of glycosylation on the extracellular passenger domain of the Escherichia coli autotransporter Ag43alpha, which induces frizzy colony morphology and cell settling. We identify 16 glycosylation sites and suggest two possible glycosylation motifs for serine and threonine residues. Glycosylation stabilizes against thermal and chemical denaturation and increases refolding kinetics. Unexpectedly, glycosylation also reduces the stabilizing effect of Ca(2+) ions, removes the ability of Ca(2+) to promote cell adhesion, reduces the ability of Ag43alpha-containing cells to form bacterial amyloid and increases the susceptibility of the resulting amyloid to proteolysis. In addition, our results indicate that Ag43alpha folds without a stable

intermediate, unlike pertactin, indicating that autotransporters may arrive at the native state by a variety of different mechanisms despite a common overall structure. A small but significant fraction of Ag43alpha can survive intact in the periplasm if expressed without the beta-domain, suggesting that it is able to adopt a protease-resistant structure prior to translocation across the membrane. The present study demonstrates that glycosylation may play significant roles in structural and functional properties of bacterial autotransporters at many different levels.

Konishi, T., F. Taguchi, et al. (2009). "Structural characterization of an O-linked tetrasaccharide from Pseudomonas syringae pv. tabaci flagellin." Carbohydr Res **344**(16): 2250-4.

The flagellin of Pseudomonas syringae pv. tabaci is a glycoprotein that contains O-linked oligosaccharides composed of rhamnosyl and 4,6-dideoxy-4-(3-hydroxybutanamido)-2-O-methylglucosyl residues. These O-linked glycans are released by hydrazinolysis and then labeled at their reducing ends with 2-aminopyridine (PA). A PA-labeled trisaccharide and a PA-labeled tetrasaccharide are isolated by normal-phase high-performance liquid chromatography. These oligosaccharides are structurally characterized using mass spectrometry and NMR spectroscopy. Our data show that P. syringae pv. tabaci flagellin is glycosylated with a tetrasaccharide, 4,6-dideoxy-4-(3-hydroxybutanamido)-2-O-methyl-Glcp-(1-->3)-alpha-L-Rhap-(1-->2)-alpha-L-Rhap-(1-->2)-alpha-L-Rha-(1-->, as well a trisaccharide, 4,6-dideoxy-4-(3-hydroxybutanamido)-2-O-methyl-Glcp-(1-->3)-alpha-L-Rhap-(1-->2)-alpha-L-Rha-(1-->, which was identified in a previous study.

Kosma, P., C. Neuninger, et al. (1995). "Glycan structure of the S-layer glycoprotein of Bacillus sp. L420-91." Glycoconj J **12**(1): 99-107.

Preliminary taxonomic characterization of isolate L420-91 has revealed that this organism is closely related to the species Bacillus aneurinolyticus. The bacterium is covered by a squarely arranged crystalline surface layer composed of identical glycoprotein subunits with an apparent molecular mass in the range of 109 kDa. A total carbohydrate content of approximately 3.5% (wt/wt) was determined in the purified surface layer glycoprotein. Glycopeptides were obtained after exhaustive Pronase digestion and purification including gel filtration, ion exchange chromatography and HPLC. From the combined evidence of composition analysis. Smith degradation and nuclear magnetic resonance spectroscopy experiments we propose the following structure for the glycan chain of the surface layer glycoprotein: [formula: see text]

Kosma, P., T. Wugeditsch, et al. (1995). "Glycan structure of a heptose-containing S-layer glycoprotein of Bacillus thermoaerophilus." Glycobiology **5**(8): 791-6.

The characterization of the S-layer glycoprotein of Bacillus thermoaerophilus revealed unexpected novelties. The isolation and purification procedure had to be changed due to complete solubility in aqueous buffers of the constituting S-layer protomers. Upon degradation of the S-layer glycoprotein by pronase and

purification of the products by gel filtration, ion-exchange chromatography, chromatofocusing and HPLC, one representative glycopeptide fraction was selected for further characterization. From the combined evidence of composition analysis, chemical degradation, NMR spectroscopy experiments and comparison with synthesized model substance, we propose the following repeating unit structure of the glycan chain: -->4)-alpha-L-Rhap-(1-->3)-beta-D-glycero-D-manno-Hepp-(1--> This is the first description of heptose residues occurring as a constituent of S-layer glycoproteins of gram-positive eubacteria.

Kowarik, M., N. M. Young, et al. (2006). "Definition of the bacterial N-glycosylation site consensus sequence." EMBO J **25**(9): 1957-66.

The Campylobacter jejuni pgl locus encodes an N-linked protein glycosylation machinery that can be functionally transferred into Escherichia coli. In this system, we analyzed the elements in the C. jejuni N-glycoprotein AcrA required for accepting an N-glycan. We found that the eukaryotic primary consensus sequence for N-glycosylation is N terminally extended to D/E-Y-N-X-S/T (Y, X not equalP) for recognition by the bacterial oligosaccharyltransferase (OST) PglB. However, not all consensus sequences were N-glycosylated when they were either artificially introduced or when they were present in non-C. jejuni proteins. We were able to produce recombinant glycoproteins with engineered N-glycosylation sites and confirmed the requirement for a negatively charged side chain at position -2 in C. jejuni N-glycoproteins. N-glycosylation of AcrA by the eukaryotic OST in Saccharomyces cerevisiae occurred independent of the acidic residue at the -2 position. Thus, bacterial N-glycosylation site selection is more specific than the eukaryotic equivalent with respect to the polypeptide acceptor sequence.

Kupcu, Z., L. Marz, et al. (1984). "Evidence for the glycoprotein nature of the crystalline cell wall surface layer of Bacillus stearothermophilus strain NRS2004/3a." FEBS Lett **173**(1): 185-90.

The surface layer of Bacillus stearothermophilus strain NRS2004/3a was isolated and chemically characterized. The results of these initial studies lead to the conclusion that the cell surface protein is glycosylated.

Kus, J. V., J. Kelly, et al. (2008). "Modification of Pseudomonas aeruginosa Pa5196 type IV Pilins at multiple sites with D-Araf by a novel GT-C family Arabinosyltransferase, TfpW." J Bacteriol **190**(22): 7464-78.

Pseudomonas aeruginosa Pa5196 produces type IV pilins modified with unusual alpha1,5-linked d-arabinofuranose (alpha1,5-D-Araf) glycans, identical to those in the lipoarabinomannan and arabinogalactan cell wall polymers from Mycobacterium spp. In this work, we identify a second strain of P. aeruginosa, PA7, capable of expressing arabinosylated pilins and use a combination of site-directed mutagenesis, electrospray ionization mass spectrometry (MS), and electron transfer dissociation MS to identify the exact sites and extent of pilin modification in strain Pa5196. Unlike previously characterized type IV pilins that are glycosylated at a single position, those from strain Pa5196 were modified at

multiple sites, with modifications of alphabeta-loop residues Thr64 and Thr66 being important for normal pilus assembly. Trisaccharides of alpha1,5-D-Araf were the principal modifications at Thr64 and Thr66, with additional mono- and disaccharides identified on Ser residues within the antiparallel beta sheet region of the pilin. TfpW was hypothesized to encode the pilin glycosyltransferase based on its genetic linkage to the pilin, weak similarity to membrane-bound GT-C family glycosyltransferases (which include the Mycobacterium arabinosyltransferases EmbA/B/C), and the presence of characteristic motifs. Loss of TfpW or mutation of key residues within the signature GT-C glycosyltransferase motif completely abrogated pilin glycosylation, confirming its involvement in this process. A Pa5196 pilA mutant complemented with other Pseudomonas pilins containing potential sites of modification expressed nonglycosylated pilins, showing that TfpW's pilin substrate specificity is restricted. TfpW is the prototype of a new type IV pilin posttranslational modification system and the first reported gram-negative member of the GT-C glycosyltransferase family.

Kus, J. V., E. Tullis, et al. (2004). "Significant differences in type IV pilin allele distribution among Pseudomonas aeruginosa isolates from cystic fibrosis (CF) versus non-CF patients." Microbiology **150**(Pt 5): 1315-26.

Type IV pili (TFP) are important colonization factors of the opportunistic pathogen Pseudomonas aeruginosa, involved in biofilm formation and attachment to host cells. This study undertook a comprehensive analysis of TFP alleles in more than 290 environmental, clinical, rectal and cystic fibrosis (CF) isolates of P. aeruginosa. Based on the results, a new system of nomenclature is proposed, in which P. aeruginosa TFP are divided into five distinct phylogenetic groups. Each pilin allele is stringently associated with characteristic, distinct accessory genes that allow the identification of the allele by specific PCR. The invariant association of the pilin and accessory genes implies horizontal transfer of the entire locus. Analysis of pilin allele distribution among isolates from various sources revealed a striking bias in the prevalence of isolates with group I pilin genes from CF compared with non-CF human sources ($P<0.0001$), suggesting this particular pilin type, which can be post-translationally modified by glycosylation via the action of TfpO (PilO), may confer a colonization or persistence advantage in the CF host. This allele was also predominant in paediatric CF isolates (29 of 43; 67.4 %), showing that this bias is apparent early in colonization. Group I pilins were also the most common type found in environmental isolates tested. To the authors' knowledge, this is the first example of a P. aeruginosa virulence factor allele that is strongly associated with CF isolates.

Lara, M., L. Servin-Gonzalez, et al. (2004). "Expression, secretion, and glycosylation of the 45- and 47-kDa glycoprotein of Mycobacterium tuberculosis in Streptomyces lividans." Appl Environ Microbiol **70**(2): 679-85.

The gene encoding the 45/47 kDa glycoprotein (Rv1860) of Mycobacterium tuberculosis was expressed in Streptomyces lividans under its own promoter and

under the thiostrepton-inducible Streptomyces promoter PtipA. The recombinant protein was released into the culture medium and, like the native protein, migrated as a double band at 45 and 47 kDa in sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) gels. However, in contrast to the native protein, only the 47-kDa recombinant protein could be labeled with concanavalin A (ConA). Carbohydrate digestion with jack bean alpha-D-mannosidase resulted in a reduction in the molecular mass of the recombinant protein upper band and completely eliminated ConA binding. Two-dimensional gel electrophoresis revealed only one isoelectric point for the recombinant protein. Comparative fingerprinting analysis of the individually purified upper and lower recombinant protein bands, treated under the same conditions with specific proteases, resulted in similar peptide patterns, and the peptides had the same N-terminal sequence, suggesting that migration of the recombinant protein as two bands in SDS-PAGE gels could be due to differences in glycosylation. Mass spectrometry analysis of the recombinant protein indicated that as in native protein, both the N-terminal and C-terminal domains of the recombinant protein are glycosylated. Furthermore, it was determined that antibodies of human tuberculosis patients reacted mainly against the carbohydrate residues of the glycoprotein. Altogether, these observations show that expression of genes for mycobacterial antigens in S. lividans is very useful for elucidation of the functional role and molecular mechanisms of glycosylation in bacteria.

Larsen, J. C., C. Szymanski, et al. (2004). "N-linked protein glycosylation is required for full competence in Campylobacter jejuni 81-176." J Bacteriol **186**(19): 6508-14.

The recent sequencing of the virulence plasmid of Campylobacter jejuni 81-176 revealed the presence of genes homologous to type IV secretion systems (TFSS) that have subsequently been found in Helicobacter pylori and Wolinella succinogenes. Mutational analyses of some of these genes have implicated their involvement in intestinal epithelial cell invasion and natural competence. In this report, we demonstrate that one of these type IV secretion homologs, Cjp3/VirB10, is a glycoprotein. Treatment with various glycosidases and binding to soybean agglutinin indicated that the structure of the glycan present on VirB10 contains a terminal GalNAc, consistent with previous reports of N-linked glycans in C. jejuni. Site-directed mutagenesis of five putative N-linked glycosylation sites indicated that VirB10 is glycosylated at two sites, N32 and N97. Mutants in the N-linked general protein glycosylation (pgl) system of C. jejuni are significantly reduced in natural transformation, which is likely due, in part, to lack of glycosylation of VirB10. The natural transformation defect in a virB10 mutant can be complemented in trans by using a plasmid expressing wild-type VirB10 or an N32A substitution but not by using a mutant expressing VirB10 with an N97A substitution. Taken together, these results suggest that glycosylation of VirB10 specifically at N97 is required for the function of the TFSS and for full competence in C. jejuni 81-176.

Lechner, J. and M. Sumper (1987). "The primary structure of a procaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria." J Biol

Chem **262**(20): 9724-9.

The hexagonally patterned surface layer of halobacteria consists of a true glycoprotein. This procaryotic glycoprotein has recently been shown to exhibit novel features with respect to saccharide structure and saccharide biosynthesis. The primary structure and the location of glycosylation sites were determined by cloning and sequencing of the glycoprotein gene of Halobacterium halobium. According to the predicted amino acid sequence, the glycoprotein is synthesized with a N-terminal leader sequence of 34 amino acid residues reminiscent of eucaryotic and procaryotic signal peptides. A hydrophobic stretch of 21 amino acid residues at the C terminus probably serves as a transmembrane domain. 14 threonine residues are clustered adjacent to this membrane anchor and linked to these threonines are all the disaccharides of the cell surface glycoprotein. 12 N-glycosylation sites are distributed over the polypeptide chain.

Lechner, J., F. Wieland, et al. (1985). "Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. Purification and identification of sulfated dolichyl monophosphoryl tetrasaccharides from halobacteria." J Biol Chem **260**(2): 860-6.

A novel type of N-glycosidic linkage, asparaginyl glucose, occurs in the cell surface glycoprotein of halobacteria (Wieland, F., Heitzer, R., and Schaefer, W. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 5470-5474). Sulfated oligosaccharides containing glucuronic acids are attached to the polypeptide chain via this linkage. Here we describe the isolation and chemical characterization of lipid-linked precursors of these saccharides, and these have the following new features. Rather than the bacterial undecaprenol, a C60-dolichol is the carrier lipid. The oligosaccharide is bound to this lipid via a monophosphate, rather than a pyrophosphate bridge. Sulfation of the saccharides is completed while they are linked to lipid and does not occur after transfer of the saccharides to protein.

Lechner, J., F. Wieland, et al. (1985). "Transient methylation of dolichyl oligosaccharides is an obligatory step in halobacterial sulfated glycoprotein biosynthesis." J Biol Chem **260**(15): 8984-9.

Biosynthesis of sulfated saccharides that are linked to asparagine residues in the cell surface glycoprotein of Halobacterium halobium via a glucose residue involves sulfated dolichyl-monophosphoryl oligosaccharide intermediates (Lechner, J., Wieland, F., and Sumper, M. (1985) J. Biol. Chem. 260, 860-866). During isolation and characterization of these lipid oligosaccharides we detected a group of related compounds containing additional unidentified sugar residues. Here we report that: 1) the unknown sugar residues were 3-O-methylglucose, linked peripherally to the lipid-saccharide intermediates; 2) the 3-O-methylglucose residues in the oligosaccharides occur only at the lipid-linked level but are absent at the protein-linked level; 3) cell surface glycoprotein biosynthesis in Halobacteria in vivo is drastically depressed when S-adenosylmethionine-dependent methylation is inhibited, indicating that methylation is an obligatory step during glycoprotein synthesis. We propose a mechanism for the transport of lipid oligosaccharides through the cell membrane,

involving an intermediate stage in which the saccharide moieties are transiently modified with 3-O-methylglucose.

Lin, X. and J. Tang (1990). "Purification, characterization, and gene cloning of thermopsin, a thermostable acid protease from Sulfolobus acidocaldarius." J Biol Chem **265**(3): 1490-5.

A thermostable, acid proteolytic activity has been found to be associated with the cells and in the culture medium of Sulfolobus acidocaldarius, an archaebacterium. This acid protease, which has been named thermopsin, was purified to homogeneity from the culture medium by a five-step procedure including column chromatographies on DEAE-Sepharose CL-6B, phenyl-Sepharose CL-4B, Sephadex G-100, monoQ (fast protein liquid chromatography), and gel filtration (high pressure liquid chromatography). The purified thermopsin produced a single band on sodium dodecyl sulfate-polyacrylamide gel electrophoresis and the proteolytic activity was associated with the band. Thermopsin is a single-chain protein as indicated by gel electrophoresis and by a single NH2-terminal sequence. It has maximal proteolytic activity at pH 2 and 90 degrees C. A genomic library of S. acidocaldarius was prepared and screened by an oligonucleotide probe designed from the NH2-terminal sequence of thermopsin. Five positive clones were isolated. From these clones the thermopsin gene was mapped and sequenced. The nucleotide sequence showed that the thermopsin structure is encoded in 1020 bases. In the deduced protein sequence, there are 41 amino acid residues (including the initiation Met) preceding the NH2-terminal position of thermopsin. Most of these residues appear to be characteristic of a leader sequence. However, the presence in this region of a short pro sequence cannot be ruled out. Thermopsin contains a single cysteine at residue 237 that is not essential for activity (Fusek, M., Lin, X.-L., Tang, J. (1990) J. Biol. Chem. 265, 1496-1501. Thermopsin has no apparent sequence similarity to aspartic proteases of the pepsin family nor to pepstatin-insensitive acid protease (Maita, T., Nagata, S., Matsuda, G., Murata, S., Oda, K., Murao, S., and Tsura, D. (1984) J. Biochem. 95, 465-475) and thus may represent a new class of acid proteases. Also absent is the characteristic active site aspartyl sequence of aspartic proteases. There are 11 potential N-glycosylation sites on each thermopsin molecule. The molecular weight estimated from gel filtration (45,000) is larger than that calculated from the sequence (32,651), suggesting that thermopsin is the sequence (32,651), suggesting that thermopsin is glycosylated at at least some of these 11 sites.

Linton, D., E. Allan, et al. (2002). "Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in Campylobacter jejuni." Mol Microbiol **43**(2): 497-508.

It was demonstrated recently that there is a system of general protein glycosylation in the human enteropathogen Campylobacter jejuni. To characterize such glycoproteins, we identified a lectin, Soybean agglutinin (SBA), which binds to multiple C. jejuni proteins on Western blots. Binding of lectin SBA was disrupted by mutagenesis of genes within the previously identified protein

glycosylation locus. This lectin was used to purify putative glycoproteins selectively and, after sodium dodecyl sulphatepolyacrylamide gel electrophoresis (SDS-PAGE), Coomassie-stained bands were cut from the gels. The bands were digested with trypsin, and peptides were identified by mass spectrometry and database searching. A 28kDa band was identified as PEB3, a previously characterized immunogenic cell surface protein. Bands of 32 and 34kDa were both identified as a putative periplasmic protein encoded by the C. jejuni NCTC 11168 coding sequence Cj1670c. We have named this putative glycoprotein CgpA. We constructed insertional knockout mutants of both the peb3 and cgpA genes, and surface protein extracts from mutant and wild-type strains were analysed by one- and two-dimensional polyacrylamide gel electrophoresis (PAGE). In this way, we were able to identify the PEB3 protein as a 28 kDa SBA-reactive and immunoreactive glycoprotein. The cgpA gene encoded SBA-reactive and immunoreactive proteins of 32 and 34 kDa. By using specific exoglycosidases, we demonstrated that the SBA binding property of acid-glycine extractable C. jejuni glycoproteins, including PEB3 and CgpA, is a result of the presence of alpha-linked N-acetylgalactosamine residues. These data confirm the existence, and extend the boundaries, of the previously identified protein glycosylation locus of C. jejuni. Furthermore, we have identified two such glycoproteins, the first non-flagellin campylobacter glycoproteins to be identified, and demonstrated that their glycan components contain alpha-linked N-acetylgalactosamine residues.

Linton, D., A. V. Karlyshev, et al. (2000). "Multiple N-acetyl neuraminic acid synthetase (neuB) genes in Campylobacter jejuni: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide." Mol Microbiol **35**(5): 1120-34.

N-acetyl neuraminic acid (NANA) is a common constituent of Campylobacter jejuni lipo-oligosaccharide (LOS). Such structures often mimic human gangliosides and are thought to be involved in the triggering of Guillain-Barre syndrome (GBS) and Miller-Fisher syndrome (MFS) following C. jejuni infection. Analysis of the C. jejuni NCTC 11168 genome sequence identified three putative NANA synthetase genes termed neuB1, neuB2 and neuB3. The NANA synthetase activity of all three C. jejuni neuB gene products was confirmed by complementation experiments in an Escherichia coli neuB-deficient strain. Isogenic mutants were created in all three neuB genes, and for one such mutant (neuB1) LOS was shown to have increased mobility. C. jejuni NCTC 11168 wild-type LOS bound cholera toxin, indicating the presence of NANA in a LOS structure mimicking the ganglioside GM1. This property was lost in the neuB1 mutant. Gas chromatography-mass spectrometry and fast atom bombardment-mass spectrometry analysis of LOS from wild-type and the neuB1 mutant strain demonstrated the lack of NANA in the latter. Expression of the neuB1 gene in E. coli confirmed that NeuB1 was capable of in vitro NANA biosynthesis through condensation of N-acetyl-D-mannosamine and phosphoenolpyruvate. Southern analysis demonstrated that the neuB1 gene was confined to strains of C. jejuni with LOS containing a single NANA residue. Mutagenesis of neuB2 and neuB3 did not affect LOS, but neuB3 mutants were aflagellate and non-motile. No

phenotype was evident for neuB2 mutants in strain NCTC 11168, but for strain G1 the flagellin protein from the neuB2 mutant showed an apparent reduction in molecular size relative to the wild type. Thus, the neuB genes of C. jejuni appear to be involved in the biosynthesis of at least two distinct surface structures: LOS and flagella.

