Structures and functions of proteins that utilize and modify Wall Teichoic Acid

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Cengiz Koç
aus Hanau / Hessen

Tübingen
2016
Tag der mündlichen Qualifikation: Oktober/November 2016

Dekan: Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter: Prof. Dr. Thilo Stehle
2. Berichterstatter: PD Dr. Frank Essmann
1 Contents
I ABBREVIATIONS.................................................................................................. C
II Summary (english)............................................................................................. E
III Summary (german)........................................................................................... F
IV List of publications .......................................................................................... H
V Personal contribution ....................................................................................... I

2 INTRODUCTION ................................................................................................. 1
2.1 Gram-positive Bacteria ..................................................................................... 1
  2.1.1 Staphylococci ............................................................................................... 2
    2.1.1.1 β-hemolysis and Coagulase-negative Staphylococci ......................... 2
    2.1.1.2 MRSA & MSSA .................................................................................... 4
  2.1.2 Actinobacteria .............................................................................................. 5
  2.1.3 Anaerobic Bacilli .......................................................................................... 5
2.2 The Gram-Positive Cell Envelope and the Cell Wall ......................... 7
  2.2.1 Glycosyltransferases and biofilm .............................................................. 9
  2.2.2 Autolysins and Penicillin binding proteins (PBPs) ............................... 9
2.3 Wall teichoic acid (WTA) & lipoteichoic acid (LTA) .................. 12
  2.3.1 WTA de novo biosynthesis .................................................................... 13
  2.3.2 TarM/TarS glycosylation .......................................................................... 14
  2.3.3 Immune evasion via WTA alanylation .................................................... 15
2.4 Immunogenic relevance of cell surface molecules ....................... 17
2.5 Mobile genetic elements (MGE) ................................................................. 19
2.6 Bacteriophages ............................................................................................... 21
  2.6.1 Lytic cycle and lysogenic cycle ............................................................... 22
  2.6.2 The order Caudovirales .......................................................................... 23
    2.6.2.1 Siphoviridae ....................................................................................... 23
    2.6.2.2 Myoviridae ......................................................................................... 24
  2.6.3 S. aureus Pathogenicity Islands (SaPIs) .................................................. 26
  2.6.4 Siphoviridial architecture ....................................................................... 27
    2.6.4.1 The siphoviridial tail end polymorphism ........................................ 27
    2.6.4.2 Receptor binding proteins ............................................................... 29
    2.6.4.3 Teichoic acids as receptors for bacterial viruses ............................ 30
2.7 Aims of Thesis ............................................................................................... 31