Lloyd, R. C., B. G. Davis, et al. (2000). "Site-selective glycosylation of subtilisin Bacillus lentus causes dramatic increases in esterase activity." Bioorg Med Chem **8**(7): 1537-44.
Using site directed mutagenesis combined with chemical modification, we have developed a general and versatile method for the glycosylation of proteins which is virtually unlimited in the scope of proteins and glycans that may be conjugated and in which the site of glycosylation and the nature of the introduced glycan can be carefully controlled. We have demonstrated the applicability of this method through the synthesis of a library of 48 glycosylated forms of the serine protease subtilisin Bacillus lentus (SBL) as single, pure species. As part of our ongoing program to tailor the activity of SBL for use in peptide synthesis, we have screened these enzymes for activity against the esterase substrate succinyl-Ala-Ala-Pro-Phe-S-benzyl. Gratifyingly, 22 enzymes displayed greater than wild type (WT) activity. Glycosylation at positions 62, in the S2 pocket, resulted in five glycosylated forms of SBL that were 1.3- to 1.9-fold more active than WT. At position 217, in the S1' pocket, all glycosylations increased kcat/KM up to a remarkable 8.4-fold greater than WT for the glucosylated enzyme L217C-S-beta-Glc(Ac)3. Furthermore, the ratio of amidase to esterase activity, (kcat/KM)esterase/(kcat/KM)amidase (E/A), is increased relative to wild type for all 48 glycosylated forms of SBL. Again, the most dramatic changes are observed at positions 62 and 217 and L217C-S-beta-Glc(Ac)3 has an E/A that is 17.2-fold greater than WT. The tailored specificity and high activity of this glycoform can be rationalized by molecular modeling analysis, which suggests that the carbohydrate moiety occupies the S1' leaving group pocket and enhances the rate of deacylation of the acyl-enzyme intermediate. These glycosylated enzymes are ideal candidates for use as catalysts in peptide synthesis as they have greatly increased (kcat,KM)esterase and severely reduced (kcat/KM)amidase and will favor the formation of the amide bond over hydrolysis.

Logan, S. M. (2006). "Flagellar glycosylation - a new component of the motility repertoire?" Microbiology **152**(Pt 5): 1249-62.
The biosynthesis, assembly and regulation of the flagellar apparatus has been the subject of extensive studies over many decades, with considerable attention devoted to the peritrichous flagella of Escherichia coli and Salmonella enterica. The characterization of flagellar systems from many other bacterial species has revealed subtle yet distinct differences in composition, regulation and mode of assembly of this important subcellular structure. Glycosylation of the major structural protein, the flagellin, has been shown most recently to be an important component of numerous flagellar systems in both Archaea and Bacteria, playing either an integral role in assembly or for a number of bacterial pathogens a role

in virulence. This review focuses on the structural diversity in flagellar glycosylation systems and demonstrates that as a consequence of the unique assembly processes, the type of glycosidic linkage found on archaeal and bacterial flagellins is distinctive.

Logan, S. M., J. P. Hui, et al. (2009). "Identification of novel carbohydrate modifications on Campylobacter jejuni 11168 flagellin using metabolomics-based approaches." FEBS J **276**(4): 1014-23.

It is well known that the flagellin of Campylobacter jejuni is extensively glycosylated by pseudaminic acid and the related acetamindino derivative, in addition to flagellin glycosylation being essential for motility and colonization of host cells. Recently, the use of metabolomics permitted the unequivocal characterization of unique flagellin modifications in Campylobacter, including novel legionaminic acid sugars in Campylobacter coli, which had been impossible to ascertain in earlier studies using proteomics-based approaches. To date, the precise identities of the flagellin glycosylation modifications have only been elucidated for C. jejuni 81-176 and C. coli VC167 and those present in the first genome-sequenced strain C. jejuni 11168 remain elusive due to lability and respective levels of individual glycan modifications. We report the characterization of the carbohydrate modifications on C. jejuni 11168 flagellin using metabolomics-based approaches. Detected as their corresponding CMP-linked precursors, structural information on the flagellin modifications was obtained using a combination of MS and NMR spectroscopy. In addition to the pseudaminic acid and legionaminic acid sugars known to be present on Campylobacter flagellin, two unusual 2,3-di-O-methylglyceric acid modifications of a nonulosonate sugar were identified. By performing a metabolomic analysis of selected isogenic mutants of genes from the flagellin glycosylation locus of this pathogen, these novel CMP-linked precursors were confirmed to be di-O-methylglyceric acid derivatives of pseudaminic acid and the related acetamidino sugar. This is the first comprehensive analysis of the flagellar modifications in C. jejuni 11168 and structural elucidation of di-O-methylglyceric acid derivatives of pseudaminic acid on Campylobacter flagellin.

Logan, S. M., J. F. Kelly, et al. (2002). "Structural heterogeneity of carbohydrate modifications affects serospecificity of Campylobacter flagellins." Mol Microbiol **46**(2): 587-97.

Flagellin from Campylobacter coli VC167 is post-translationally modified at > or = 16 amino acid residues with pseudaminic acid and three related derivatives. The predominant modification was 5,7-diacetamido-3,5,7,9 - tetradeoxy - l - glycero - l - manno - nonulosonic acid (pseudaminic acid, Pse5Ac7Ac), a modification that has been described previously on flagellin from Campylobacter jejuni 81-176. VC167 lacked two modi-fications present in 81-176 and instead had two unique modifications of masses 431 and 432 Da. Flagellins from both C. jejuni 81-176 and C. coli VC167 were also modified with an acetamidino form of pseudaminic acid (PseAm), but tandem mass spectrometry indicated that the structure of PseAm differed in the two strains. Synthesis of PseAm in C. coli VC167 requires

a minimum of six ptm genes. In contrast, PseAm is synthesized in C. jejuni 81-176 via an alternative pathway using the product of the pseA gene. Mutation of the ptm genes in C. coli VC167 can be detected by changes in apparent Mr of flagellin in SDS-PAGE gels, changes in isoelectric focusing (IEF) patterns and loss of immunoreactivity with antiserum LAH2. These changes corresponded to loss of both 315 Da and 431 Da modifications from flagellin. Complementation of the VC167 ptm mutants with the 81-176 pseA gene in trans resulted in flagellins containing both 315 and 431 Da modifications, but these flagellins remained unreactive in LAH2 antibody, suggesting that the unique form of PseAm encoded by the ptm genes contributes to the serospecificity of the flagellar filament.

Logan, S. M., T. J. Trust, et al. (1989). "Evidence for posttranslational modification and gene duplication of Campylobacter flagellin." J Bacteriol **171**(6): 3031-8.

A gene encoding a flagellin protein of Campylobacter coli VC167 has been cloned and sequenced. The gene was identified in a pBR322 library by hybridization to a synthetic oligonucleotide probe corresponding to amino acids 4 to 9 of the N-terminal sequence obtained by direct chemical analysis (S. M. Logan, L. A. Harris, and T. J. Trust, J. Bacteriol. 169:5072-5077, 1987). The DNA was sequenced and shown to contain an open reading frame encoding a protein with a molecular weight of 58,945 and a length of 572 amino acids. The deduced amino acid sequence was identical to the published N-terminal amino acid sequence of VC167 flagellin and to four internal regions whose partial sequences were obtained by direct chemical analysis of two tryptic and two cyanogen bromide peptides of VC167 flagellin. The C. coli flagellin protein contains posttranslationally modified serine residues, most of which occur within a region containing two 9-amino-acid repeating peptides separated by 34 unique amino acids. Comparisons with the sequences of flagellins from other bacterial species revealed conserved residues at the amino- and carboxy-terminal regions. Hybridization data suggest the presence of a second flagellin copy located adjacent to the first on the VC167 chromosome.

Lupas, A., H. Engelhardt, et al. (1994). "Domain structure of the Acetogenium kivui surface layer revealed by electron crystallography and sequence analysis." J Bacteriol **176**(5): 1224-33.

The three-dimensional structure of the Acetogenium kivui surface layer (S-layer) has been determined to a resolution of 1.7 nm by electron crystallographic techniques. Two independent reconstructions were made from layers negatively stained with uranyl acetate and Na-phosphotungstate. The S-layer has p6 symmetry with a center-to-center spacing of approximately 19 nm. Within the layer, six monomers combine to form a ring-shaped core surrounded by a fenestrated rim and six spokes that point towards the axis of threefold symmetry and provide lateral connectivity to other hexamers in the layer. The structure of the A. kivui S-layer protein is very similar to that of the Bacillus brevis middle wall protein, with which it shares an N-terminal domain of homology. This domain is found in several other extracellular proteins, including the S-layer proteins from Bacillus sphaericus and Thermus thermophilus, Omp alpha from Thermotoga

maritima, an alkaline cellulase from Bacillus strain KSM-635, and xylanases from Clostridium thermocellum and Thermoanaerobacter saccharolyticum, and may serve to anchor these proteins to the peptidoglycan. To our knowledge, this is the first example of a domain conserved in several S-layer proteins.

Magidovich, H., S. Yurist-Doutsch, et al. "AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of Haloferax volcanii." Mol Microbiol **76**(1): 190-9.

While pathways for N-glycosylation in Eukarya and Bacteria have been solved, considerably less is known of this post-translational modification in Archaea. In the halophilic archaeon Haloferax volcanii, proteins encoded by the agl genes are involved in the assembly and attachment of a pentasaccharide to select asparagine residues of the S-layer glycoprotein. AglP, originally identified based on the proximity of its encoding gene to other agl genes whose products were shown to participate in N-glycosylation, was proposed, based on sequence homology, to serve as a methyltransferase. In the present report, gene deletion and mass spectrometry were employed to reveal that AglP is responsible for adding a 14 Da moiety to a hexuronic acid found at position four of the pentasaccharide decorating the Hfx. volcanii S-layer glycoprotein. Subsequent purification of a tagged version of AglP and development of an in vitro assay to test the function of the protein confirmed that AglP is a S-adenosyl-L-methionine-dependent methyltransferase.

Maita, N., J. Nyirenda, et al. "Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases." J Biol Chem **285**(7): 4941-50.

Oligosaccharyltransferase (OST) catalyzes the transfer of an oligosaccharide from a lipid donor to an asparagine residue in nascent polypeptide chains. In the bacterium Campylobacter jejuni, a single-subunit membrane protein, PglB, catalyzes N-glycosylation. We report the 2.8 A resolution crystal structure of the C-terminal globular domain of PglB and its comparison with the previously determined structure from the archaeon Pyrococcus AglB. The two distantly related oligosaccharyltransferases share unexpected structural similarity beyond that expected from the sequence comparison. The common architecture of the putative catalytic sites revealed a new catalytic motif in PglB. Site-directed mutagenesis analyses confirmed the contribution of this motif to the catalytic function. Bacterial PglB and archaeal AglB constitute a protein family of the catalytic subunit of OST along with STT3 from eukaryotes. A structure-aided multiple sequence alignment of the STT3/PglB/AglB protein family revealed three types of OST catalytic centers. This novel classification will provide a useful framework for understanding the enzymatic properties of the OST enzymes from Eukarya, Archaea, and Bacteria.

Marceau, M., K. Forest, et al. (1998). "Consequences of the loss of O-linked glycosylation of meningococcal type IV pilin on piliation and pilus-mediated adhesion." Mol Microbiol **27**(4): 705-15.

Pili, which are assembled from protein subunits called pilin, are indispensable for

the adhesion of capsulated Neisseria meningitidis (MC) to eukaryotic cells. Both MC and Neisseria gonorrhoeae (GC) pilins are glycosylated, but the effect of this modification is unknown. In GC, a galactose alpha-1,3-N-acetyl glucosamine is O-linked to Ser-63, whereas in MC, an O-linked trisaccharide is present between residues 45 and 73 of pilin. As Ser-63 was found to be conserved in pilin variants from different strains, it was replaced by Ala in two MC variants to test the possible role of this residue in pilin glycosylation and modulation of pili function. The mutated alleles were stably expressed in MC, and the proteins they encoded migrated more quickly than the normal protein during SDS-PAGE. As controls, neighbouring Asn-61 and Ser-62 were replaced by an Ala with no effect on electrophoretic mobility. Silver staining of purified pilin obtained from MC after oxidation with periodic acid confirmed the loss of glycosylation in the Ser-63-->Ala pilin variants. Mass spectrometry of HPLC-purified trypsin-digested peptides of pilin and Ser-63-->Ala pilin confirmed that peptide 45-73 has the molecular size of a glycopeptide in the wild type. In strains producing non-glycosylated pilin variants, we observed that (i) no truncated S pilin monomer was produced; (ii) piliation was slightly increased; and (iii) presumably as a consequence, adhesiveness for epithelial cells was increased 1.6- to twofold in these derivatives. In addition, pilin monomers and/or individual pilus fibres, obtained after solubilization of a crude pili preparation in a high pH buffer, were reassociated into insoluble aggregates of pili more completely with non-glycosylated variants than with the normal pilin. Taken together, these data eliminate a major role for pilin glycosylation in piliation and subsequent pilus-mediated adhesion, but they demonstrate that glycosylation facilitates solubilization of pilin monomers and/or individual pilus fibres.

McNally, D. J., A. J. Aubry, et al. (2007). "Targeted metabolomics analysis of Campylobacter coli VC167 reveals legionaminic acid derivatives as novel flagellar glycans." J Biol Chem **282**(19): 14463-75.

Glycosylation of Campylobacter flagellin is required for the biogenesis of a functional flagella filament. Recently, we used a targeted metabolomics approach using mass spectrometry and NMR to identify changes in the metabolic profile of wild type and mutants in the flagellar glycosylation locus, characterize novel metabolites, and assign function to genes to define the pseudaminic acid biosynthetic pathway in Campylobacter jejuni 81-176 (McNally, D. J., Hui, J. P., Aubry, A. J., Mui, K. K., Guerry, P., Brisson, J. R., Logan, S. M., and Soo, E. C. (2006) J. Biol. Chem. 281, 18489-18498). In this study, we use a similar approach to further define the glycome and metabolomic complement of nucleotide-activated sugars in Campylobacter coli VC167. Herein we demonstrate that, in addition to CMP-pseudaminic acid, C. coli VC167 also produces two structurally distinct nucleotide-activated nonulosonate sugars that were observed as negative ions at m/z 637 and m/z 651 (CMP-315 and CMP-329). Hydrophilic interaction liquid chromatography-mass spectrometry yielded suitable amounts of the pure sugar nucleotides for NMR spectroscopy using a cold probe. Structural analysis in conjunction with molecular modeling identified the sugar moieties as acetamidino and N-methylacetimidoyl derivatives of

legionaminic acid (Leg5Am7Ac and Leg5AmNMe7Ac). Targeted metabolomic analyses of isogenic mutants established a role for the ptmA-F genes and defined two new ptm genes in this locus as legionaminic acid biosynthetic enzymes. This is the first report of legionaminic acid in Campylobacter sp. and the first report of legionaminic acid derivatives as modifications on a protein.

McNally, D. J., J. P. Hui, et al. (2006). "Functional characterization of the flagellar glycosylation locus in Campylobacter jejuni 81-176 using a focused metabolomics approach." J Biol Chem **281**(27): 18489-98.

Bacterial genome sequencing has provided a wealth of genetic data. However, the definitive functional characterization of hypothetical open reading frames and novel biosynthetic genes remains challenging. This is particularly true for genes involved in protein glycosylation because the isolation of their glycan moieties is often problematic. We have developed a focused metabolomics approach to define the function of flagellin glycosylation genes in Campylobacter jejuni 81-176. A capillary electrophoresis-electrospray mass spectrometry and precursor ion scanning method was used to examine cell lysates of C. jejuni 81-176 for sugar nucleotides. Novel nucleotide-activated intermediates of the pseudaminic acid (Pse5NAc7NAc) pathway and its acetamidino derivative (PseAm) were found to accumulate within select isogenic mutants, and use of a hydrophilic interaction liquid chromatography-mass spectrometry method permitted large scale purifications of the intermediates. NMR with cryo probe (cold probe) technology was utilized to complete the structural characterization of microgram quantities of CMP-5-acetamido-7-acetamidino-3,5,7,9-tetradeoxy-L-glycero-alpha-L-manno-n onulosonic acid (CMP-Pse5NAc7Am), which is the first report of Pse modified at C7 with an acetamidino group in Campylobacter, and UDP-2,4-diacetamido-2,4,6-trideoxy-alpha-D-glucopyranose, which is a bacillosamine derivative found in the N-linked proteinglycan. Using this focused metabolomics approach, pseB, pseC, pseF, pseI, and for the first time pseA, pseG, and pseH were found to be directly involved in either the biosynthesis of CMP-Pse5NAc7NAc or CMP-Pse5NAc7Am. In contrast, it was shown that pseD, pseE, Cj1314c, Cj1315c, Cjb1301, Cj1334, Cj1341c, and Cj1342c have no role in the CMP-Pse5NAc7NAc or CMP-Pse5NAc7Am pathways. These results demonstrate the usefulness of this approach for targeting compounds within the bacterial metabolome to assign function to genes, identify metabolic intermediates, and elucidate novel biosynthetic pathways.

Mengele, R. and M. Sumper (1992). "Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles." J Biol Chem **267**(12): 8182-5.

The outer surface of the moderate halophilic archaebacterium Haloferax volcanii (formerly named Halobacterium volcanii) is covered with a hexagonally packed surface (S) layer glycoprotein. The polypeptide (794 amino acid residues) contains 7 N-glycosylation sites. Four of these sites were isolated as glycopeptides and the structure of one of the corresponding saccharides was determined. Oligosaccharides consisting of beta-1,4-linked glucose residues are

attached to the protein via the linkage unit asparaginyl-glucose. In the related glycoprotein from the extreme halophile Halobacterium halobium, the glucose residues are replaced by sulfated glucuronic acid residues, causing a drastic increase in surface charge density. This is discussed in terms of a recent model explaining the stability of halophilic proteins.

Mescher, M. F. and J. L. Strominger (1976). "Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium." J Biol Chem **251**(7): 2005-14.

The glycoprotein which accounts for approximately 50% of the protein and all of the nonlipid carbohydrate of the cell envelope of Halobacterium salinarium (Mescher, M. F., Strominger, J. L., and Watson S. W. (1974) J. Bacteriol. 120, 945-954) has been purified and partially characterized. The glycoprotein has an apparent molecular weight of 200,000, is extremely acidic, and has a carbohydrate content of approximately 10 to 12%. The carbohydrate included neutral hexoses, amino sugar, and uronic acid. Information regarding the number, composition, and mode of attachment of the carbohydrate chains was obtained by isolation and examination of the glycopeptides derived from degradation of cell envelope protein with trypsin and pronase. Trypsin digestion resulted in two glycopeptides. One of these was large (approximately 55,000 daltons) and had most of the neutral hexose linked to it. The carbohydrate moieties consisted of di- and trisaccharides of glucosylgalactose and (uronic acid, glucose)-galactose attached via O-glycosidic linkages between galactose and threonine. The other tryptic glycopeptide had a relatively large heterosaccharide attached to it via an alkaline-stable linkage. The heterosaccharide contained 1 glucose, 8 to 9 galactose, 1 mannose, and 10 to 11 glucosamine residues, and approximately 6 residues of an unidentified amino augar. The alkaline stability of the linkage and the amino acid composition of glycopeptides resulting from Pronase digestion of the tryptic glycopeptide showed that the heterosaccharide was attached to an asparagine residue, presumably via an N-glycosylamine bond to the amide group. The intact glycoprotein has a single N-linked heterosaccharide, 22 to 24 O-linked disaccharides, and 12 to 14 O-linked trisaccharides per molecule. N- and O-glycosidic linkages are the most common carbohydrate-protein linkages in mammalian glycoproteins but, to our knowledge, this is the first report of either type of linkage in a prokaryotic cell envelope protein.

Mescher, M. F., J. L. Strominger, et al. (1974). "Protein and carbohydrate composition of the cell envelope of Halobacterium salinarium." J Bacteriol **120**(2): 945-54.

The isolated cell envelope of Halobacterium salinarium strain 1 contained 15 to 20 proteins that were resolved by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. All but one of these proteins had molecular weights of 130,000 or less and together accounted for 50 to 60% of the total envelope protein. The remaining 40 to 50% of the envelope protein was accounted for by a single protein with an apparent molecular weight of approximately 194,000 that stained for carbohydrate with periodate-Schiff

reagent. The proteolytic enzymes trypsin and Pronase were used to show that the carbohydrate is covalently bound to the protein. Separation of amino sugar- and hexose-containing tryptic peptides by gel filtration indicated that all of the nonlipid carbohydrate of the cell envelope is covalently bound to protein. The results of partial purification by phenol extraction indicated that both the amino sugar and hexose are bound to the 194,000-molecular-weight protein. Exposure of isolated cell envelopes to low salt concentration resulted in solubilization of a majority of the envelope proteins. A relatively small number of proteins, including the high-molecular-weight, carbohydrate-containing protein, remained bound to the sedimentable cell membrane fraction.

Messner, P. (1997). "Bacterial glycoproteins." Glycoconj J **14**(1): 3-11.
Glycoproteins are a diverse group of complex macromolecules that are present in virtually all forms of life. Their presence in prokaryotes, however, has been demonstrated, and accepted, only recently. Bacterial glycoproteins have been identified in many archaeobacteria and in eubacteria. They comprise a wide range of different cell envelope components such as membrane-associated glycoproteins, surface-associated glycoproteins and crystalline surface layers (S-layers), as well as secreted glycoproteins and exoenzymes. Even their occurrence in the cytoplasm cannot yet be ruled out. This minireview tries to cover the whole subject as completely as possible and refers to available information on presence, structure, biosynthesis, and molecular biology of bacterial glycoproteins.

Messner, P. and C. Schaffer (2003). "Prokaryotic glycoproteins." Fortschr Chem Org Naturst **85**: 51-124.

Messner, P., K. Steiner, et al. (2008). "S-layer nanoglycobiology of bacteria." Carbohydr Res **343**(12): 1934-51.
Cell surface layers (S-layers) are common structures of the bacterial cell envelope with a lattice-like appearance that are formed by a self-assembly process. Frequently, the constituting S-layer proteins are modified with covalently linked glycan chains facing the extracellular environment. S-layer glycoproteins from organisms of the Bacillaceae family possess long, O-glycosidically linked glycans that are composed of a great variety of sugar constituents. The observed variations already exceed the display found in eukaryotic glycoproteins. Recent investigations of the S-layer protein glycosylation process at the molecular level, which has lagged behind the structural studies due to the lack of suitable molecular tools, indicated that the S-layer glycoprotein glycan biosynthesis pathway utilizes different modules of the well-known biosynthesis routes of lipopolysaccharide O-antigens. The genetic information for S-layer glycan biosynthesis is usually present in S-layer glycosylation (slg) gene clusters acting in concert with housekeeping genes. To account for the nanometer-scale cell surface display feature of bacterial S-layer glycosylation, we have coined the neologism 'nanoglycobiology'. It includes structural and biochemical aspects of S-layer glycans as well as molecular data on the machinery underlying the

glycosylation event. A key aspect for the full potency of S-layer nanoglycobiology is the unique self-assembly feature of the S-layer protein matrix. Being aware that in many cases the glycan structures associated with a protein are the key to protein function, S-layer protein glycosylation will add a new and valuable component to an 'S-layer based molecular construction kit'. In our long-term research strategy, S-layer nanoglycobiology shall converge with other functional glycosylation systems to produce 'functional' S-layer neoglycoproteins for diverse applications in the fields of nanobiotechnology and vaccine technology. Recent advances in the field of S-layer nanoglycobiology have made our overall strategy a tangible aim of the near future.

Michel, G., K. Pojasek, et al. (2004). "The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery." J Biol Chem **279**(31): 32882-96.

Chondroitinase B from Pedobacter heparinus is the only known enzyme strictly specific for dermatan sulfate and is a widely used enzymatic tool for the structural characterization of glycosaminoglycans. This beta-helical polysaccharide lyase belongs to family PL-6 and cleaves the beta(1,4) linkage of dermatan sulfate in a random manner, yielding 4,5-unsaturated dermatan sulfate disaccharides as the product. The previously reported structure of its complex with a dermatan sulfate disaccharide product identified the -1 and -2 subsites of the catalytic groove. We present here the structure of chondroitinase B complexed with several dermatan sulfate and chondroitin sulfate oligosaccharides. In particular, the soaking of chondroitinase B crystals with a dermatan sulfate hexasaccharide results in a complex with two dermatan sulfate disaccharide reaction products, enabling the identification of the +2 and +1 subsites. Unexpectedly, this structure revealed the presence of a calcium ion coordinated by sequence-conserved acidic residues and by the carboxyl group of the l-iduronic acid at the +1 subsite. Kinetic and site-directed mutagenesis experiments have subsequently demonstrated that chondroitinase B absolutely requires calcium for its activity, indicating that the protein-Ca(2+)-oligosaccharide complex is functionally relevant. Modeling of an intact tetrasaccharide in the active site of chondroitinase B provided a better understanding of substrate specificity and the role of Ca(2+) in enzymatic activity. Given these results, we propose that the Ca(2+) ion neutralizes the carboxyl moiety of the l-iduronic acid at the cleavage site, whereas the conserved residues Lys-250 and Arg-271 act as Bronsted base and acid, respectively, in the lytic degradation of dermatan sulfate by chondroitinase B.

Michell, S. L., A. O. Whelan, et al. (2003). "The MPB83 antigen from Mycobacterium bovis contains O-linked mannose and (1-->3)-mannobiose moieties." J Biol Chem **278**(18): 16423-32.

Mycobacterium tuberculosis and Mycobacterium bovis, the causative agents of human and bovine tuberculosis, have been reported to express a range of surface and secreted glycoproteins, although only one of these has been subjected to detailed structural analysis. We describe the use of a genetic system, in conjunction with lectin binding, to characterize the points of

attachment of carbohydrate moieties to the polypeptide backbone of a second mycobacterial glycoprotein, antigen MPB83 from M. bovis. Biochemical and structural analysis of the native MPB83 protein and derived peptides demonstrated the presence of 3 mannose units attached to two threonine residues. Mannose residues were joined by a (1 --> 3) linkage, in contrast to the (1 --> 2) linkage previously observed in antigen MPT32 from M. tuberculosis and the (1 --> 2) and (1 --> 6) linkages in other mycobacterial glycolipids and polysaccharides. The identification of glycosylated antigens within the M. tuberculosis complex raises the possibility that the carbohydrate moiety of these glycoproteins might be involved in pathogenesis, either by interaction with mannose receptors on host cells, or as targets or modulators of the cell-mediated immune response. Given such a possibility characterization of mycobacterial glycoproteins is a step toward understanding their functional role and elucidating the mechanisms of mycobacterial glycosylation.

Miller, W. L., M. J. Matewish, et al. (2008). "Flagellin glycosylation in Pseudomonas aeruginosa PAK requires the O-antigen biosynthesis enzyme WbpO." J Biol Chem **283**(6): 3507-18.

Pseudomonas aeruginosa PAK (serotype O6) produces a single polar, glycosylated flagellum composed of a-type flagellin. To determine whether or not flagellin glycosylation in this serotype requires O-antigen genes, flagellin was isolated from the wild type, three O-antigen-deficient mutants wbpL, wbpO, and wbpP, and a wbpO mutant complemented with a plasmid containing a wild-type copy of wbpO. Flagellin from the wbpO mutant was smaller (42 kDa) than that of the wild type (45 kDa), or other mutants strains, and exhibited an altered isoelectric point (pI 4.8) when compared with PAK flagellin (pI 4.6). These differences were because of the truncation of the glycan moiety in the wbpO- flagellin. Thus, flagellin glycosylation in P. aeruginosa PAK apparently requires a functional WbpO but not WbpP. Because WbpP was previously proposed to catalyze a metabolic step in the biosynthesis of B-band O-antigen that precedes the action of WbpO, these results prompted us to reevaluate the two-step pathway catalyzed by WbpO and WbpP. Results from WbpO-WbpP-coupled enzymatic assays showed that either WbpO or WbpP is capable of initiating the two-step pathway; however, the kinetic parameters favored the WbpO reaction to occur first, converting UDP-N-acetyl-D-glucosamine to UDP-N-acetyl-D-glucuronic acid prior to the conversion to UDP-N-acetyl-D-galacturonic acid by WbpP. This is the first report to show that a C4 epimerase could utilize UDP-N-acetylhexuronic acid as a substrate.

Moormann, C., I. Benz, et al. (2002). "Functional substitution of the TibC protein of enterotoxigenic Escherichia coli strains for the autotransporter adhesin heptosyltransferase of the AIDA system." Infect Immun **70**(5): 2264-70.

The plasmid-encoded AIDA (adhesin involved in diffuse adherence) autotransporter protein derived from diffuse-adhering clinical Escherichia coli isolate 2787 and the TibA (enterotoxigenic invasion locus B) protein encoded by the chromosomal tib locus of enterotoxigenic E. coli (ETEC) strain H10407 are

posttranslationally modified by carbohydrate substituents. Analysis of the AIDA-I adhesin showed that the modification involved heptose residues. AIDA-I is modified by the heptosyltransferase activity of the product of the aah gene, which is located directly upstream of adhesin-encoding gene aidA. The carbohydrate modification of the TibA adhesin/invasin is mediated by the TibC protein but has not been elucidated. Based on the sequence similarities between TibC and AAH (autotransporter adhesin heptosyltransferase) and between the TibA and the AIDA proteins we hypothesized that the AIDA system and the Tib system encoded by the tib locus are structurally and functionally related. Here we show that (i) TibC proteins derived from different ETEC strains appear to be highly conserved, (ii) recombinant TibC proteins can substitute for the AAH heptosyltransferase in introducing the heptosyl modification to AIDA-I, (iii) this modification is functional in restoring the adhesive function of AIDA-I, (iv) a single amino acid substitution at position 358 completely abolishes this activity, and (v) antibodies directed at the functionally active AIDA-I recognize a protein resembling modified TibA in ETEC strains. In summary, we conclude that, like AAH, TibC represents an example of a novel class of heptosyltransferases specifically transferring heptose residues onto multiple sites of a protein backbone. A potential consensus sequence for the modification site is suggested.

Muir, E. M., I. Fyfe, et al. "Modification of N-glycosylation sites allows secretion of bacterial chondroitinase ABC from mammalian cells." J Biotechnol **145**(2): 103-10.
Although many eukaryotic proteins have been secreted by transfected bacterial cells, little is known about how a bacterial protein is treated as it passes through the secretory pathway when expressed in a eukaryotic cell. The eukaryotic N-glycosylation system could interfere with folding and secretion of prokaryotic proteins whose sequence has not been adapted for glycosylation in structurally appropriate locations. Here we show that such interference does indeed occur for chondroitinase ABC from the bacterium Proteus vulgaris, and can be overcome by eliminating potential N-glycosylation sites. Chondroitinase ABC was heavily glycosylated when expressed in mammalian cells or in a mammalian translation system, and this process prevented secretion of functional enzyme. Directed mutagenesis of selected N-glycosylation sites allowed efficient secretion of active chondroitinase. As these proteoglycans are known to inhibit regeneration of axons in the mammalian central nervous system, the modified chondroitinase gene is a potential tool for gene therapy to promote neural regeneration, ultimately in human spinal cord injury.

Nita-Lazar, M., M. Wacker, et al. (2005). "The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation." Glycobiology **15**(4): 361-7.
In the Gram-negative bacterium Campylobacter jejuni there is a pgl (protein glycosylation) locus-dependent general N-glycosylation system of proteins. One of the proteins encoded by pgl locus, PglB, a homolog of the eukaryotic oligosaccharyltransferase component Stt3p, is proposed to function as an oligosaccharyltransferase in this prokaryotic system. The sequence requirements

of the acceptor polypeptide for N-glycosylation were analyzed by reverse genetics using the reconstituted glycosylation of the model protein AcrA in Escherichia coli. As in eukaryotes, the N-X-S/T sequon is an essential but not a sufficient determinant for N-linked protein glycosylation. This conclusion was supported by the analysis of a novel C. jejuni glycoprotein, HisJ. Export of the polypeptide to the periplasm was required for glycosylation. Our data support the hypothesis that eukaryotic and bacterial N-linked protein glycosylation are homologous processes.

Nothaft, H. and C. M. Szymanski "Protein glycosylation in bacteria: sweeter than ever." Nat Rev Microbiol **8**(11): 765-78.

Investigations into bacterial protein glycosylation continue to progress rapidly. It is now established that bacteria possess both N-linked and O-linked glycosylation pathways that display many commonalities with their eukaryotic and archaeal counterparts as well as some unexpected variations. In bacteria, protein glycosylation is not restricted to pathogens but also exists in commensal organisms such as certain Bacteroides species, and both the N-linked and O-linked glycosylation pathways can modify multiple proteins. Improving our understanding of the intricacies of bacterial protein glycosylation systems should lead to new opportunities to manipulate these pathways in order to engineer glycoproteins with potential value as novel vaccines.

Novotny, R., C. Schaffer, et al. (2004). "S-layer glycan-specific loci on the chromosome of Geobacillus stearothermophilus NRS 2004/3a and dTDP-L-rhamnose biosynthesis potential of G. stearothermophilus strains." Microbiology **150**(Pt 4): 953-65.

The approximately 16.5 kb surface layer (S-layer) glycan biosynthesis (slg) gene cluster of the Gram-positive thermophile Geobacillus stearothermophilus NRS 2004/3a has been sequenced. The cluster is located immediately downstream of the S-layer structural gene sgsE and consists of 13 ORFs that have been identified by database sequence comparisons. The cluster encodes dTDP-L-rhamnose biosynthesis (rml operon), required for building up the polyrhamnan S-layer glycan, as well as for assembly and export of the elongated glycan chain, and its transfer to the S-layer protein. This is the first report of a gene cluster likely to be involved in the glycosylation of an S-layer protein. There is evidence that this cluster is transcribed as a polycistronic unit, whereas sgsE is transcribed monocistronically. To get insights into the regulatory mechanisms underlying glycosylation of the S-layer protein, the influence of growth temperature on the S-layer was investigated in seven closely related G. stearothermophilus strains, of which only strain NRS 2004/3a possessed a glycosylated S-layer. Chromosomal DNA preparations of these strains were screened for the presence of the rml operon, because L-rhamnose is a frequent constituent of S-layer glycans. From rml-positive strains, flanking regions of the operon were sequenced. Comparison with the slg gene cluster of G. stearothermophilus NRS 2004/3a revealed sequence homologies between adjacent genes. The temperature inducibility of S-layer protein glycosylation was investigated in those strains by raising the growth temperature from 55 degrees C to 67 degrees C; no change of either the

protein banding pattern or the glycan staining behaviour was observed on SDS-PAGE gels, although the sgsE transcript was several-fold more abundant at 67 degrees C. Cell-free extracts of the strains were capable of converting dTDP-D-glucose to dtdp-L-rhamnose. Taken together, the results indicate that the rml locus is highly conserved among G. stearothermophilus strains, and that in the investigated rml-containing strains, dTDP-L-rhamnose is actively synthesized in vitro. However, in contrast to previous reports for G. stearothermophilus wild-type strains, an increase in growth temperature did not switch an S-layer protein phenotype to an S-layer glycoprotein phenotype, via the de novo generation of a new S-layer gene sequence.

Olivier, N. B., M. M. Chen, et al. (2006). "In vitro biosynthesis of UDP-N,N'-diacetylbacillosamine by enzymes of the Campylobacter jejuni general protein glycosylation system." Biochemistry **45**(45): 13659-69.

In Campylobacter jejuni 2,4-diacetamido-2,4,6-trideoxy-alpha-d-glucopyranose, termed N,N'-diacetylbacillosamine (Bac2,4diNAc), is the first carbohydrate in the glycoprotein N-linked heptasaccharide. With uridine diphosphate-N-acetylglucosamine (UDP-GlcNAc) as a starting point, two enzymes of the general protein glycosylation (Pgl) pathway in C. jejuni (PglF and PglE) have recently been shown to modify this sugar nucleotide to form UDP-2-acetamido-4-amino-2,4,6-trideoxy-alpha-d-glycopyranose (UDP-4-amino-sugar) [Schoenhofen, I. C., et al. (2006) J. Biol. Chem. 281, 723-732]. PglD has been proposed to catalyze the final step in N,N'-diacetylbacillosamine synthesis by N-acetylation of the UDP-4-amino-sugar at the C4 position. We have cloned, overexpressed, and purified PglD from the pgl locus of C. jejuni NCTC 11168 and identified it as the acetyltransferase that modifies the UDP-4-amino-sugar to form UDP-N,N'-diacetylbacillosamine, utilizing acetyl-coenzyme A as the acetyl group donor. The UDP-N,N'-diacetylbacillosamine product was purified from the reaction by reverse phase C18 HPLC and the structure determined by NMR analysis. Additionally, the full-length PglF was overexpressed and purified in the presence of detergent as a GST fusion protein, allowing for derivation of kinetic parameters. We found that the UDP-4-amino-sugar was readily synthesized from UDP-GlcNAc in a coupled reaction using PglF and PglE. We also demonstrate the in vitro biosynthesis of the complete heptasaccharide lipid-linked donor by coupling the action of eight enzymes (PglF, PglE, PglD, PglC, PglA, PglJ, PglH, and PglI) in the Pgl pathway in a single reaction vessel.

Oman, T. J., J. M. Boettcher, et al. "Sublancin is not a lantibiotic but an S-linked glycopeptide." Nat Chem Biol **7**(2): 78-80.

Sublancin is shown to be an S-linked glycopeptide containing a glucose attached to a cysteine residue, establishing a new post-translational modification. The activity of the S-glycosyl transferase was reconstituted in vitro, and the enzyme is shown to have relaxed substrate specificity, allowing the preparation of analogs of sublancin. Glycosylation is essential for its antimicrobial activity.

Ozbek, S., J. F. Muller, et al. (2005). "Favourable mediation of crystal contacts by

cocoamidopropylbetaine (CAPB)." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 477-80.

Crystals of excellent quality are a prerequisite for high-resolution X-ray data. However, in refinement protocols of crystallization conditions it is often difficult to obtain the right combination of, for example, protein concentration, drop size, temperature and additives. A novel approach for optimizing crystal contacts in a most favourable fashion by performing crystallization setups with the zwitterionic surfactant cocoamidoproylbetaine (CAPB) is introduced. In the presence of this surfactant, highly diffracting crystals were obtained. Here, data from a right-handed coiled coil (RHCC) in complex with CAPB at 1.4 A resolution are presented. The addition of CAPB using otherwise identical crystallization conditions and the same X-ray source caused an improvement in resolution from 2.9 to 1.4 A.

Parge, H. E., K. T. Forest, et al. (1995). "Structure of the fibre-forming protein pilin at 2.6 A resolution." Nature **378**(6552): 32-8.

The crystallographic structure of Neisseria gonorrhoeae pilin, which assembles into the multifunctional pilus adhesion and virulence factor, reveals an alpha-beta roll fold with a striking 85 A alpha-helical spine and an O-linked disaccharide. Key residues stabilize interactions that allow sequence hypervariability, responsible for pilin's celebrated antigenic variation, within disulphide region beta-strands and connections. Pilin surface shape, hydrophobicity and sequence variation constrain pilus assembly to the packing of flat subunit faces against alpha 1 helices. Helical fibre assembly is postulated to form a core of coiled alpha 1 helices banded by beta-sheet, leaving carbohydrate and hypervariable sequence regions exposed to solvent.

Paul, G., F. Lottspeich, et al. (1986). "Asparaginyl-N-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan." J Biol Chem **261**(3): 1020-4.

The cell surface glycoprotein of Halobacteria contains two different types of sulfated saccharides: hexuronic acid-containing oligosaccharides linked to the protein via asparaginylglucose, and a serially repeated saccharide unit containing amino sugars that resembles the animal glycosaminoglycans. Here we report that 1) the sulfated repeating unit saccharide is linked to the cell surface glycoprotein via asparaginyl-N-acetylgalactosamine, 2) the amino acid sequence surrounding this linkage region is -Asn-Ala-Ser-, and thus in agreement with the acceptor sequence ASN-X-Thr(Ser) common to all eucaryotic N-glycosidically bound saccharides determined so far; 3) in addition to galactose, galacturonic acid, N-acetylglucosamine, and N-acetylgalactosamine, the methylated hexuronic acid 3-O-methylgalacturonic acid occurs as a stoichiometric constituent of the sulfated building block of the glycosaminoglycan chain.

Peters, J., W. Baumeister, et al. (1996). "Hyperthermostable surface layer protein tetrabrachion from the archaebacterium Staphylothermus marinus: evidence for the presence of a right-handed coiled coil derived from the primary structure." J Mol Biol **257**(5): 1031-41.

The scaffold of the surface layer covering the hyperthermophilic archaebacterium

Staphylothermus marinus is formed by an extended filiform glycoprotein complex, tetrabrachion, which is anchored in the cell membrane at one end of a 70 nm stalk and branches at the other end into four arms of 24 nm length. The arms from a canopy-like meshwork by end-to-end contacts, enclosing a "quasi-periplasmic space". The primary structure of the complex, obtained by an approach based entirely on the polymerase chain reaction, shows that the light and the heavy chains are encoded in this order in a single gene and are generated by internal proteolytic cleavage. One light chain associates with the N-terminal part of a heavy chain to form one of the four arms of the complex, comprising about 1000 residues. Following a glycine-rich linker of about ten residues, the C-terminal 500 residues of the four heavy chains converge to form a four-stranded parallel coiled coil, which ends in a transmembrane segment. The sequence of the coiled coil is exceptional in that the heptad repeat of hydrophobic residues typical for left-handed coiled coils shifts to an undecad repeat after an internal proline residue, indicating that the C-terminal part of the sequence forms a right-handed coiled coil. Such a periodicity has not been detected in coiled coils to date. The almost flawless pattern of aliphatic residues, mainly leucine and isoleucine, throughout the hydrophobic core of the stalk provide one explanation for its exceptional stability.

Peters, J., M. Nitsch, et al. (1995). "Tetrabrachion: a filamentous archaebacterial surface protein assembly of unusual structure and extreme stability." J Mol Biol **245**(4): 385-401.

The surface (S-) layer of the hyperthermophilic archaebacterium Staphylothermus marinus was isolated, dissected into separate domains by chemical and proteolytic methods, and analyzed by spectroscopic, electron microscopic and biochemical techniques. The S-layer is formed by a poorly ordered meshwork of branched, filiform morphological subunits resembling dandelion seed-heads. A morphological subunit (christened by us tetrabrachion) consists of a 70 nm long, almost perfectly straight stalk ending in four straight arms of 24 nm length that provide lateral connectivity by end-to-end contacts. At 32 nm from the branching point, tetrabrachion carries two globular particles of 10 nm diameter that have both tryptic and chymotryptic protease activity. Tetrabrachion is built by a tetramer of M(r) 92,000 polypeptides that form a parallel, four-stranded alpha-helical rod and separate at one end into four strands. These strands interact in a 1:1 stoichiometry with polypeptides of M(r) 85,000 to form the arms. The arms are composed entirely of beta-sheets. All S-layer components contain bound carbohydrates (glucose, mannose, and glucosamine) at a ratio of 38 g/100 g protein for the complete tetrabrachion-protease complex. The unique structure of tetrabrachion is reflected in an extreme thermal stability in the presence of strong denaturants (1% (w/v) SDS of 6M guanidine): the arms, which are stabilized by intramolecular disulphide bridges, melt around 115 degrees C under non-reducing conditions, whereas the stalk sustains heating up to about 130 degrees C. Complete denaturation of the stalk domain requires treatment with 70% (v/v) sulfuric acid or with fuming trifluoromethanesulfonic acid. The globular protease can be heated to 90

degrees C in 6M guanidine and to 120 degrees C in 1% SDS and represents one of the most stable proteases characterized to date.

Peters, J., S. Rudolf, et al. (1992). "Evidence for tyrosine-linked glycosaminoglycan in a bacterial surface protein." Biol Chem Hoppe Seyler **373**(4): 171-6.

The S-layer protein of Acetogenium kivui was subjected to proteolysis with different proteases and several high molecular mass glycosaminoglycan peptides containing glucose, galactosamine and an unidentified sugar-related component were separated by molecular sieve chromatography and reversed-phase HPLC and subjected to N-terminal sequence analysis. By methylation analysis glucose was found to be uniformly 1,6-linked, whereas galactosamine was exclusively 1,4-linked. Hydrazinolysis and subsequent amino-acid analysis as well as two-dimensional NMR spectroscopy were used to demonstrate that in these peptides carbohydrate was covalently linked to tyrosine. As all of the four Tyr-glycosylation sites were found to be preceded by valine, a new recognition sequence for glycosylation is suggested.

Peyfoon, E., B. Meyer, et al. "The S-layer glycoprotein of the crenarchaeote Sulfolobus acidocaldarius is glycosylated at multiple sites with chitobiose-linked N-glycans." Archaea **2010**.

Glycosylation of the S-layer of the crenarchaea Sulfolobus acidocaldarius has been investigated using glycoproteomic methodologies. The mature protein is predicted to contain 31 N-glycosylation consensus sites with approximately one third being found in the C-terminal domain spanning residues L(1004)-Q(1395). Since this domain is rich in Lys and Arg and therefore relatively tractable to glycoproteomic analysis, this study has focused on mapping its N-glycosylation. Our analysis identified nine of the 11 consensus sequence sites, and all were found to be glycosylated. This constitutes a remarkably high glycosylation density in the C-terminal domain averaging one site for each stretch of 30-40 residues. Each of the glycosylation sites observed was shown to be modified with a heterogeneous family of glycans, with the largest having a composition Glc(1)Man(2)GlcNAc(2) plus 6-sulfoquinovose (QuiS), consistent with the tribranched hexasaccharide previously reported in the cytochrome b(558/566) of S. acidocaldarius. S. acidocaldarius is the only archaeal species whose N-glycans are known to be linked via the chitobiose core disaccharide that characterises the N-linked glycans of Eukarya.

Pfoestl, A., A. Hofinger, et al. (2003). "Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-galactose in Aneurinibacillus thermoaerophilus L420-91T." J Biol Chem **278**(29): 26410-7.

The glycan chain of the S-layer protein of Aneurinibacillus thermoaerophilus L420-91T (DSM 10154) consists of d-rhamnose and 3-acetamido-3,6-dideoxy-d-galactose (d-Fucp3NAc). Thymidine diphosphate-activated d-Fucp3NAc serves as precursor for the assembly of structural polysaccharides in Gram-positive and Gram-negative organisms. The biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-d-galactose (dTDP-d-Fucp3NAc) involves five enzymes. The first two

steps of the reaction are catalyzed by enzymes that are part of the well studied dTDP-l-rhamnose biosynthetic pathway, namely d-glucose-1-phosphate thymidyltransferase (RmlA) and dTDP-d-glucose-4,6-dehydratase (RmlB). The enzymes catalyzing the last three synthesis reactions have not been characterized biochemically so far. These steps include an isomerase, a transaminase, and a transacetylase. We identified all five genes involved by chromosome walking in the Gram-positive organism A. thermoaerophilus L420-91T and overexpressed the three new enzymes heterologously in Escherichia coli. The activities of these enzymes were monitored by reverse phase high performance liquid chromatography, and the intermediate products formed were characterized by 1H and 13C nuclear magnetic resonance spectroscopy analysis. Alignment of the newly identified proteins with known sequences revealed that the elucidated pathway in this Gram-positive organism may also be valid in the biosynthesis of the O-antigen of lipopolysaccharides of Gram-negative organisms. The key enzyme in the biosynthesis of dTDP-d-Fucp3NAc has been identified as an isomerase, which converts the 4-keto educt into the 3-keto product, with concomitant epimerization at C-4 to produce a 6-deoxy-d-xylo configuration. This is the first report of the functional characterization of the biosynthesis of dTDP-d-Fucp3NAc and description of a novel type of isomerase capable of synthesizing dTDP-6-deoxy-d-xylohex-3-ulose from dTDP-6-deoxy-d-xylohex-4-ulose.

Plavner, N. and J. Eichler (2008). "Defining the topology of the N-glycosylation pathway in the halophilic archaeon Haloferax volcanii." J Bacteriol **190**(24): 8045-52.
In Eukarya, N glycosylation involves the actions of enzymes working on both faces of the endoplasmic reticulum membrane. The steps of bacterial N glycosylation, in contrast, transpire essentially on the cytoplasmic side of the plasma membrane, with only transfer of the assembled glycan to the target protein occurring on the external surface of the cell. For Archaea, virtually nothing is known about the topology of enzymes involved in assembling those glycans that are subsequently N linked to target proteins on the external surface of the cell. To remedy this situation, subcellular localization and topology predictive algorithms, protease accessibility, and immunoblotting, together with cysteine modification following site-directed mutagenesis, were enlisted to define the topology of Haloferax volcanii proteins experimentally proven to participate in the N-glycosylation process. AglJ and AglD, involved in the earliest and latest stages, respectively, of assembly of the pentasaccharide decorating the H. volcanii S-layer glycoprotein, were shown to present their soluble N-terminal domain, likely containing the putative catalytic site of each enzyme, to the cytosol. The same holds true for Alg5-B, Dpm1-A, and Mpg1-D, proteins putatively involved in this posttranslational event. The results thus point to the assembly of the pentasaccharide linked to certain Asn residues of the H. volcanii S-layer glycoprotein as occurring within the cell.

Plummer, T. H., Jr., A. L. Tarentino, et al. (1995). "Novel, specific O-glycosylation of secreted Flavobacterium meningosepticum proteins. Asp-Ser and Asp-Thr-Thr

consensus sites." J Biol Chem **270**(22): 13192-6.

A new type of O-linked oligosaccharide has been discovered on several proteins secreted by the Gram-negative bacterium Flavobacterium meningosepticum, including Endo F2 (three sites), Endo F3 (one site), and a P40 protease (one site). The oligosaccharide moiety is covalently attached via a mannose residue to a serine or threonine at consensus sites corresponding to Asp-Ser* or Asp-Thr*-Thr. Preliminary characterization by mass spectroscopy revealed an oligosaccharide of 1244 Da at each of the proposed glycosylation sites. Collision-associated dissociation analysis showed a characteristic daughter ion series of m/z 218, 394, and 556, indicative of a common Flavobacterium oligosaccharide. Compositional analysis demonstrated an unusual profile of monosaccharides, including hexoses, methylated hexoses, and uronic acid derivatives.

Power, P. M., L. F. Roddam, et al. (2000). "Genetic characterization of pilin glycosylation in Neisseria meningitidis." Microbiology **146 (Pt 4)**: 967-79.

Pili of Neisseria meningitidis are a key virulence factor, being the major adhesin of this capsulate organism and contributing to specificity for the human host. Pili are post-translationally modified by addition of an O-linked trisaccharide, Gal(beta1-4)Gal(alpha1-3)2,4-diacetimido-2,4,6-trideoxyhexose++ +. In a previous study the authors identified and characterized a gene, pglA, encoding a galactosyltransferase involved in pilin glycosylation. In this study a set of random genomic sequences from N. meningitidis strain MC58 was used to search for further genes involved in pilin glycosylation. Initially, an open reading frame was identified, and designated pglD (pilin glycosylation gene D), which was homologous to genes involved in polysaccharide biosynthesis. The region adjacent to this gene was cloned and nucleotide sequence analysis revealed two further genes, pglB and pglC, which were also homologous with genes involved in polysaccharide biosynthesis. Insertional mutations were constructed in pglB, pglC and pglD in N. meningitidis C311#3, a strain with well-defined LPS and pilin-linked glycan structures, to determine whether these genes had a role in the biosynthesis of either of these molecules. Analysis of these mutants revealed that there was no alteration in the phenotype of LPS in any of the mutant strains as judged by SDS-PAGE gel migration. In contrast, increased gel migration of the pilin subunit molecules of pglB, pglC and pglD mutants by Western analysis was observed. Pilin from each of the pglB, pglC and pglD mutants did not react with a terminal-galactose-specific stain, confirming that the gel migration differences were due to the alteration or absence of the pilin-linked trisaccharide structure in these mutants. In addition, antisera specific for the C311#3 trisaccharide failed to react with pilin from the pglB, pglC, pglD and galE mutants. Analysis of nucleotide sequence homologies has suggested specific roles for pglB, pglC and pglD in the biosynthesis of the 2,4-diacetimido-2,4,6-trideoxyhexose structure.

Power, P. M., K. L. Seib, et al. (2006). "Pilin glycosylation in Neisseria meningitidis occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in Escherichia coli." Biochem Biophys Res Commun **347**(4): 904-8.

Pili (type IV fimbriae) of Neisseria meningitidis are glycosylated by the addition of O-linked sugars. Recent work has shown that PglF, a protein with homology to O-antigen 'flippases', is required for the biosynthesis of the pilin-linked glycan and suggests pilin glycosylation occurs in a manner analogous to the wzy-dependent addition of O-antigen to the core-LPS. O-Antigen ligases are crucial in this pathway for the transfer of undecraprenol-linked sugars to the LPS-core in Gram-negative bacteria. An O-antigen ligase homologue, pglL, was identified in N. meningitidis. PglL mutants showed no change in LPS phenotypes but did show loss of pilin glycosylation, confirming PglL is essential for pilin O-linked glycosylation in N. meningitidis.

Rangarajan, E. S., S. Bhatia, et al. (2007). "Structural context for protein N-glycosylation in bacteria: The structure of PEB3, an adhesin from Campylobacter jejuni." Protein Sci **16**(5): 990-5.

Campylobacter jejuni is unusual among bacteria in possessing a eukaryotic-like system for N-linked protein glycosylation at Asn residues in sequons of the type Asp/Glu-Xaa-Asn-Xaa-Ser/Thr. However, little is known about the structural context of the glycosylated sequons, limiting the design of novel recombinant glycoproteins. To obtain more information on sequon structure, we have determined the crystal structure of the PEB3 (Cj0289c) dimer. PEB3 has the class II periplasmic-binding protein fold, with each monomer having two domains with a ligand-binding site containing citrate located between them, and overall resembles molybdate- and sulfate-binding proteins. The sequon around Asn90 is located within a surface-exposed loop joining two structural elements. The three key residues are well exposed on the surface; hence, they may be accessible to the PglB oligosaccharyltransferase in the folded state.

Reinhold, B. B., C. R. Hauer, et al. (1995). "Detailed structural analysis of a novel, specific O-linked glycan from the prokaryote Flavobacterium meningosepticum." J Biol Chem **270**(22): 13197-203.

In the preceding paper, preliminary analysis revealed a new type of O-linked oligosaccharide of 1244 Da at each of two proposed glycosylation sites on several proteins secreted by the Gram-negative bacterium Flavobacterium meningosepticum (Plummer, T. H., Jr., Tarentino, A. L., and Hauer, C. R. (1995) J. Biol. Chem. 270, 13192-13196). In this report we detail the linkage, sequence, and branching of this unusual heptasaccharide by electrospray (ES) ionization mass spectrometry (MS), and collision-induced dissociation (CID). The proposed structure was supported by a combination of isotopic labeling, composition and methylation analysis, and the preparation of several chemical analogs and derivatives with each product evaluated by MS and CID. The singly branched structure contained seven residues, including three different uronyl analogs: a methylated rhamnose and mannose, a glucose, and a reducing terminal mannose. Only pyranose ring forms were detected ((2-OMe)Man1-4GlcNAcU1-4GlcU1-4Glc1-4(2-OMe)G lcU-4 [(2-OMe)Rham1-2]Man).

Ristl, R., K. Steiner, et al. "The s-layer glycome-adding to the sugar coat of bacteria." Int

J Microbiol **2011**.

The amazing repertoire of glycoconjugates present on bacterial cell surfaces includes lipopolysaccharides, capsular polysaccharides, lipooligosaccharides, exopolysaccharides, and glycoproteins. While the former are constituents of Gram-negative cells, we review here the cell surface S-layer glycoproteins of Gram-positive bacteria. S-layer glycoproteins have the unique feature of self-assembling into 2D lattices providing a display matrix for glycans with periodicity at the nanometer scale. Typically, bacterial S-layer glycans are O-glycosidically linked to serine, threonine, or tyrosine residues, and they rely on a much wider variety of constituents, glycosidic linkage types, and structures than their eukaryotic counterparts. As the S-layer glycome of several bacteria is unravelling, a picture of how S-layer glycoproteins are biosynthesized is evolving. X-ray crystallography experiments allowed first insights into the catalysis mechanism of selected enzymes. In the future, it will be exciting to fully exploit the S-layer glycome for glycoengineering purposes and to link it to the bacterial interactome.

Romain, F., C. Horn, et al. (1999). "Deglycosylation of the 45/47-kilodalton antigen complex of Mycobacterium tuberculosis decreases its capacity to elicit in vivo or in vitro cellular immune responses." Infect Immun **67**(11): 5567-72.

A protection against a challenge with Mycobacterium tuberculosis is induced by previous immunization with living attenuated mycobacteria, usually bacillus Calmette-Guerin (BCG). The 45/47-kDa antigen complex (Apa) present in culture filtrates of BCG of M. tuberculosis has been identified and isolated based on its ability to interact mainly with T lymphocytes and/or antibodies induced by immunization with living bacteria. The protein is glycosylated. A large batch of Apa was purified from M. tuberculosis culture filtrate to determine the extent of glycosylation and its role on the expression of the immune responses. Mass spectrometry revealed a spectrum of glycosylated molecules, with the majority of species bearing six, seven, or eight mannose residues (22, 24, and 17%, respectively), while others three, four, or five mannoses (5, 9, and 14%, respectively). Molecules with one, two, or nine mannoses were rare (1.5, 3, and 3%, respectively), as were unglycosylated species (in the range of 1%). To eliminate the mannose residues linked to the protein, the glycosylated Apa molecules were chemically or enzymatically treated. The deglycosylated antigen was 10-fold less active than native molecules in eliciting delayed-type hypersensitivity reactions in guinea pigs immunized with BCG. It was 30-fold less active than native molecules when assayed in vitro for its capacity to stimulate T lymphocytes primed in vivo. The presence of the mannose residues on the Apa protein was essential for the antigenicity of the molecules in T-cell-dependent immune responses in vitro and in vivo.

Santos-Silva, T., J. M. Dias, et al. (2007). "Crystal structure of the 16 heme cytochrome from Desulfovibrio gigas: a glycosylated protein in a sulphate-reducing bacterium." J Mol Biol **370**(4): 659-73.

Sulphate-reducing bacteria have a wide variety of periplasmic cytochromes

involved in electron transfer from the periplasm to the cytoplasm. HmcA is a high molecular mass cytochrome of 550 amino acid residues that harbours 16 c-type heme groups. We report the crystal structure of HmcA isolated from the periplasm of Desulfovibrio gigas. Crystals were grown using polyethylene glycol 8K and zinc acetate, and diffracted beyond 2.1 A resolution. A multiple-wavelength anomalous dispersion experiment at the iron absorption edge enabled us to obtain good-quality phases for structure solution and model building. DgHmcA has a V-shape architecture, already observed in HmcA isolated from Desulfovibrio vulgaris Hildenborough. The presence of an oligosaccharide molecule covalently bound to an Asn residue was observed in the electron density maps of DgHmcA and confirmed by mass spectrometry. Three modified monosaccharides appear at the highly hydrophobic vertex, possibly acting as an anchor of the protein to the cytoplasmic membrane.

Sartain, M. J. and J. T. Belisle (2009). "N-Terminal clustering of the O-glycosylation sites in the Mycobacterium tuberculosis lipoprotein SodC." Glycobiology **19**(1): 38-51.
SodC is one of two superoxide dismutases produced by Mycobacterium tuberculosis. This protein was previously shown to contribute to virulence and to act as a B-cell antigen. SodC is also a putative lipoprotein, and like other Sec-translocated mycobacterial proteins it was suggested to be modified with glycosyl units. To definitively define the glycosylation of SodC, we applied an approach that combined site-directed mutagenesis, lectin binding, and mass spectrometry. This resulted in identification of six O-glycosylated residues within a 13-amino-acid region near the N-terminus. Each residue was modified with one to three hexose units, and the most dominant SodC glycoform was modified with nine hexose units. In addition to O-glycosylation of threonine residues, this study provides the first evidence of serine O-glycosylation in mycobacteria. When combined with bioinformatic analyses, the clustering of O-glycosylation appeared to occur in a region of SodC with a disordered structure and not in regions important to the enzymatic activity of SodC. The use of recombinant amino acid substitutions to alter glycosylation sites provided further evidence that glycosylation influences proteolytic processing and ultimately positioning of cell wall proteins.

Sasisekharan, R., M. Bulmer, et al. (1993). "Cloning and expression of heparinase I gene from Flavobacterium heparinum." Proc Natl Acad Sci U S A **90**(8): 3660-4.
Heparinases, enzymes that cleave heparin and heparin sulfate, are implicated in physiological and pathological functions ranging from wound healing to tumor metastasis and are useful in deheparinization therapies. We report the cloning of the heparinase I (EC 4.2.2.7) gene from Flavobacterium heparinum using PCR. Two degenerate oligonucleotides, based on the amino acid sequences derived from tryptic peptides of purified heparinase, were used to generate a 600-bp probe by PCR amplification using Flavobacterium genomic DNA as the template. This probe was used to screen a Flavobacterium genomic DNA library in pUC18. The open reading frame of heparinase I is 1152 bp in length, encoding a precursor protein of 43.8 kDa. Eleven of the tryptic peptides (approximately 35%

of the total amino acids) mapped onto the open reading frame. The amino acid sequence reveals a consensus heparin binding domain and a 21-residue leader peptide with a characteristic Ala-(Xaa)-Ala cleavage site. Recombinant heparinase was expressed in Escherichia coli as a soluble protein, using the T7 polymerase pET expression system. The recombinant heparinase cleavage of heparin was identical to that of native heparinase.

Schaffer, C. and P. Messner (2004). "Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology." Glycobiology **14**(8): 31R-42R.

Bacterial cell surface layers, referred to simply as S-layers, have been described for all major phylogenetic groups of bacteria, which may indicate their pivotal role for a bacterium in its natural habitat. They have the unique ability to assemble into two-dimensional crystalline arrays that completely cover the bacterial cells. Glycosylation represents the most frequent modification of S-layer proteins. S-layer glycoproteins constitute a class of glycoconjugates first isolated in the mid-1970s, but S-layer glycoprotein research is still being regarded as an "exotic field of glycobiology," possibly because of its "noneukaryotic" character. Extensive work over the past 30 years provided evidence of an enormous diversity of S-layer glycoproteins that have been created in nature over 3 billion years of prokaryotic evolution. These glycoconjugates are substantially different from eukaryotic glycoproteins, with regard to both composition and structure; nevertheless, some general structural concepts may be deduced. The awareness of the high application potential of S-layer glycoproteins, especially in combination with their intrinsic cell surface display feature, in the field of modern nanobiotechnology as a base for glycoengineering has recently led to the investigation of the S-layer protein glycosylation process at the molecular level, which has lagged behind the structural studies due to the lack of suitable molecular tools. From that work an even more interesting picture of this class of glycoconjugates is emerging. The availability of purified enzymes from S-layer glycan biosynthesis pathways exhibiting increased stabilities and/or rare sugar specificities in conjunction with preliminary genomic data on S-layer glycan biosynthesis clusters will pave the way for the rational design of S-layer neoglycoproteins.

Schaffer, C., N. Muller, et al. (1999). "Complete glycan structure of the S-layer glycoprotein of Aneurinibacillus thermoaerophilus GS4-97." Glycobiology **9**(4): 407-14.

Isolate GS4-97 was purified from an extraction juice sample of an Austrian beet sugar factory and affiliated to the newly described species Aneurinibacillus thermoaerophilus. It is closely related to the type strain of this species, A.thermoaerophilus L420-91(T), and possesses a square surface layer (S-layer) array composed of identical glycoprotein monomers as its outermost cell envelope component. By sodium dodecyl sulfate-polyacrylamide gel electrophoresis, the purified S-layer showed an apparent molecular mass of approximately 109,000. After thorough proteolytic degradation of this material by pronase E and purification of the reaction mixture by gel permeation,

chromatofocusing, and reversed-phase chromatography, a homogeneous glycopeptide fraction was obtained which was subjected to one- and two-dimensional nuclear magnetic resonance spectroscopy. The combined chemical and spectroscopic evidence, together with N-terminal sequencing, suggest the following structure of the O-glycosidically linked S-layer glycan chain of the glycopeptide: This is the first description of a beta-d-GalNAc-Thr linkage in glycoproteins.

Schaffer, C., T. Wugeditsch, et al. (2002). "The surface layer (S-layer) glycoprotein of Geobacillus stearothermophilus NRS 2004/3a. Analysis of its glycosylation." J Biol Chem **277**(8): 6230-9.

Geobacillus stearothermophilus NRS 2004/3a possesses an oblique surface layer (S-layer) composed of glycoprotein subunits as the outermost component of its cell wall. In addition to the elucidation of the complete S-layer glycan primary structure and the determination of the glycosylation sites, the structural gene sgsE encoding the S-layer protein was isolated by polymerase chain reaction-based techniques. The open reading frame codes for a protein of 903 amino acids, including a leader sequence of 30 amino acids. The mature S-layer protein has a calculated molecular mass of 93,684 Da and an isoelectric point of 6.1. Glycosylation of SgsE was investigated by means of chemical analyses, 600-MHz nuclear magnetic resonance spectroscopy, and matrix-assisted laser desorption ionization-time of flight mass spectrometry. Glycopeptides obtained after Pronase digestion revealed the glycan structure [-->2)-alpha-L-Rhap-(1-->3)-beta-L-Rhap-(1-->2)-alpha-L-Rhap-(1-->](n = 13-18), with a 2-O-methyl group capping the terminal trisaccharide repeating unit at the non-reducing end of the glycan chains. The glycan chains are bound via the disaccharide core -->3)-alpha-l-Rhap-(1-->3)-alpha-L-Rhap-(L--> and the linkage glycose beta-D-Galp in O-glycosidic linkages to the S-layer protein SgsE at positions threonine 620 and serine 794. This S-layer glycoprotein contains novel linkage regions and is the first one among eubacteria whose glycosylation sites have been characterized.

Scherman, H., D. Kaur, et al. (2009). "Identification of a polyprenylphosphomannosyl synthase involved in the synthesis of mycobacterial mannosides." J Bacteriol **191**(21): 6769-72.

We report on the identification of a glycosyltransferase (GT) from Mycobacterium tuberculosis H37Rv, Rv3779, of the membranous GT-C superfamily responsible for the direct synthesis of polyprenyl-phospho-mannopyranose and thus indirectly for lipoarabinomannan, lipomannan, and the higher-order phosphatidyl-myo-inositol mannosides.

Schirm, M., S. K. Arora, et al. (2004). "Structural and genetic characterization of glycosylation of type a flagellin in Pseudomonas aeruginosa." J Bacteriol **186**(9): 2523-31.

Type a flagellins from two strains of Pseudomonas aeruginosa, strains PAK and JJ692, were found to be glycosylated with unique glycan structures. In both

cases, two sites of O-linked glycosylation were identified on each monomer, and these sites were localized to the central, surface-exposed domain of the monomer in the assembled filament. The PAK flagellin was modified with a heterogeneous glycan comprising up to 11 monosaccharide units that were O linked through a rhamnose residue to the protein backbone. The flagellin of JJ692 was less complex and had a single rhamnose substitution at each site. The role of the glycosylation island gene cluster in the production of each of these glycosyl moieties was investigated. These studies revealed that the orfA and orfN genes were required for attachment of the heterologous glycan and the proximal rhamnose residue, respectively.

Schirm, M., M. Kalmokoff, et al. (2004). "Flagellin from Listeria monocytogenes is glycosylated with beta-O-linked N-acetylglucosamine." J Bacteriol **186**(20): 6721-7.

Glycan staining of purified flagellin from Listeria monocytogenes serotypes 1/2a, 1/2b, 1/2c, and 4b suggested that the flagellin protein from this organism is glycosylated. Mass spectrometry analysis demonstrated that the flagellin protein of L. monocytogenes is posttranslationally modified with O-linked N-acetylglucosamine (GlcNAc) at up to six sites/monomer. The sites of glycosylation are all located in the central, surface-exposed region of the protein monomer. Immunoblotting with a monoclonal antibody specific for beta-O-linked GlcNAc confirmed that the linkage was in the beta configuration, this residue being a posttranslational modification commonly observed in eukaryote nuclear and cytoplasmic proteins.

Schirm, M., E. C. Soo, et al. (2003). "Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori." Mol Microbiol **48**(6): 1579-92.

Mass spectrometry analyses of the complex polar flagella from Helicobacter pylori demonstrated that both FlaA and FlaB proteins are post-translationally modified with pseudaminic acid (Pse5Ac7Ac, 5,7-diacetamido-3,5,7,9-tetradeoxy-l-glycero-l-manno -n o n-ulosonic acid). Unlike Campylobacter, flagellar glycosylation in Helicobacter displays little heterogeneity in isoform or glycoform distribution, although all glycosylation sites are located in the central core region of the protein monomer in a manner similar to that found in Campylobacter. Bioinformatic analysis revealed five genes (HP0840, HP0178, HP0326A, HP0326B, HP0114) homologous to other prokaryote genes previously reported to be involved in motility, flagellar glycosylation or polysaccharide biosynthesis. Insertional mutagenesis of four of these homologues in Helicobacter (HP0178, HP0326A, HP0326B, HP0114) resulted in a non-motile phenotype, no structural flagella filament and only minor amounts of flagellin protein detectable by Western immunoblot. However, mRNA levels for the flagellin structural genes remained unaffected by each mutation. In view of the combined bioinformatic and structural evidence indicating a role for these gene products in glycan biosynthesis, subsequent investigations focused on the functional characterization of the respective gene products. A novel approach was devised to identify biosynthetic sugar nucleotide precursors from intracellular

metabolic pools of parent and isogenic mutants using capillary electrophoresis-electrospray mass spectrometry (CE-ESMS) and precursor ion scanning. HP0326A, HP0326B and the HP0178 gene products are directly involved in the biosynthesis of the nucleotide-activated form of Pse, CMP-Pse. Mass spectral analyses of the cytosolic extract from the HP0326A and HP0326B isogenic mutants revealed the accumulation of a mono- and a diacetamido trideoxyhexose UDP sugar nucleotide precursor.

Schmidt, M. A., L. W. Riley, et al. (2003). "Sweet new world: glycoproteins in bacterial pathogens." Trends Microbiol **11**(12): 554-61.

In eukaryotes, the combinatorial potential of carbohydrates is used for the modulation of protein function. However, despite the wealth of cell wall and surface-associated carbohydrates and glycoconjugates, the accepted dogma has been that prokaryotes are not able to glycosylate proteins. This has now changed and protein glycosylation in prokaryotes is an accepted fact. Intriguingly, in Gram-negative bacteria most glycoproteins are associated with virulence factors of medically significant pathogens. Also, important steps in pathogenesis have been linked to the glycan substitution of surface proteins, indicating that the glycosylation of bacterial proteins might serve specific functions in infection and pathogenesis and interfere with inflammatory immune responses. Therefore, the carbohydrate modifications and glycosylation pathways of bacterial proteins will become new targets for therapeutic and prophylactic measures. Here we discuss recent findings on the structure, genetics and function of glycoproteins of medically important bacteria and potential applications of bacterial glycosylation systems for the generation of novel glycoconjugates.

Schoenhofen, I. C., V. V. Lunin, et al. (2006). "Structural and functional characterization of PseC, an aminotransferase involved in the biosynthesis of pseudaminic acid, an essential flagellar modification in Helicobacter pylori." J Biol Chem **281**(13): 8907-16.

Helicobacter pylori flagellin is heavily glycosylated with the novel sialic acid-like nonulosonate, pseudaminic acid (Pse). The glycosylation process is essential for assembly of functional flagellar filaments and consequent bacterial motility. Because motility is a key virulence factor for this and other important pathogens, the Pse biosynthetic pathway offers potential for novel therapeutic targets. From recent NMR analyses, we determined that the conversion of UDP-alpha-D-Glc-NAc to the central intermediate in the pathway, UDP-4-amino-4,6-dideoxy-beta-L-AltNAc, proceeds by formation of UDP-2-acetamido-2,6-dideoxy-beta-L-arabino-4-hexulose by the dehydratase/epimerase PseB (HP0840) followed with amino transfer by the aminotransferase, PseC (HP0366). The central role of PseC in the H. pylori Pse biosynthetic pathway prompted us to determine crystal structures of the native protein, its complexes with pyridoxal phosphate alone and in combination with the UDP-4-amino-4,6-dideoxy-beta-L-AltNAc product, the latter being converted to the external aldimine form in the active site of the enzyme. In the binding site, the AltNAc sugar ring adopts a 4C1 chair conformation, which is different from the predominant 1C4 form found in solution. The enzyme forms a homodimer where each monomer contributes to the active

site, and these structures have permitted the identification of key residues involved in stabilization, and possibly catalysis, of the beta-L-arabino intermediate during the amino transfer reaction. The essential role of Lys183 in the catalytic event was confirmed by site-directed mutagenesis. This work presents for the first time a nucleotide-sugar aminotransferase co-crystallized with its natural ligand, and, in conjunction with the recent functional characterization of this enzyme, these results will assist in elucidating the aminotransferase reaction mechanism within the Pse biosynthetic pathway.

Schwarz, F., C. Lizak, et al. "Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo." Glycobiology **21**(1): 45-54.

A number of proteobacteria carry the genetic information to perform N-linked glycosylation, but only the protein glycosylation (pgl) pathway of Campylobacter jejuni has been studied to date. Here, we report that the pgl gene cluster of Campylobacter lari encodes for a functional glycosylation machinery that can be reconstituted in Escherichia coli. We determined that the N-glycan produced in this system consisted of a linear hexasaccharide. We found that the oligosaccharyltransferase (OST) of C. lari conserved a predominant specificity for the primary sequence D/E-X(-1)-N-X(+1)-S/T (where X(-1) and X(+1) can be any amino acid but proline). At the same time, we observed that this enzyme exhibited a relaxed specificity toward the acceptor site and modified asparagine residues of a protein at sequences DANSG and NNNST. Moreover, C. lari pgl glycosylated a native E. coli protein. Bacterial N-glycosylation appears as a useful tool to establish a molecular description of how single-subunit OSTs perform selection of glycosyl acceptor sites.

Scott, N. E., D. R. Bogema, et al. (2009). "Mass spectrometric characterization of the surface-associated 42 kDa lipoprotein JlpA as a glycosylated antigen in strains of Campylobacter jejuni." J Proteome Res **8**(10): 4654-64.

Campylobacter jejuni is the most common cause of bacterial gastroenteritis in the developed world. Immunoproteomics highlighted a 42-45 kDa antigen that comigrated on two-dimensional (2-DE) gels with the C. jejuni major outer membrane protein (MOMP). Predictive analysis revealed two candidates for the identity of the antigen, the most likely of which was the surface-associated lipoprotein, JlpA. Recombinant JlpA (rJlpA) reacted with patient sera, confirming that JlpA is antigenic. Polyclonal antibodies raised against rJlpA reacted against 3 JlpA mass variants from multiple C. jejuni. These variants differed by approximately 1.5 kDa, suggesting the presence of the N-linked C. jejuni glycan on two sites. Soybean agglutinin affinity and 2-DE purified 2 JlpA glycoforms (43.5 and 45 kDa). Their identities were confirmed using mass spectrometry following trypsin digest. Glycopeptides within JlpA variants were identified by proteinase-K digestion, graphite micropurification and MS-MS. Sites of glycosylation were confirmed as asparagines 107 and 146, both of which are flanked by the N-linked sequon. Sequence analysis confirmed that the N146 sequon is conserved in all C. jejuni genomes examined to date, while the N107 sequon is absent in the reference strain NCTC 11168. Western blotting

confirmed the presence of only a single JlpA glycoform in both virulent (O) and avirulent (GS) isolates of NCTC 11168. MS analysis showed that JlpA exists as 3 discrete forms, unmodified, glycosylated at N146, and glycosylated at both N(146/107), suggesting glycan addition at N146 is necessary for N107 glycosylation. Glycine extracts and Western blotting revealed that doubly glycosylated JlpA was the predominant form on the C. jejuni JHH1 surface; however, glycosylation is not required for antigenicity. This is the first study to identify N-linked glycosylation of a surface-exposed C. jejuni virulence factor and to show strain variation in glycosylation sites.

Shams-Eldin, H., B. Chaban, et al. (2008). "Identification of the archaeal alg7 gene homolog (encoding N-acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in Saccharomyces cerevisiae." J Bacteriol **190**(6): 2217-20.

The Mv1751 gene product is thought to catalyze the first step in the N-glycosylation pathway in Methanococcus voltae. Here, we show that a conditional lethal mutation in the alg7 gene (N-acetylglucosamine-1-phosphate transferase) in Saccharomyces cerevisiae was successfully complemented with Mv1751, highlighting a rare case of cross-domain complementation.

Shaya, D., A. Tocilj, et al. (2006). "Crystal structure of heparinase II from Pedobacter heparinus and its complex with a disaccharide product." J Biol Chem **281**(22): 15525-35.

Heparinase II depolymerizes heparin and heparan sulfate glycosaminoglycans, yielding unsaturated oligosaccharide products through an elimination degradation mechanism. This enzyme cleaves the oligosaccharide chain on the nonreducing end of either glucuronic or iduronic acid, sharing this characteristic with a chondroitin ABC lyase. We have determined the first structure of a heparin-degrading lyase, that of heparinase II from Pedobacter heparinus (formerly Flavobacterium heparinum), in a ligand-free state at 2.15 A resolution and in complex with a disaccharide product of heparin degradation at 2.30 A resolution. The protein is composed of three domains: an N-terminal alpha-helical domain, a central two-layered beta-sheet domain, and a C-terminal domain forming a two-layered beta-sheet. Heparinase II shows overall structural similarities to the polysaccharide lyase family 8 (PL8) enzymes chondroitin AC lyase and hyaluronate lyase. In contrast to PL8 enzymes, however, heparinase II forms stable dimers, with the two active sites formed independently within each monomer. The structure of the N-terminal domain of heparinase II is also similar to that of alginate lyases from the PL5 family. A Zn2+ ion is bound within the central domain and plays an essential structural role in the stabilization of a loop forming one wall of the substrate-binding site. The disaccharide binds in a long, deep canyon formed at the top of the N-terminal domain and by loops extending from the central domain. Based on structural comparison with the lyases from the PL5 and PL8 families having bound substrates or products, the disaccharide found in heparinase II occupies the "+1" and "+2" subsites. The structure of the enzyme-product complex, combined with data from previously characterized

mutations, allows us to propose a putative chemical mechanism of heparin and heparan-sulfate degradation.

Sherlock, O., U. Dobrindt, et al. (2006). "Glycosylation of the self-recognizing Escherichia coli Ag43 autotransporter protein." J Bacteriol **188**(5): 1798-807.
Glycosylation is a common modulation of protein function in eukaryotes and is biologically important. However, in bacteria protein glycosylation is rare, and relatively few bacterial glycoproteins are known. In Escherichia coli only two glycoproteins have been described to date. Here we introduce a novel member to this exclusive group, namely, antigen 43 (Ag43), a self-recognizing autotransporter protein. By mass spectrometry Ag43 was demonstrated to be glycosylated by addition of heptose residues at several positions in the passenger domain. Glycosylation of Ag43 by the action of the Aah and TibC glycosyltransferases was observed in laboratory strains. Importantly, Ag43 was also found to be glycosylated in a wild-type strain, suggesting that Ag43-glycosylation may be a widespread phenomenon. Glycosylation of Ag43 does not seem to interfere with its self-associating properties. However, the glycosylated form of Ag43 enhances bacterial binding to human cell lines, whereas the nonglycosylated version of Ag43 does not to confer this property.

Sherlock, O., M. A. Schembri, et al. (2004). "Novel roles for the AIDA adhesin from diarrheagenic Escherichia coli: cell aggregation and biofilm formation." J Bacteriol **186**(23): 8058-65.
Diarrhea-causing Escherichia coli strains are responsible for numerous cases of gastrointestinal disease and constitute a serious health problem throughout the world. The ability to recognize and attach to host intestinal surfaces is an essential step in the pathogenesis of such strains. AIDA is a potent bacterial adhesin associated with some diarrheagenic E. coli strains. AIDA mediates bacterial attachment to a broad variety of human and other mammalian cells. It is a surface-displayed autotransporter protein and belongs to the selected group of bacterial glycoproteins; only the glycosylated form binds to mammalian cells. Here, we show that AIDA possesses self-association characteristics and can mediate autoaggregation of E. coli cells. We demonstrate that intercellular AIDA-AIDA interaction is responsible for bacterial autoaggregation. Interestingly, AIDA-expressing cells can interact with antigen 43 (Ag43)-expressing cells, which is indicative of an intercellular AIDA-Ag43 interaction. Additionally, AIDA expression dramatically enhances biofilm formation by E. coli on abiotic surfaces in flow chambers.

Smedley, J. G., 3rd, E. Jewell, et al. (2005). "Influence of pilin glycosylation on Pseudomonas aeruginosa 1244 pilus function." Infect Immun **73**(12): 7922-31.
The opportunistic pathogen Pseudomonas aeruginosa is a leading cause of nosocomial pneumonia. Among its virulence factors, the type IV pili of P. aeruginosa strain 1244 contain a covalently linked, three-sugar glycan of previously unknown significance. The work described in this paper was carried out to determine the influence of the P. aeruginosa 1244 pilin glycan on pilus

function, as well as a possible role in pathogenesis. To accomplish this, a deletion was introduced into the pilO gene of this organism. The isogenic knockout strain produced, 1244G7, was unable to glycosylate pilin but could produce pili normal in appearance and quantity. In addition, this strain had somewhat reduced twitching motility, was sensitive to pilus-specific bacteriophages, and could form a normal biofilm. Analysis of whole cells and isolated pili from wild-type P. aeruginosa strain 1244 by transmission electron microscopy with a glycan-specific immunogold label showed that this saccharide was distributed evenly over the fiber surface. The presence of the pilin glycan reduced the hydrophobicity of purified pili as well as whole cells. With regard to pathogenicity, P. aeruginosa strains producing glycosylated pili were commonly found among clinical isolates and particularly among those strains isolated from sputum. Competition index analysis using a mouse respiratory model comparing strains 1244 and 1244G7 indicated that the presence of the pilin glycan allowed for significantly greater survival in the lung environment. These results collectively suggest that the pilin glycan is a significant virulence factor and may aid in the establishment of infection.

Spagnolo, L., I. Toro, et al. (2004). "Unique features of the sodC-encoded superoxide dismutase from Mycobacterium tuberculosis, a fully functional copper-containing enzyme lacking zinc in the active site." J Biol Chem **279**(32): 33447-55.

The sodC-encoded Mycobacterium tuberculosis superoxide dismutase (SOD) shows high sequence homology to other members of the copper/zinc-containing SOD family. Its three-dimensional structure is reported here, solved by x-ray crystallography at 1.63-A resolution. Metal analyses of the recombinant protein indicate that the native form of the enzyme lacks the zinc ion, which has a very important structural and functional role in all other known enzymes of this class. The absence of zinc within the active site is due to significant rearrangements in the zinc subloop, including deletion or mutation of the metal ligands His115 and His123. Nonetheless, the enzyme has a catalytic rate close to the diffusion limit; and unlike all other copper/zinc-containing SODs devoid of zinc, the geometry of the copper site is pH-independent. The protein shows a novel dimer interface characterized by a long and rigid loop, which confers structural stability to the enzyme. As the survival of bacterial pathogens within their host critically depends on their ability to recruit zinc in highly competitive environments, we propose that the observed structural rearrangements are required to build up a zinc-independent but fully active and stable copper-containing SOD.

Steiner, K., R. Novotny, et al. (2007). "Functional characterization of the initiation enzyme of S-layer glycoprotein glycan biosynthesis in Geobacillus stearothermophilus NRS 2004/3a." J Bacteriol **189**(7): 2590-8.

The glycan chain of the S-layer glycoprotein of Geobacillus stearothermophilus NRS 2004/3a is composed of repeating units [-->2)-alpha-l-Rhap-(1-->3)-beta-l-Rhap-(1-->2)-alpha-l-Rhap-(1-->], with a 2-O-methyl modification of the terminal trisaccharide at the nonreducing end of the glycan chain, a core saccharide composed of two or three alpha-l-rhamnose residues, and a beta-d-galactose

residue as a linker to the S-layer protein. In this study, we report the biochemical characterization of WsaP of the S-layer glycosylation gene cluster as a UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase that primes the S-layer glycoprotein glycan biosynthesis of Geobacillus stearothermophilus NRS 2004/3a. Our results demonstrate that the enzyme transfers in vitro a galactose-1-phosphate from UDP-galactose to endogenous phosphoryl-polyprenol and that the C-terminal half of WsaP carries the galactosyltransferase function, as already observed for the UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WbaP from Salmonella enterica. To confirm the function of the enzyme, we show that WsaP is capable of reconstituting polysaccharide biosynthesis in WbaP-deficient strains of Escherichia coli and Salmonella enterica serovar Typhimurium.

Steiner, K., R. Novotny, et al. (2008). "Molecular basis of S-layer glycoprotein glycan biosynthesis in Geobacillus stearothermophilus." J Biol Chem **283**(30): 21120-33.

The Gram-positive bacterium Geobacillus stearothermophilus NRS 2004/3a possesses a cell wall containing an oblique surface layer (S-layer) composed of glycoprotein subunits. O-Glycans with the structure [-->2)-alpha-L-Rhap-(1-->3)-beta-L-Rhap-(1-->2)-alpha-L-Rhap-(1-->](n) (= 13-18), a2-O-methyl group capping the terminal repeating unit at the nonreducing end and a -->2)-alpha-L-Rhap-[(1-->3)-alpha-L-Rhap](n) (= 1-2)(1-->3)- adaptor are linked via a beta-D-Galp residue to distinct sites of the S-layer protein SgsE. S-layer glycan biosynthesis is encoded by a polycistronic slg (surface layer glycosylation) gene cluster. Four assigned glycosyltransferases named WsaC-WsaF, were investigated by a combined biochemical and NMR approach, starting from synthetic octyl-linked saccharide precursors. We demonstrate that three of the enzymes are rhamnosyltransferases that are responsible for the transfer of L-rhamnose from a dTDP-beta-L-Rha precursor to the nascent S-layer glycan, catalyzing the formation of the alpha1,3- (WsaC and WsaD) and beta1,2-linkages (WsaF) present in the adaptor saccharide and in the repeating units of the mature S-layer glycan, respectively. These enzymes work in concert with a multifunctional methylrhamnosyltransferase (WsaE). The N-terminal portion of WsaE is responsible for the S-adenosylmethionine-dependent methylation reaction of the terminal alpha1,3-linked L-rhamnose residue, and the central and C-terminal portions are involved in the transfer of L-rhamnose from dTDP-beta-L-rhamnose to the adaptor saccharide to form the alpha1,2- and alpha1,3-linkages during S-layer glycan chain elongation, with the methylation and the glycosylation reactions occurring independently. Characterization of these enzymes thus reveals the complete molecular basis for S-layer glycan biosynthesis.

Steiner, K., G. Pohlentz, et al. (2006). "New insights into the glycosylation of the surface layer protein SgsE from Geobacillus stearothermophilus NRS 2004/3a." J Bacteriol **188**(22): 7914-21.

The surface of Geobacillus stearothermophilus NRS 2004/3a cells is covered by an oblique surface layer (S-layer) composed of glycoprotein subunits. To this S-layer glycoprotein, elongated glycan chains are attached that are composed of [--

>2)-alpha-l-Rhap-(1-->3)-beta-l-Rhap-(1-->2)-alpha-L-Rhap-(1-->] repeating units, with a 2-O-methyl modification of the terminal trisaccharide at the nonreducing end of the glycan chain and a core saccharide as linker to the S-layer protein. On sodium dodecyl sulfate-polyacrylamide gels, four bands appear, of which three represent glycosylated S-layer proteins. In the present study, nanoelectrospray ionization time-of-flight mass spectrometry (MS) and infrared matrix-assisted laser desorption/ionization orthogonal time-of-flight mass spectrometry were adapted for analysis of this high-molecular-mass and water-insoluble S-layer glycoprotein to refine insights into its glycosylation pattern. This is a prerequisite for artificial fine-tuning of S-layer glycans for nanobiotechnological applications. Optimized MS techniques allowed (i) determination of the average masses of three glycoprotein species to be 101.66 kDa, 108.68 kDa, and 115.73 kDa, (ii) assignment of nanoheterogeneity to the S-layer glycans, with the most prevalent variation between 12 and 18 trisaccharide repeating units, and the possibility of extension of the already-known -->3)-alpha-l-Rhap-(1-->3)-alpha-l-Rhap-(1--> core by one additional rhamnose residue, and (iii) identification of a third glycosylation site on the S-layer protein, at position threonine-590, in addition to the known sites threonine-620 and serine-794. The current interpretation of the S-layer glycoprotein banding pattern is that in the 101.66-kDa glycoprotein species only one glycosylation site is occupied, in the 108.68-kDa glycoprotein species two glycosylation sites are occupied, and in the 115.73-kDa glycoprotein species three glycosylation sites are occupied, while the 94.46-kDa band represents nonglycosylated S-layer protein.

Stepper, J., S. Shastri, et al. "Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins." FEBS Lett **585**(4): 645-50.

O-Glycosylation is a ubiquitous eukaryotic post-translational modification, whereas early reports of S-linked glycopeptides have never been verified. Prokaryotes also glycosylate proteins, but there are no confirmed examples of sidechain glycosylation in ribosomal antimicrobial polypeptides collectively known as bacteriocins. Here we show that glycocin F, a bacteriocin secreted by Lactobacillus plantarum KW30, is modified by an N-acetylglucosamine beta-O-linked to Ser18, and an N-acetylhexosamine S-linked to C-terminal Cys43. The O-linked N-acetylglucosamine is essential for bacteriostatic activity, and the C-terminus is required for full potency (IC(50) 2 nM). Genomic context analysis identified diverse putative glycopeptide bacteriocins in Firmicutes. One of these, the reputed lantibiotic sublancin, was shown to contain a hexose S-linked to Cys22.

Stetefeld, J., M. Jenny, et al. (2000). "Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer." Nat Struct Biol **7**(9): 772-6.

The crystal structure of a polypeptide chain fragment from the surface layer protein tetrabrachion from Staphylothermus marinus has been determined at 1.8 A resolution. As proposed on the basis of the presence of 11-residue repeats, the polypeptide chain fragment forms a parallel right-handed coiled coil structure. Complementary hydrophobic interactions and complex networks of surface salt

bridges result in an extremely thermostable tetrameric structure with remarkable properties. In marked contrast to left-handed coiled coil tetramers, the right-handed coiled coil reveals large hydrophobic cavities that are filled with water molecules. As a consequence, the packing of the hydrophobic core differs markedly from that of a right-handed parallel coiled coil tetramer that was designed on the basis of left-handed coiled coil structures.

Stimson, E., M. Virji, et al. (1995). "Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose." Mol Microbiol **17**(6): 1201-14.

Neisseria meningitidis pili are filamentous protein structures that are essential adhesins in capsulate bacteria. Pili of adhesion variants of meningococcal strain C311 contain glycosyl residues on pilin (PilE), their major structural subunit. Despite the presence of three potential N-linked glycosylation sites, none appears to be occupied in these pilins. Instead, a novel O-linked trisaccharide substituent, not previously found as a constituent of glycoproteins, is present within a peptide spanning amino acid residues 45 to 73 of the PilE molecule. This structure contains a terminal 1-4-linked digalactose moiety covalently linked to a 2,4-diacetamido-2,4,6-trideoxyhexose sugar which is directly attached to pilin. Pilins derived from galactose epimerase (galE) mutants lack the digalactosyl moiety, but retain the diacetamidotrideoxyhexose substitution. Both parental (#3) pilins and those derived from a hyper-adherent variant (#16) contained identical sugar substitutions in this region of pilin, and galE mutants of #3 were similar to the parental phenotype in their adherence to host cells. These studies have confirmed our previous observations that meningococcal pili are glycosylated and provided the first structural evidence for the presence of covalently linked carbohydrate on pili. In addition, they have revealed a completely novel protein/saccharide linkage.

Sugiyama, S., Y. Matsuo, et al. (1996). "The 1.8-A X-ray structure of the Escherichia coli PotD protein complexed with spermidine and the mechanism of polyamine binding." Protein Sci **5**(10): 1984-90.

The PotD protein from Escherichia coli is one of the components of the polyamine transport system present in the periplasm. This component specifically binds either spermidine or putrescine. The crystal structure of the E. coli PotD protein complexed with spermidine was solved at 1.8 A resolution and revealed the detailed substrate-binding mechanism. The structure provided the detailed conformation of the bound spermidine. Furthermore, a water molecule was clearly identified in the binding site lying between the amino-terminal domain and carboxyl-terminal domain. Through this water molecule, the bound spermidine molecule forms two hydrogen bonds with Thr 35 and Ser 211. Another periplasmic component of polyamine transport, the PotF protein, exhibits 35% sequence identity with the PotD protein, and it binds only putrescine, not spermidine. To understand these different substrate specificities, model building of the PotF protein was performed on the basis of the PotD crystal structure. The hypothetical structure suggests that the side chain of Lys 349 in PotF inhibits spermidine binding because of the repulsive forces between its positive charge

and spermidine. On the other hand, putrescine could be accommodated into the binding site without any steric hindrance because its molecular size is much smaller than that of spermidine, and the positively charged amino group is relatively distant from Lys 349.

Sugiyama, S., D. G. Vassylyev, et al. (1996). "Crystal structure of PotD, the primary receptor of the polyamine transport system in Escherichia coli." J Biol Chem **271**(16): 9519-25.

PotD protein is a periplasmic binding protein and the primary receptor of the polyamine transport system, which regulates the polyamine content in Escherichia coli. The crystal structure of PotD in complex with spermidine has been solved at 2.5-A resolution. The PotD protein consists of two domains with an alternating beta-alpha-beta topology. The polyamine binding site is in a central cleft lying in the interface between the domains. In the cleft, four acidic residues recognize the three positively charged nitrogen atoms of spermidine, while five aromatic side chains anchor the methylene backbone by van der Waals interactions. The overall fold of PotD is similar to that of other periplasmic binding proteins, and in particular to the maltodextrin-binding protein from E. coli, despite the fact that sequence identity is as low as 20%. The comparison of the PotD structure with the two maltodextrin-binding protein structures, determined in the presence and absence of the substrate, suggests that spermidine binding rearranges the relative orientation of the PotD domains to create a more compact structure.

Sumper, M., E. Berg, et al. (1990). "Primary structure and glycosylation of the S-layer protein of Haloferax volcanii." J Bacteriol **172**(12): 7111-8.

The outer surface of the archaebacterium Haloferax volcanii (formerly named Halobacterium volcanii) is covered with a hexagonally packed surface (S) layer. The gene coding for the S-layer protein was cloned and sequenced. The mature polypeptide is composed of 794 amino acids and is preceded by a typical signal sequence of 34 amino acid residues. A highly hydrophobic stretch of 20 amino acids at the C-terminal end probably serves as a transmembrane domain. Clusters of threonine residues are located adjacent to this membrane anchor. The S-layer protein is a glycoprotein containing both N- and O-glycosidic bonds. Glucosyl-(1----2)-galactose disaccharides are linked to threonine residues. The primary structure and the glycosylation pattern of the S-layer glycoproteins from Haloferax volcanii and from Halobacterium halobium were compared and found to exhibit distinct differences, despite the fact that three-dimensional reconstructions from electron micrographs revealed no structural differences at least to the 2.5-nm level attained so far (M. Kessel, I. Wildhaber, S. Cohe, and W. Baumeister, EMBO J. 7:1549-1554, 1988).

Szymanski, C. M., D. H. Burr, et al. (2002). "Campylobacter protein glycosylation affects host cell interactions." Infect Immun **70**(4): 2242-4.

Campylobacter jejuni 81-176 pgl mutants impaired in general protein glycosylation showed reduced ability to adhere to and invade INT407 cells and to

colonize intestinal tracts of mice.

Szymanski, C. M. and B. W. Wren (2005). "Protein glycosylation in bacterial mucosal pathogens." Nat Rev Microbiol **3**(3): 225-37.

In eukaryotes, glycosylated proteins are ubiquitous components of extracellular matrices and cellular surfaces. Their oligosaccharide moieties are implicated in a wide range of cell-cell and cell-matrix recognition events that are required for biological processes ranging from immune recognition to cancer development. Glycosylation was previously considered to be restricted to eukaryotes; however, through advances in analytical methods and genome sequencing, there have been increasing reports of both O-linked and N-linked protein glycosylation pathways in bacteria, particularly amongst mucosal-associated pathogens. Studying glycosylation in relatively less-complicated bacterial systems provides the opportunity to elucidate and exploit glycoprotein biosynthetic pathways. We will review the genetic organization, glycan structures and function of glycosylation systems in mucosal bacterial pathogens, and speculate on how this knowledge may help us to understand glycosylation processes in more complex eukaryotic systems and how it can be used for glycoengineering.

Taguchi, F., R. Shimizu, et al. (2003). "Post-translational modification of flagellin determines the specificity of HR induction." Plant Cell Physiol **44**(3): 342-9.

Flagellin, a constituent of the flagellar filament, is a potent elicitor of hypersensitive cell death in plant cells. Flagellins of Pseudomonas syringae pvs. glycinea and tomato induce hypersensitive cell death in their non-host tobacco plants, whereas those of P. syringae pv. tabaci do not remarkably induce it in its host tobacco plants. However, the deduced amino acid sequences of flagellins from pvs. tabaci and glycinea are identical, indicating that post-translational modification of flagellins plays an important role in determining hypersensitive reaction (HR)-inducibility. To investigate genetically the role of modification of flagellin in HR-induction, biological and phytopathological phenotypes of a flagella-defective Delta fliC mutant and Delta fliC mutants complemented by the introduction of the flagellin gene (fliC) from different pathovars of P. syringae were investigated. The Delta fliC mutant of pv. tabaci lost flagella, motility, the ability to induce HR cell death in non-host tomato cells and virulence toward host tobacco plants, whereas all pv. tabaci complemented by the introduction of the fliC gene of pvs. tabaci, glycinea or tomato recovered all the abilities that the Delta fliC mutant had lost. These results indicate that post-translational modification of flagellins is strongly correlated with the ability to cause HR cell death.

Taguchi, F., K. Takeuchi, et al. (2006). "Identification of glycosylation genes and glycosylated amino acids of flagellin in Pseudomonas syringae pv. tabaci." Cell Microbiol **8**(6): 923-38.

A glycosylation island is a genetic region required for glycosylation. The glycosylation island of flagellin in Pseudomonas syringae pv. tabaci 6605 consists of three orfs: orf1, orf2 and orf3. Orf1 and orf2 encode putative

glycosyltransferases, and their deletion mutants, Deltaorf1 and Deltaorf2, exhibit deficient flagellin glycosylation or produce partially glycosylated flagellin respectively. Digestion of glycosylated flagellin from wild-type bacteria and non-glycosylated flagellin from Deltaorf1 mutant using aspartic N-peptidase and subsequent HPLC analysis revealed candidate glycosylated amino acids. By generation of site-directed Ser/Ala-substituted mutants, all glycosylated amino acid residues were identified at positions 143, 164, 176, 183, 193 and 201. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (MS) analysis revealed that each glycan was about 540 Da. While all glycosylation-defective mutants retained swimming ability, swarming ability was reduced in the Deltaorf1, Deltaorf2 and Ser/Ala-substituted mutants. All glycosylation mutants were also found to be impaired in the ability to adhere to a polystyrene surface and in the ability to cause disease in tobacco. Based on the predicted tertiary structure of flagellin, S176 and S183 are expected to be located on most external surface of the flagellum. Thus the effect of Ala-substitution of these serines is stronger than that of other serines. These results suggest that glycosylation of flagellin in P. syringae pv. tabaci 6605 is required for bacterial virulence. It is also possible that glycosylation of flagellin may mask elicitor function of flagellin molecule.

Taguchi, F., M. Yamamoto, et al. "Defects in flagellin glycosylation affect the virulence of Pseudomonas syringae pv. tabaci 6605." Microbiology **156**(Pt 1): 72-80.

Flagellar motility and its glycosylation are indispensable for the virulence of Pseudomonas syringae pv. tabaci 6605. Six serine residues of the flagellin protein at positions 143, 164, 176, 183, 193 and 201 are glycosylated, and the glycan structure at 201 was determined to consist of a trisaccharide of two L-rhamnosyl residues and a modified 4-amino-4,6-dideoxyglucosyl (viosamine) residue. To investigate the glycan structures attached to the other serine residues and to identify the glycans important for virulence, Ser/Ala-substituted mutants were generated. Six mutant strains that each retained a single glycosylated serine residue were generated by replacing five of the six serine residues with alanine residues. MALDI-TOF mass analysis of flagellin proteins revealed that the major component of each glycan was a trisaccharide basically similar to that at position 201, but with heterogeneity in glycoform distribution. Swarming motility and amounts of acylhomoserine lactones (AHLs) as quorum-sensing signal molecules were significantly reduced, especially in the S143-5S/A, S164-5S/A and S201-5S/A mutants, whereas tolerance to antibiotics was increased in these three mutants. All the mutants showed lower ability to cause disease on host tobacco plants. These results supported our previous finding that glycosylation of the most externally located sites on the surface of the flagellin molecule, such as S176 and S183, is required for virulence in P. syringae pv. tabaci 6605. Furthermore, it is speculated that flagellum-dependent motility might be correlated with quorum sensing and antibiotic resistance.

Takeuchi, K., H. Ono, et al. (2007). "Flagellin glycans from two pathovars of Pseudomonas syringae contain rhamnose in D and L configurations in different ratios

and modified 4-amino-4,6-dideoxyglucose." J Bacteriol **189**(19): 6945-56.

Flagellins from Pseudomonas syringae pv. glycinea race 4 and Pseudomonas syringae pv. tabaci 6605 have been found to be glycosylated. Glycosylation of flagellin is essential for bacterial virulence and is also involved in the determination of host specificity. Flagellin glycans from both pathovars were characterized, and common sites of glycosylation were identified on six serine residues (positions 143, 164, 176, 183, 193, and 201). The structure of the glycan at serine 201 (S201) of flagellin from each pathovar was determined by sugar composition analysis, mass spectrometry, and (1)H and (13)C nuclear magnetic resonance spectroscopy. These analyses showed that the S201 glycans from both pathovars were composed of a common unique trisaccharide consisting of two rhamnosyl (Rha) residues and one modified 4-amino-4,6-dideoxyglucosyl (Qui4N) residue, beta-D-Quip4N(3-hydroxy-1-oxobutyl)2Me-(1-->3)-alpha-L-Rhap-(1-->2)-alpha- L-Rhap. Furthermore, mass analysis suggests that the glycans on each of the six serine residues are composed of similar trisaccharide units. Determination of the enantiomeric ratio of Rha from the flagellin proteins showed that flagellin from P. syringae pv. tabaci 6605 consisted solely of L-Rha, whereas P. syringae pv. glycinea race 4 flagellin contained both L-Rha and D-Rha at a molar ratio of about 4:1. Taking these findings together with those from our previous study, we conclude that these flagellin glycan structures may be important for the virulence and host specificity of P. syringae.

Tarentino, A. L., G. Quinones, et al. (1993). "Multiple endoglycosidase F activities expressed by Flavobacterium meningosepticum endoglycosidases F2 and F3. Molecular cloning, primary sequence, and enzyme expression." J Biol Chem **268**(13): 9702-8.

The genes for Flavobacterium meningosepticum Endo (endoglycosidase) F2 and Endo F3 were cloned, and their nucleotide sequences were determined. The deduced amino acid sequences were verified independently to a large extent by direct peptide microsequencing of 66 and 84% of native Endo F2 and Endo F3, respectively. Structurally, the Endo F2 and Endo F3 genes code for a typically long leader sequence of 45 and 39 amino acids, respectively, and, in both cases, a mature protein of 290 amino acids. Comparative structural analysis demonstrated minimum overall homology (15-30%) between Endo F1, Endo F2, and Endo F3, but revealed distinct clusters of identical residues distributed throughout the entire sequence, which represent motifs for binding and hydrolysis of beta 1,4-di-N-acetylchitobiosyl linkages in complex carbohydrates. The mobility of native Endo F2 and Endo F3 on SDS-polyacrylamide gel electrophoresis, unlike Endo F1, did not correlate with the molecular weights determined from the coding region of the corresponding genes. Mass spectrometry confirmed that Endo F2 and Endo F3 were heterogeneous and contained approximately 4000 and 1200 daltons of mass not accounted for in the gene structure. We presume that Endo F2 and Endo F3 are variably post-translationally modified during secretion by possible linkage to the hydroxyl of serine.

Tarentino, A. L., G. Quinones, et al. (1995). "Molecular cloning and sequence analysis of flavastacin: an O-glycosylated prokaryotic zinc metalloendopeptidase." Arch Biochem Biophys **319**(1): 281-5.

A new zinc metalloendopeptidase that cleaves peptides on the amino-terminal side of aspartic acid was isolated from the cultural filtrate of Flavobacterium meningosepticum. The gene for this new enzyme was cloned into pBluescript, and the complete nucleotide sequence was determined. Over 40% of the deduced amino acid sequence was verified independently by direct protein microsequencing. The most important structural features of this new enzyme include (i) the presence of an unusual O-linked oligosaccharide of unknown function located at a unique consensus site near the C-terminus and (ii) a characteristic extended zinc-binding site and corresponding Met-turn that places this metalloendopeptidase in the astacin family. This is the first example of a prokaryotic enzyme related to the eukaryotic astacin group; it is being designated hereafter as flavastacin.

Thibault, P., S. M. Logan, et al. (2001). "Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin." J Biol Chem **276**(37): 34862-70.

Flagellins from three strains of Campylobacter jejuni and one strain of Campylobacter coli were shown to be extensively modified by glycosyl residues, imparting an approximate 6000-Da shift from the molecular mass of the protein predicted from the DNA sequence. Tryptic peptides from C. jejuni 81-176 flagellin were subjected to capillary liquid chromatography-electrospray mass spectrometry with a high/low orifice stepping to identify peptide segments of aberrant masses together with their corresponding glycosyl appendages. These modified peptides were further characterized by tandem mass spectrometry and preparative high performance liquid chromatography followed by nano-NMR spectroscopy to identify the nature and precise site of glycosylation. These analyses have shown that there are 19 modified Ser/Thr residues in C. jejuni 81-176 flagellin. The predominant modification found on C. jejuni flagellin was O-linked 5,7-diacetamido-3,5,7,9-tetradeoxy-l-glycero-l-manno-nonulosonic acid (pseudaminic acid, Pse5Ac7Ac) with additional heterogeneity conferred by substitution of the acetamido groups with acetamidino and hydroxyproprionyl groups. In C. jejuni 81-176, the gene Cj1316c, encoding a protein of unknown function, was shown to be involved in the biosynthesis and/or the addition of the acetamidino group on Pse5Ac7Ac. Glycosylation is not random, since 19 of the total 107 Ser/Thr residues are modified, and all but one of these are restricted to the central, surface-exposed domain of flagellin when folded in the filament. The mechanism of attachment appears unrelated to a consensus peptide sequence but is rather based on surface accessibility of Ser/Thr residues in the folded protein.

Totten, P. A. and S. Lory (1990). "Characterization of the type a flagellin gene from Pseudomonas aeruginosa PAK." J Bacteriol **172**(12): 7188-99.

Flagella in procaryotes are complex structures requiring the coordinate expression of over 50 genes, including flagellin, the major repeating structural

protein. We have previously shown that a functional RpoN gene product is required for expression of flagellin in Pseudomonas aeruginosa PAK (P. A. Totten and S. Lory, J. Bacteriol. 172:389-396, 1990) and have now cloned, sequenced, and determined the transcriptional start site of the structural gene for this flagellin. The clones containing this gene produced a protein that reacted on Western immunoblots with polyclonal and four different monoclonal antibodies to purified flagella. However, this flagellin protein in Escherichia coli was slightly smaller (41 kDa) than flagellin protein produced in P. aeruginosa PAK (45 kDa), indicating degradation in E. coli or modification in P. aeruginosa. Comparison of the deduced amino acid sequence of this gene with the amino acid sequences of other flagellins revealed a conservation in the N- and C-terminal domains, suggesting conservation of secretion or assembly signals between these organisms. The sequence 5' of the structural gene contained potential RpoN-specific promoters as well as a promoter sequence recognized by RpoF (sigma 28), the alternative sigma factor required for expression of flagellin genes in E. coli (and Bacillus subtilis). Deletion analysis of the promoter region as well as transcriptional start site mapping implicated the RpoF, and not the RpoN, consensus sequences as the functional promoter for the flagellin gene. Models for the involvement of both RpoN and RpoF in the expression of flagellin in P. aeruginosa are presented.

Twine, S. M., C. J. Paul, et al. (2008). "Flagellar glycosylation in Clostridium botulinum." FEBS J **275**(17): 4428-44.

Flagellins from Clostridium botulinum were shown to be post-translationally modified with novel glycan moieties by top-down MS analysis of purified flagellin protein from strains of various toxin serotypes. Detailed analyses of flagellin from two strains of C. botulinum demonstrated that the protein is modified by a novel glycan moiety of mass 417 Da in O-linkage. Bioinformatic analysis of available C. botulinum genomes identified a flagellar glycosylation island containing homologs of genes recently identified in Campylobacter coli that have been shown to be responsible for the biosynthesis of legionaminic acid derivatives. Structural characterization of the carbohydrate moiety was completed utilizing both MS and NMR spectroscopy, and it was shown to be a novel legionaminic acid derivative, 7-acetamido-5-(N-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-alpha-D-galacto-nonulosonic acid, (alphaLeg5GluNMe7Ac). Electron transfer dissociation MS with and without collision-activated dissociation was utilized to map seven sites of O-linked glycosylation, eliminating the need for chemical derivatization of tryptic peptides prior to analysis. Marker ions for novel glycans, as well as a unique C-terminal flagellin peptide marker ion, were identified in a top-down analysis of the intact protein. These ions have the potential for use in for rapid detection and discrimination of C. botulinum cells, indicating botulinum neurotoxin contamination. This is the first report of glycosylation of Gram-positive flagellar proteins by the 'sialic acid-like' nonulosonate sugar, legionaminic acid.

Twine, S. M., C. W. Reid, et al. (2009). "Motility and flagellar glycosylation in Clostridium difficile." J Bacteriol **191**(22): 7050-62.

In this study, intact flagellin proteins were purified from strains of Clostridium difficile and analyzed using quadrupole time of flight and linear ion trap mass spectrometers. Top-down studies showed the flagellin proteins to have a mass greater than that predicted from the corresponding gene sequence. These top-down studies revealed marker ions characteristic of glycan modifications. Additionally, diversity in the observed masses of glycan modifications was seen between strains. Electron transfer dissociation mass spectrometry was used to demonstrate that the glycan was attached to the flagellin protein backbone in O linkage via a HexNAc residue in all strains examined. Bioinformatic analysis of C. difficile genomes revealed diversity with respect to glycan biosynthesis gene content within the flagellar biosynthesis locus, likely reflected by the observed flagellar glycan diversity. In C. difficile strain 630, insertional inactivation of a glycosyltransferase gene (CD0240) present in all sequenced genomes resulted in an inability to produce flagellar filaments at the cell surface and only minor amounts of unmodified flagellin protein.

Upreti, R. K., M. Kumar, et al. (2003). "Bacterial glycoproteins: functions, biosynthesis and applications." Proteomics **3**(4): 363-79.

Although widely distributed in eukaryotic cells glycoproteins appear to be rare in prokaryotic organisms. The prevalence of the misconception that bacteria do not glycosylate their proteins has been a subject matter of discussion for a long time. Glycoconjugates that are linked to proteins or peptides, generated by the ribosomal translational mechanism have been reported only in the last two to three decades in a few prokaryotic organisms. Most studied prokaryotic glycoproteins are the S-layer glycoproteins of Archeabacteria. Apart from these, membrane-associated, surface-associated, secreted glycoproteins and exoenzymes glycoproteins are also well documented in both, Archea and Eubacteria. From the recent literature, it is now clear that prokaryotes are capable of glycosylating proteins. In general, prokaryotes are deprived of the cellular organelles required for glycosylation. In prokaryotes many different glycoprotein structures have been observed that display much more variation than that observed in eukaryotes. Besides following similar mechanisms in the process of glycosylation, prokaryotes have also been shown to use mechanisms that are different from those found in eukaryotes. The knowledge pertaining to the functional aspects of prokaryotic glycoproteins is rather scarce. This review summarizes developments and understanding relating to characteristics, synthesis, and functions of prokaryotic glycoproteins. An extensive summary of glycosylation that has been reported to occur in bacteria has also been tabulated. Various possible applications of these diverse biomolecules in biotechnology, vaccine development, pharmaceutics and diagnostics are also touched upon.

VanDyke, D. J., J. Wu, et al. (2009). "Identification of genes involved in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon Methanococcus maripaludis." Mol Microbiol **72**(3): 633-44.

Recently, the flagellin proteins of Methanococcus maripaludis were found to harbour an N-linked tetrasaccharide composed of N-acetylgalactosamine, di-

acetylated glucuronic acid, an acetylated and acetamidino-modified mannuronic acid linked to threonine, and a novel terminal sugar [(5S)-2-acetamido-2,4-dideoxy-5-O-methyl-alpha-L-erythro-hexos-5-ulo-1,5-p yranose]. To identify genes involved in the assembly and attachment of this glycan, in-frame deletions were constructed in putative glycan assembly genes. Successful deletion of genes encoding three glycosyltransferases and an oligosaccharyltransferase (Stt3p homologue) resulted in flagellins of decreased molecular masses as evidenced by immunoblotting, indicating partial or completely absent glycan structures. Deletion of the oligosaccharyltransferase or the glycosyltransferase responsible for the transfer of the second sugar in the chain resulted in flagellins that were not assembled into flagella filaments, as evidenced by electron microscopy. Deletions of the glycosyltransferases responsible for the addition of the third and terminal sugars in the glycan were confirmed by mass spectrometry analysis of purified flagellins from these mutants. Although flagellated, these mutants had decreased motility as evidenced by semi-swarm plate analysis with the presence of each additional sugar improving movement capabilities.

VanDyke, D. J., J. Wu, et al. (2008). "Identification of a putative acetyltransferase gene, MMP0350, which affects proper assembly of both flagella and pili in the archaeon Methanococcus maripaludis." J Bacteriol **190**(15): 5300-7.

Glycosylation is a posttranslational modification utilized in all three domains of life. Compared to eukaryotic and bacterial systems, knowledge of the archaeal processes involved in glycosylation is limited. Recently, Methanococcus voltae flagellin proteins were found to have an N-linked trisaccharide necessary for proper flagellum assembly. Current analysis by mass spectrometry of Methanococcus maripaludis flagellin proteins also indicated the attachment of an N-glycan containing acetylated sugars. To identify genes involved in sugar biosynthesis in M. maripaludis, a putative acetyltransferase was targeted for in-frame deletion. Deletion of this gene (MMP0350) resulted in a flagellin molecular mass shift to a size comparable to that expected for underglycosylated or completely nonglycoslyated flagellins, as determined by immunoblotting. Assembled flagellar filaments were not observed by electron microscopy. Interestingly, the deletion also resulted in defective pilus anchoring. Mutant cells with a deletion of MMP0350 had very few, if any, pili attached to the cell surface compared to a nonflagellated but piliated strain. However, pili were obtained from culture supernatants of this strain, indicating that the defect was not in pilus assembly but in stable attachment to the cell surface. Complementation of MMP0350 on a plasmid restored pilus attachment, but it was unable to restore flagellation, likely because the mutant ceased to make detectable flagellin. These findings represent the first report of a biosynthetic gene involved in flagellin glycosylation in archaea. Also, it is the first gene to be associated with pili, linking flagellum and pilus structure and assembly through posttranslational modifications.

Veith, A., A. Klingl, et al. (2009). "Acidianus, Sulfolobus and Metallosphaera surface layers: structure, composition and gene expression." Mol Microbiol **73**(1): 58-72.

The cell walls of Sulfolobales species consist of proteinaceous S-layers assembled from two polypeptides, SlaA and SlaB. We isolated the large S-layer protein of Acidianus ambivalens and both S-layer subunits of Sulfolobus solfataricus and Metallosphaera sedula, respectively. The slaAB genes, lying adjacently in the chromosomes, are constitutively transcribed as bicistronic operons in A. ambivalens and S. solfataricus. A smaller slaA transcript appeared in Northern hybridizations of A. ambivalens RNA. PCRs experiments showed that 80-85% of the transcripts stop at an oligo-T terminator downstream of slaA while 15-20% are read through to a second terminator downstream of slaB. The bicistronic operons including promoter and terminator regions are conserved in the Sulfolobales. While no SlaA homologue is found outside the Sulfolobales, SlaB is distantly similar to S-layer proteins of other Crenarchaeota, e.g. the Staphylothermus marinus tetrabrachion. Molecular modelling suggests SlaBs to be composed of 2-3 consecutive beta sandwich domains, a coiled-coil domain of 15-17 nm in length and a C-terminal transmembrane helix. Electron microscopy shows crystalline protein arrays with triangular and hexagonal pores. We propose that the mushroom-shaped 'unit cells' of the Sulfolobales' S-layers consist of three SlaBs anchoring the complex in the membrane and six SlaAs forming the detergent-resistant outer sacculus.

Venugopal, H., P. J. Edwards, et al. "Structural, dynamic, and chemical characterization of a novel S-glycosylated bacteriocin." Biochemistry **50**(14): 2748-55.

Bacteriocins are bacterial peptides with specific activity against competing species. They hold great potential as natural preservatives and for their probiotic effects. We show here nuclear magnetic resonance-based evidence that glycocin F, a 43-amino acid bacteriocin from Lactobacillus plantarum, contains two beta-linked N-acetylglucosamine moieties, attached via side chain linkages to a serine via oxygen, and to a cysteine via sulfur. The latter linkage is novel and has helped to establish a new type of post-translational modification, the S-linked sugar. The peptide conformation consists primarily of two alpha-helices held together by a pair of nested disulfide bonds. The serine-linked sugar is positioned on a short loop sequentially connecting the two helices, while the cysteine-linked sugar presents at the end of a long disordered C-terminal tail. The differing chemical and conformational stabilities of the two N-actetylglucosamine moieties provide clues about the possible mode of action of this bacteriostatic peptide.

Verma, A., M. Schirm, et al. (2006). "Glycosylation of b-Type flagellin of Pseudomonas aeruginosa: structural and genetic basis." J Bacteriol **188**(12): 4395-403.

The flagellin of Pseudomonas aeruginosa can be classified into two major types-a-type or b-type-which can be distinguished on the basis of molecular weight and reactivity with type-specific antisera. Flagellin from the a-type strain PAK was shown to be glycosylated with a heterogeneous O-linked glycan attached to Thr189 and Ser260. Here we show that b-type flagellin from strain PAO1 is also posttranslationally modified with an excess mass of up to 700 Da, which cannot be explained through phosphorylation. Two serine residues at positions 191 and 195 were found to be modified. Each site had a deoxyhexose to which is linked a

unique modification of 209 Da containing a phosphate moiety. In comparison to strain PAK, which has an extensive flagellar glycosylation island of 14 genes in its genome, the equivalent locus in PAO1 comprises of only four genes. PCR analysis and sequence information suggested that there are few or no polymorphisms among the islands of the b-type strains. Mutations were made in each of the genes, PA1088 to PA1091, and the flagellin from these isogenic mutants was examined by mass spectrometry to determine whether they were involved in posttranslational modification of the type-b flagellin. While mutation of PA1088, PA1089, and PA1090 genes altered the composition of the flagellin glycan, only unmodified flagellin was produced by the PA1091 mutant strain. There were no changes in motility or lipopolysaccharide banding in the mutants, implying a role that is limited to glycosylation.

Vik, A., F. E. Aas, et al. (2009). "Broad spectrum O-linked protein glycosylation in the human pathogen Neisseria gonorrhoeae." Proc Natl Acad Sci U S A **106**(11): 4447-52.

Protein glycosylation is an important element of biologic systems because of its significant effects on protein properties and functions. Although prominent within all domains of life, O-linked glycosylation systems modifying serine and threonine residues within bacteria and eukaryotes differ substantially in target protein selectivity. In particular, well-characterized bacterial systems have been invariably dedicated to modification of individual proteins or related subsets thereof. Here we characterize a general O-linked glycosylation system that targets structurally and functionally diverse groups of membrane-associated proteins in the gram-negative bacterium Neisseria gonorrhoeae, the etiologic agent of the human disease gonorrhea. The 11 glycoproteins identified here are implicated in activities as varied as protein folding, disulfide bond formation, and solute uptake, as well as both aerobic and anaerobic respiration. Along with their common trafficking within the periplasmic compartment, the protein substrates share quasi-related domains bearing signatures of low complexity that were demonstrated to encompass sites of glycan occupancy. Thus, as in eukaryotes, the broad scope of this system is dictated by the relaxed specificity of the glycan transferase as well as the bulk properties and context of the protein-targeting signal rather than by a strict amino acid consensus sequence. Together, these findings reveal previously unrecognized commonalities linking O-linked protein glycosylation in distantly related life forms.

Vinogradov, E., M. B. Perry, et al. (2003). "The structure of the glycopeptides from the fish pathogen Flavobacterium columnare." Carbohydr Res **338**(23): 2653-8.

Proteolytic digestion of the phenol-water extraction product of the fish pathogen Flavobacterium columnare afforded a mixture of glycopeptides in which the oligosaccharide moiety was an unusual hexasaccharide composed of 4-O-methyl-2-acetamido-2-deoxy-D-glucuronic acid (GlcNAcA), D-glucuronic acid (D-GlcA), 2,3-di-O-acetyl-D-xylose (D-Xyl), 2-O-methyl-D-glucuronic acid (D-GlcA), D-mannose (D-Man), and 2-O-methyl-L-rhamnose (L-Rha). By the application of high-resolution 1D and 2D NMR, mass spectrometry, and chemical analysis, the hexasaccharide structure was determined to be: [carbohydrate structure--see

text] where all monosaccharides have the D-configuration except for 2-O-methyl-L-rhamnose; and were in the pyranose form. Only one carbohydrate structure was found. The peptide part was represented by tri- to hepta-peptides with a minimal common tripeptide fragment Asp-Ser-Ala, extended with Ala and Val.

Virji, M., J. R. Saunders, et al. (1993). "Pilus-facilitated adherence of Neisseria meningitidis to human epithelial and endothelial cells: modulation of adherence phenotype occurs concurrently with changes in primary amino acid sequence and the glycosylation status of pilin." Mol Microbiol **10**(5): 1013-28.

Adherence of capsulate Neisseria meningitidis to endothelial and epithelial cells is facilitated in variants that express pili. Whereas piliated variants of N. meningitidis strain C311 adhered to endothelial cells in large numbers (> 150 bacteria/cell), derivatives containing specific mutations that disrupt pilE encoding the pilin subunit were both non-piliated and failed to adhere to endothelial cells (< 1 bacterium/cell). In addition, meningococcal pili recognized human endothelial and epithelial cells but not cells originating from other animals. Variants of strain C311 were obtained that expressed pilins of reduced apparent M(r) and exhibited a marked increase in adherence to epithelial cells. Structural analysis of pilins from two hyper-adherent variants and the parent strain were carried out by DNA sequencing of their pilE genes. Deduced molecular weights of pilins were considerably lower compared with their apparent M(r) values on SDS-PAGE. Hyper-adherent pilins shared unique changes in sequence including substitution of Asn-113 for Asp-113 and changes from Asn-Asp-Thr-Asp to Thr-Asp-Ala-Lys at residues 127-130 in mature pilin. Asn residues 113 and 127 of 'parental' pilin both form part of the typical eukaryotic N-glycosylation motif Asn-X-Ser/Thr and could potentially be glycosylated post-translationally. The presence of carbohydrate on pilin was demonstrated and when pilins were deglycosylated, their migration on SDS-PAGE increased, supporting the notion that variable glycosylation accounts for discrepancies in apparent and deduced molecular weights. Functionally distinct pilins produced by two fully piliated variants of a second strain (MC58) differed only in that the putative glycosylation motif Asn-60-Asn-61-Thr-62 in an adherent variant was replaced with Asp-60-Asn-61-Ser-62 in a non-adherent variant. Fully adherent backswitchers obtained from the non-adherent variant always regained Asn-60 but retained Ser-62. We propose, therefore, that functional variations in N. meningitidis pili may be modulated in large part by primary amino acid sequence changes that ablate or create N-linked glycosylation sites on the pilin subunit.

Virji, M., E. Stimson, et al. (1996). "Posttranslational modifications of meningococcal pili. Identification of a common trisaccharide substitution on variant pilins of strain C311." Ann N Y Acad Sci **797**: 53-64.

Neisseria meningitidis pili are filamentous protein structures that are essential adhesins in capsulate bacteria. Pili of adhesion variants of meningococcal strain C311 contain glycosyl residues on pilin (PilE), their major structural subunit. Recent studies have shown that a novel O-linked trisaccharide substituent, not previously found as a constituent of glycoproteins, is present within a peptide

spanning amino acid residues 50 to 73 of the PilE molecule. The structure was shown to be Gal beta 1-4 Gal alpha 1-3 diacetamidotrideoxyhexose which is directly attached to pilin. Pilins derived from galactose epimerase (galE) mutants lack the digalactosyl moiety, but retain the diacetamidotrideoxyhexose substitution. These studies confirm our previous observations that meningococcal pili are glycosylated and provide the first structural evidence for the presence of covalently linked carbohydrate on pili. We have identified a completely novel protein/carbohydrate linkage on a multimeric protein that is an essential virulence determinant in N. meningitidis.

Voisin, S., R. S. Houliston, et al. (2005). "Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of Methanococcus voltae." J Biol Chem **280**(17): 16586-93.

The flagellum of Methanococcus voltae is composed of four structural flagellin proteins FlaA, FlaB1, FlaB2, and FlaB3. These proteins possess a total of 15 potential N-linked sequons (NX(S/T)) and show a mass shift on an SDS-polyacrylamide gel indicating significant post-translational modification. We describe here the structural characterization of the flagellin glycan from M. voltae using mass spectrometry to examine the proteolytic digests of the flagellin proteins in combination with NMR analysis of the purified glycan using a sensitive, cryogenically cooled probe. Nano-liquid chromatography-tandem mass spectrometry analysis of the proteolytic digests of the flagellin proteins revealed that they are post-translationally modified with a novel N-linked trisaccharide of mass 779 Da that is composed of three sugar residues with masses of 318, 258, and 203 Da, respectively. In every instance the glycan is attached to the peptide through the asparagine residue of a typical N-linked sequon. The glycan modification has been observed on 14 of the 15 sequon sites present on the four flagellin structural proteins. The novel glycan structure elucidated by NMR analysis was shown to be a trisaccharide composed of beta-ManpNAcA6Thr-(1-4)-beta-Glc-pNAc3NAcA-(1-3)-beta-GlcpNAc linked to Asn. In addition, the same trisaccharide was identified on a tryptic peptide of the S-layer protein from this organism implicating a common N-linked glycosylation pathway.

Voisin, S., J. V. Kus, et al. (2007). "Glycosylation of Pseudomonas aeruginosa strain Pa5196 type IV pilins with mycobacterium-like alpha-1,5-linked d-Araf oligosaccharides." J Bacteriol **189**(1): 151-9.

Pseudomonas aeruginosa is a gram-negative bacterium that uses polar type IV pili for adherence to various materials and for rapid colonization of surfaces via twitching motility. Within the P. aeruginosa species, five distinct alleles encoding variants of the structural subunit PilA varying in amino acid sequence, length, and presence of posttranslational modifications have been identified. In this work, a combination of mass spectrometry and nuclear magnetic resonance spectroscopy was used to identify a novel glycan modification on the pilins of the group IV strain Pa5196. Group IV pilins continued to be modified in a lipopolysaccharide (wbpM) mutant of Pa5196, showing that, unlike group I strains, the pilins of group IV are not modified with the O-antigen unit of the

background strain. Instead, the pilin glycan was determined to be an unusual homo-oligomer of alpha-1,5-linked d-arabinofuranose (d-Araf). This sugar is uncommon in prokaryotes, occurring mainly in the cell wall arabinogalactan and lipoarabinomannan (LAM) polymers of mycobacteria, including Mycobacterium tuberculosis and Mycobacterium leprae. Antibodies raised against M. tuberculosis LAM specifically identified the glycosylated pilins from Pa5196, confirming that the glycan is antigenically, as well as chemically, identical to those of Mycobacterium. P. aeruginosa Pa5196, a rapidly growing strain of low virulence that expresses large amounts of glycosylated type IV pilins on its surface, represents a genetically tractable model system for elucidation of alternate pathways for biosynthesis of d-Araf and its polymerization into mycobacterium-like alpha-1,5-linked oligosaccharides.

Voorhorst, W. G., R. I. Eggen, et al. (1996). "Isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from the hyperthermophilic archaeon Pyrococcus furiosus." J Biol Chem **271**(34): 20426-31.

The hyperthermostable serine protease pyrolysin from the hyperthermophilic archaeon Pyrococcus furiosus was purified from membrane fractions. Two proteolytically active fractions were obtained, designated high (HMW) and low (LMW) molecular weight pyrolysin, that showed immunological cross-reaction and identical NH2-terminal sequences in which the third residue could be glycosylated. The HMW pyrolysin showed a subunit mass of 150 kDa after acid denaturation. Incubation of HMW pyrolysin at 95 degrees C resulted in the formation of LMW pyrolysin, probably as a consequence of COOH-terminal autoproteolysis. The 4194-base pair pls gene encoding pyrolysin was isolated and characterized, and its transcription initiation site was identified. The deduced pyrolysin sequence indicated a prepro-enzyme organization, with a 1249-residue mature protein composed of an NH2-terminal catalytic domain with considerable homology to subtilisin-like serine proteases and a COOH-terminal domain that contained most of the 32 possible N-glycosylation sites. The archaeal pyrolysin showed highest homology with eucaryal tripeptidyl peptidases II on the amino acid level but a different cleavage specificity as shown by its endopeptidase activity toward caseins, casein fragments including alphaS1-casein and synthetic peptides.

Wacker, M., D. Linton, et al. (2002). "N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli." Science **298**(5599): 1790-3.

N-linked protein glycosylation is the most abundant posttranslation modification of secretory proteins in eukaryotes. A wide range of functions are attributed to glycan structures covalently linked to asparagine residues within the asparagine-X-serine/threonine consensus sequence (Asn-Xaa-Ser/Thr). We found an N-linked glycosylation system in the bacterium Campylobacter jejuni and demonstrate that a functional N-linked glycosylation pathway could be transferred into Escherichia coli. Although the bacterial N-glycan differs structurally from its eukaryotic counterparts, the cloning of a universal N-linked glycosylation cassette in E. coli opens up the possibility of engineering

permutations of recombinant glycan structures for research and industrial applications.

Waddling, C. A., T. H. Plummer, Jr., et al. (2000). "Structural basis for the substrate specificity of endo-beta-N-acetylglucosaminidase F(3)." Biochemistry **39**(27): 7878-85.

Endo-beta-N-acetylglucosaminidase F(3) cleaves the beta(1-4) link between the core GlcNAc's of asparagine-linked oligosaccharides, with specificity for biantennary and triantennary complex glycans. The crystal structures of Endo F(3) and the complex with its reaction product, the biantennary octasaccharide, Gal-beta(1-4)-GlcNAc-beta(1-2)-Man-alpha(1-3)[Gal-beta(1-4)-GlcNAc-be ta(1-2)-Man-alpha(1-6)]-Man-beta(1-4)-GlcNAc, have been determined to 1.8 and 2.1 A resolution, respectively. Comparison of the structure of Endo F(3) with that of Endo F(1), which is specific for high-mannose oligosaccharides, reveals highly distinct folds and amino acid compositions at the oligosaccharide recognition sites. Binding of the oligosaccharide to the protein does not affect the protein conformation. The conformation of the oligosaccharide is similar to that seen for other biantennary oligosaccharides, with the exception of two links: the Gal-beta(1-4)-GlcNAc link of the alpha(1-3) branch and the GlcNAc-beta(1-2)-Man link of the alpha(1-6) branch. Especially the latter link is highly distorted and energetically unfavorable. Only the reducing-end GlcNAc and two Man's of the trimannose core are in direct contact with the protein. This is in contrast with biochemical data for Endo F(1) that shows that activity depends on the presence and identity of sugar residues beyond the trimannose core. The substrate specificity of Endo F(3) is based on steric exclusion of incompatible oligosaccharides rather than on protein-carbohydrate interactions that are unique to complexes with biantennary or triantennary complex glycans.

Wieland, F., W. Dompert, et al. (1980). "Halobacterial glycoprotein saccharides contain covalently linked sulphate." FEBS Lett **120**(1): 110-4.

Wieland, F., R. Heitzer, et al. (1983). "Asparaginylglucose: Novel type of carbohydrate linkage." Proc Natl Acad Sci U S A **80**(18): 5470-4.

The Halobacterial cell wall glycoprotein was recently shown to contain two types of sulfated saccharides: a repetitive saccharide and a nonrepetitive saccharide composed of glucuronic acid and glucose. A new type of N-glycosidic linkage is found in this latter type of saccharide: glucose is N-glycosidically linked to the polypeptide chain through the amido nitrogen of an asparagine residue, as shown by chemical analyses, proton magnetic resonance spectroscopy, and mass spectroscopy of an isolated asparaginyl saccharide. The only N-glycosidic linkage known so far is between the amido nitrogen of asparagine and N-acetylglucosamine.

Wieland, F., G. Paul, et al. (1985). "Halobacterial flagellins are sulfated glycoproteins." J Biol Chem **260**(28): 15180-5.

The cell-surface glycoprotein of Halobacteria contains oligosaccharides of the type Glc4----1GlcA4----1GlcA4----1GlcA (where GlcA indicates glucuronic acid)

with a sulfate group attached to each of the GlcA residues. We report here that in addition to this cell-surface glycoprotein, the halobacterial flagellar proteins (recently described by Alam, M., and Oesterhelt, D. (1984) J. Mol. Biol. 176, 459-475) also contain the same type of sulfated oligosaccharides. These flagellins have the following features. All of the individual flagellar proteins contain identical sulfated saccharide moieties linked to the amido nitrogen of Asn through a Glc residue (the novel type of N-glycosidic linkage that has been found in the cell-surface glycoprotein from Halobacteria (Wieland, F., Heitzer, R., and Schaefer, W. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 5470-5474)). The amino acid sequence of one carbohydrate-binding region is Gln-Ala-Ala-Gly-Ala-Asp-Asn-Jle-Asn-Leu-Thr-Lys. This surrounding sequence CHO is consistent with the general formula Asn-X-Thr(Ser), common to all N-linked glycopeptides determined so far. Biosynthesis of flagellar glycoconjugates involved sulfated oligosaccharides linked to dolichol monophosphate. The individual glycoproteins making up the flagella are structurally closely related to one another.

Wugeditsch, T., N. E. Zachara, et al. (1999). "Structural heterogeneity in the core oligosaccharide of the S-layer glycoprotein from Aneurinibacillus thermoaerophilus DSM 10155." Glycobiology **9**(8): 787-95.

The surface layer glycoprotein of Aneurinibacillus thermoaerophilus DSM 10155 has a total carbohydrate content of 15% (by mass), consisting of O-linked oligosaccharide chains. After proteolytic digestion of the S-layer glycoprotein byPronase E and subsequent purification of the digestion products by gel permeation chromatography, chromatofocusing and high-performance liquid chromatography two glycopeptide pools A and B with identical glycans and the repeating unit structure -->4)-alpha-l-Rha p -(1-->3)-beta-d- glycero -d- manno -Hep p -(1--> (Kosma et al., 1995b, Glycobiology, 5, 791-796) were obtained. Combined evidence from modified Edman-degradation in combination with liquid chromatography electrospray mass-spectrometry and nuclear magnetic resonance spectroscopy revealed that both glycopeptides contain equal amounts of the complete core structure alpha-l-Rha p -(1-->3)-alpha-l-Rha p -(1-->3)-beta-d-Gal p NAc-(1-->O)-Thr/Ser and the truncated forms alpha-l-Rha p -(1-->3)-beta-d-Gal p NAc-(1-->O)-Thr/Ser and beta-d-Gal p NAc-(1-->O)-Thr/Ser. All glycopeptides possessed the novel linkage types beta-d-Gal p NAc-(1-->O)-Thr/Ser. The different cores were substituted with varying numbers of disaccharide repeating units. By 300 MHz proton nuclear magnetic resonance spectroscopy the complete carbohydrate core structure of the fluorescently labeled glyco-peptide B was determined after Smith-degradation of its glycan chain. The NMR data confirmed and complemented the results of the mass spectroscopy experiments. Based on the S-layer glycopeptide structure, a pathway for its biosynthesis is suggested.

Yeo, H. J., T. Yokoyama, et al. (2007). "The structure of the Haemophilus influenzae HMW1 pro-piece reveals a structural domain essential for bacterial two-partner secretion." J Biol Chem **282**(42): 31076-84.

In pathogenic Gram-negative bacteria, many virulence factors are secreted via

the two-partner secretion pathway, which consists of an exoprotein called TpsA and a cognate outer membrane translocator called TpsB. The HMW1 and HMW2 adhesins are major virulence factors in nontypeable Haemophilus influenzae and are prototype two-partner secretion pathway exoproteins. A key step in the delivery of HMW1 and HMW2 to the bacterial surface involves targeting to the HMW1B and HMW2B outer membrane translocators by an N-terminal region called the secretion domain. Here we present the crystal structure at 1.92 A of the HMW1 pro-piece (HMW1-PP), a region that contains the HMW1 secretion domain and is cleaved and released during HMW1 secretion. Structural analysis of HMW1-PP revealed a right-handed beta-helix fold containing 12 complete parallel coils and one large extra-helical domain. Comparison of HMW1-PP and the Bordetella pertussis FHA secretion domain (Fha30) reveals limited amino acid homology but shared structural features, suggesting that diverse TpsA proteins have a common structural domain required for targeting to cognate TpsB proteins. Further comparison of HMW1-PP and Fha30 structures may provide insights into the keen specificity of TpsA-TpsB interactions.

Young, N. M., J. R. Brisson, et al. (2002). "Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni." J Biol Chem **277**(45): 42530-9.

Mass spectrometry investigations of partially purified Campylobacter jejuni protein PEB3 showed it to be partially modified with an Asn-linked glycan with a mass of 1406 Da and composed of one hexose, five N-acetylhexosamines and a species of mass 228 Da, consistent with a trideoxydiacetamidohexose. By means of soybean lectin affinity chromatography, a mixture of glycoproteins was obtained from a glycine extract, and two-dimensional gel proteomics analysis led to the identification of at least 22 glycoproteins, predominantly annotated as periplasmic proteins. Glycopeptides were prepared from the glycoprotein mixture by Pronase digestion and gel filtration. The structure of the glycan was determined by using nano-NMR techniques to be GalNAc-alpha1,4-GalNAc-alpha1,4-[Glcbeta1,3-]GalNAc-alpha1,4-GalNAc-alpha1 ,4-GalNAc-alpha1,3-Bac-beta1,N-Asn-Xaa, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglucopyranose. Protein glycosylation was abolished when the pglB gene was mutated, providing further evidence that the enzyme encoded by this gene is responsible for formation of the glycopeptide N-linkage. Comparison of the pgl locus with that of Neisseria meningitidis suggested that most of the homologous genes are probably involved in the biosynthesis of bacillosamine.

Yurist-Doutsch, S., M. Abu-Qarn, et al. (2008). "AglF, aglG and aglI, novel members of a gene island involved in the N-glycosylation of the Haloferax volcanii S-layer glycoprotein." Mol Microbiol **69**(5): 1234-45.

Proteins in all three domains of life can experience N-glycosylation. The steps involved in the archaeal version of this post-translational modification remain largely unknown. Hence, as the next step in ongoing efforts to identify components of the N-glycosylation pathway of the halophilic archaeon Haloferax volcanii, the involvement of three additional gene products in the biosynthesis of

the pentasaccharide decorating the S-layer glycoprotein was demonstrated. The genes encoding AglF, AglI and AglG are found immediately upstream of the gene encoding the archaeal oligosaccharide transferase, AglB. Evidence showing that AglF and AglI are involved in the addition of the hexuronic acid found at position three of the pentasaccharide is provided, while AglG is shown to contribute to the addition of the hexuronic acid found at position two. Given their proximities in the H. volcanii genome, the transcription profiles of aglF, aglI, aglG and aglB were considered. While only aglF and aglI share a common promoter, transcription of the four genes is co-ordinated, as revealed by determining transcript levels in H. volcanii cells raised in different growth conditions. Such changes in N-glycosylation gene transcription levels offer additional support for the adaptive role of this post-translational modification in H. volcanii.

Yurist-Doutsch, S. and J. Eichler (2009). "Manual annotation, transcriptional analysis, and protein expression studies reveal novel genes in the agl cluster responsible for N glycosylation in the halophilic archaeon Haloferax volcanii." J Bacteriol **191**(9): 3068-75.
While Eukarya, Bacteria, and Archaea are all capable of protein N glycosylation, the archaeal version of this posttranslational modification is the least understood. To redress this imbalance, recent studies of the halophilic archaeon Haloferax volcanii have identified a gene cluster encoding the Agl proteins involved in the assembly and attachment of a pentasaccharide to select Asn residues of the surface layer glycoprotein in this species. However, because the automated tools used for rapid annotation of genome sequences, including that of H. volcanii, are not always accurate, a reannotation of the agl cluster was undertaken in order to discover genes not previously recognized. In the present report, reanalysis of the gene cluster that includes aglB, aglE, aglF, aglG, aglI, and aglJ, which are known components of the H. volcanii protein N-glycosylation machinery, was undertaken. Using computer-based tools or visual inspection, together with transcriptional analysis and protein expression approaches, genes encoding AglP, AglQ, and AglR are now described.

Yurist-Doutsch, S., H. Magidovich, et al. "N-glycosylation in Archaea: on the coordinated actions of Haloferax volcanii AglF and AglM." Mol Microbiol **75**(4): 1047-58.
Like Eukarya and Bacteria, Archaea are also capable of performing N-glycosylation. In the halophilic archaeon Haloferax volcanii, N-glycosylation is mediated by the products of the agl gene cluster. In the present report, this gene cluster was expanded to include an additional sequence, aglM, shown to participate in the biosynthesis of hexuronic acids contained within a pentasaccharide decorating the S-layer glycoprotein, a reporter H. volcanii glycoprotein. In response to different growth conditions, changes in the transcription profile of aglM mirrored changes in the transcription profiles of aglF, aglG and aglI, genes encoding confirmed participants in the H. volcanii N-glycosylation pathway, thus offering support to the hypothesis that in H. volcanii, N-glycosylation serves an adaptive role. Following purification, biochemical analysis revealed AglM to function as a UDP-glucose dehydrogenase. In a scoupled reaction with AglF, a previously identified glucose-1-phosphate

uridyltransferase, UDP-glucuronic acid was generated from glucose-1-phosphate and UTP in a NAD(+)-dependent manner. These experiments thus represent the first step towards in vitro reconstitution of the archaeal N-glycosylation process.

Zahringer, U., H. Moll, et al. (2000). "Cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose." Eur J Biochem **267**(13): 4144-9.

Cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius (DSM 639) has been described as a novel highly glycosylated membrane-bound b-type hemoprotein [Hettmann, T., Schmidt, C. L., Anemuller, S., Zahringer, U., Moll, H., Petersen, A. & Schafer, G. (1998) J. Biol. Chem. 273, 12032-12040]. The purified cytochrome b558/566 was characterized by MALDI MS as a 64-kDa (glyco)protein expressing 17% glycosylation. Detailed chemical studies showed that it was exclusively O-mannosylated with monosaccharides and N-glycosylated with at least seven hexasaccharide units having the same unique structure. The hexasaccharide was released by cleavage with peptide:N-glycosidase (PNGase) F and found to consist of two residues each of Man and GlcNAc and one residue each of Glc and 6-deoxy-6-sulfoglucose (6-sulfoquinovose). The last sugar has been known as a component of glycolipids of plants and some prokaryotes, but has not been hitherto found in bacterial glycoproteins. Digestion with trypsin/pronase gave a mixture of glycopeptides with the same Asn-linked hexasaccharide chain, from which an N-glycosylated Tyr-Asn dipeptide was purified by gel chromatography and anion-exchange HPLC. Studies of the degradation products using methylation analysis, ESI MS, MALDI MS, and 1H and 13C NMR spectroscopy, including 1H,13C HMQC and NOESY experiments, established the structure of the unique Asn-linked hexasaccharide chain of cytochrome b558/566.

Zampronio, C. G., G. Blackwell, et al. "Novel Glycosylation Sites Localized in Campylobacter jejuni Flagellin FlaA by Liquid Chromatography Electron Capture Dissociation Tandem Mass Spectrometry." J Proteome Res **10**(3): 1238-45.

Glycosylation of flagellin in Campylobacter jejuni is essential for motility and virulence. It is well-known that flagellin from C. jejuni 81-176 is glycosylated by pseudaminic acid and its acetamidino derivative, and that Campylobactor coli VC167 flagellin is glycosylated by legionaminic acid and its derivatives. Recently, it was shown, by use of a metabolomics approach, that C. jejuni 11168 is glycosylated by dimethyl glyceric acid derivatives of pseudaminic acid, but the sites of glycosylation were not confirmed. Here, we apply an online liquid chromatography electron capture dissociation (ECD) tandem mass spectrometry approach to localize sites of glycosylation in flagellin from C. jejuni 11168. Flagellin A is glycosylated by a dimethyl glyceric acid derivative of pseudaminic acid at Ser181, Ser207 and either Thr464 or Thr 465; and by a dimethyl glyceric acid derivative of acetamidino pseudaminic acid at Ser181 and Ser207. For comparison, on-line liquid chromatography collision-induced dissociation of the tryptic digests was performed, but it was not possible to assign sites of glycosylation by that method.

Zayni, S., K. Steiner, et al. (2007). "The dTDP-4-dehydro-6-deoxyglucose reductase encoding fcd gene is part of the surface layer glycoprotein glycosylation gene cluster of Geobacillus tepidamans GS5-97T." Glycobiology **17**(4): 433-43.

The glycan chain of the S-layer protein of Geobacillus tepidamans GS5-97(T) consists of disaccharide repeating units composed of L-rhamnose and D-fucose, the latter being a rare constituent of prokaryotic glycoconjugates. Although biosynthesis of nucleotide-activated L-rhamnose is well established, D-fucose biosynthesis is less investigated. The conversion of alpha-D-glucose-1-phosphate into thymidine diphosphate (dTDP)-4-dehydro-6-deoxyglucose by the sequential action of RmlA (glucose-1-phosphate thymidylyltransferase) and RmlB (dTDP-glucose-4,6-dehydratase) is shared between the dTDP-D-fucose and the dTDP-L-rhamnose biosynthesis pathway. This key intermediate is processed by the dTDP-4-dehydro-6-deoxyglucose reductase Fcd to form dTDP-alpha-D-fucose. We identified the fcd gene in G. tepidamans GS5-97(T) by chromosome walking and performed functional characterization of the recombinant 308-amino acid enzyme. The in vitro activity of the enzymatic cascade (RmlB and Fcd) was monitored by high-performance liquid chromatography and the reaction product was confirmed by (1)H and (13)C nuclear magnetic resonance spectroscopy. This is the first characterization of the dTDP-alpha-D-fucopyranose biosynthesis pathway in a Gram-positive organism. fcd was identified as 1 of 20 open reading frames contained in a 17471-bp S-layer glycosylation (slg) gene cluster on the chromosome of G. tepidamans GS5-97(T). The sgtA structural gene is located immediately upstream of the slg gene cluster with an intergenic region of 247 nucleotides. By comparison of the SgtA amino acid sequence with the known glycosylation pattern of the S-layer protein SgsE of Geobacillus stearothermophilus NRS 2004/3a, two out of the proposed three glycosylation sites on SgtA could be identified by electrospray ionization quadrupole-time-of-flight mass spectrometry to be at positions Ser-792 and Thr-583.

Zeitler, R., E. Hochmuth, et al. (1998). "Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent N-glycosylation of the cell surface glycoprotein from Halobacterium halobium." Glycobiology **8**(12): 1157-64.

The archaeon Halobacterium halobium expresses a cell surface glycoprotein (CSG) with a repeating pentasaccharide unit N-glycosidically linked via N-acetylgalactosamine to Asn-2 of the polypeptide (GalNAc(1-N)Asn linkage type). This aspar-agine of the linkage unit is located within the N-terminal sequence Ala-Asn-Ala-Ser-, in accordance with the tripeptide consensus sequence Asn-Xaa-Ser/Thr typical for nearly every N-glycosylation site known so far, which are of the GlcNAc(1-N)-Asn linkage type. By a gene replacement method csg mutants were created which replace the serine residue of the consensus sequence by valine, leucine, and asparagine. Unexpectedly, this elimination of the consensus sequence did not prevent N-glycosylation. All respective mutant cell surface glycoproteins were N-glycosylated at Asn-2 with the same N-glycan chain as the wild type CSG. Asn-479 is N-glyco-sylated via a Glc(1-N)Asn linkage type in the wild type CSG. Replacement of Ser-481 in the sequence Asn-

Ser-Ser for valine prevented glycosylation of Asn-479. From these results we postulate the existence of two different N-glycosyltransferases in H.halobium, one of which does not use the typical consensus sequence Asn-Xaa-Ser/Thr necessary for all other N-glycosyltransferases described so far.