3 RESULTS AND DISCUSSION ........................................................................... 32
3.1 TarM ............................................................................................................... 32
  3.1.1 Isolation and crystallization ................................................................... 32
3.1.2 Diffraction data collection and model refinement ................. 32
3.1.3 Overall structure and domain organization of TarM ............... 32
3.1.4 Structural and mutational analysis of the active site .............. 33
3.1.5 DUF1975 ......................................................................... 34
3.2 φ11 Receptor Binding Protein Gp45 ...................................... 36
   3.2.1 Characterization of Gp45 .................................................. 36
   3.2.2 HHPRED analyses of baseplate components .......................... 36
   3.2.3 Isolation and crystallization .............................................. 37
   3.2.4 Diffraction data collection and phasing ................................. 37
   3.2.5 NCS and phase improvement ............................................ 38
   3.2.6 Gp45 structure (Rbp) ......................................................... 39
      3.2.6.1 Stem domain ................................................................ 39
      3.2.6.2 Shoulder domain ......................................................... 40
      3.2.6.3 Head domain .............................................................. 41
   3.2.7 Negative staining electron microscopy of the φ11 baseplate ....... 42
4  FUTURE RESEARCH .................................................................. 43
   4.1 TarM ................................................................................. 43
   4.2 Gp45 ................................................................................. 43
5  REFERENCES .............................................................................. 46
6  APPENDIX ................................................................................. 60
   6.1 Gp45 crystal data and phasing statistics tables ....................... 60
   6.2 Publications ......................................................................... 65
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>ACP</td>
<td>Acyl-carrier protein</td>
</tr>
<tr>
<td>asu</td>
<td>Asymmetric unit</td>
</tr>
<tr>
<td>CAMP</td>
<td>Cationic anti-microbial peptide</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CoNS</td>
<td>Coagulase-negative <em>Staphylococci</em></td>
</tr>
<tr>
<td>cpm</td>
<td>Capsid morphogenesis (gene)</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>DUF</td>
<td>Domain of unknown function</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>EOP</td>
<td>Efficiency of plaquing</td>
</tr>
<tr>
<td>G+C</td>
<td>Guanine+cytosine content of DNA</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>Gle</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
</tr>
<tr>
<td>Gp</td>
<td>Gene product</td>
</tr>
<tr>
<td>GroP</td>
<td>Glycerolphosphate</td>
</tr>
<tr>
<td>GT</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>hgt</td>
<td>Horizontal gene transfer</td>
</tr>
<tr>
<td>HMW</td>
<td>High-molecular weight</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>LMW</td>
<td>Low-molecular weight</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>LTA</td>
<td>Lipoteichoic acid</td>
</tr>
<tr>
<td>ManNAc</td>
<td>N-acetylmannosamine</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-Morpholino)ethanesulfonic acid</td>
</tr>
<tr>
<td>MGE</td>
<td>Mobile genetic elements</td>
</tr>
<tr>
<td>MIRAS</td>
<td>Multiple Isomorphous Replacement with Anomalous Scattering</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-Morpholino)propanesulfonic acid</td>
</tr>
<tr>
<td>MRSA</td>
<td>Methicillin-resistant <em>S. aureus</em></td>
</tr>
<tr>
<td>MSSA</td>
<td>Methicillin-sensitive <em>S. aureus</em></td>
</tr>
<tr>
<td>MurNAC</td>
<td>N-acetyl muramic acid</td>
</tr>
<tr>
<td>NAG</td>
<td>N-acetyl glucosamine</td>
</tr>
<tr>
<td>NAM</td>
<td>N-acetyl muramic acid</td>
</tr>
<tr>
<td>NCS</td>
<td>Non-crystallographic symmetry</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OGP</td>
<td>n-octyl-β-D-glucopyranoside</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleotide-triphosphate</td>
</tr>
<tr>
<td>PAMP</td>
<td>Pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>PBP</td>
<td>Penicillin-binding protein</td>
</tr>
<tr>
<td>PG</td>
<td>Peptidoglycan</td>
</tr>
<tr>
<td>PIA</td>
<td>Polysaccharide intercellular adhesin</td>
</tr>
<tr>
<td>PNAG</td>
<td>Poly-β-1,6-glucosamine</td>
</tr>
<tr>
<td>PS/A</td>
<td>Polysaccharide/adhesin</td>
</tr>
<tr>
<td>PRR</td>
<td>Pathogen recognition receptors</td>
</tr>
<tr>
<td>RboP</td>
<td>Ribitolphosphate</td>
</tr>
<tr>
<td>RBP</td>
<td>Receptor-binding protein</td>
</tr>
<tr>
<td>rmsd</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>SCC</td>
<td>Staphylococcal cassette chromosome</td>
</tr>
<tr>
<td>SIRAS</td>
<td>Single isomorphous replacement with anomalous scattering</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>SaPI</td>
<td>S. aureus pathogenicity islands</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning electron microscope</td>
</tr>
<tr>
<td>SP</td>
<td>Scaffolding protein</td>
</tr>
<tr>
<td>Tag</td>
<td>Teichoic acid glycerol</td>
</tr>
<tr>
<td>Tar</td>
<td>Teichoic acid ribitol</td>
</tr>
<tr>
<td>UDP</td>
<td>Uridine 5'-diphosphate</td>
</tr>
<tr>
<td>vgt</td>
<td>Vertical gene transfer</td>
</tr>
<tr>
<td>VRE</td>
<td>Vancomycin-resistant <em>Enterococci</em></td>
</tr>
<tr>
<td>VRSA</td>
<td>Vancomycin-resistant <em>S. aureus</em></td>
</tr>
<tr>
<td>WTA</td>
<td>Wall teichoic acid</td>
</tr>
</tbody>
</table>
II Summary (english)

By utilizing x-ray crystallography and electron microscopy (EM), this thesis describes the structure determination of two novel proteinaceous macromolecules together with their characterization by microbiological infection assays. These two objects of study revolve around an abundant cell surface molecule, wall teichoic acid (WTA), of *S. aureus* and its variants. These two structures are for one TarM, a glycosyltransferase involved in the final steps of the in-vivo synthesis pathway of WTA, and secondly Gp45, a receptor binding protein located in the baseplate of siphovirus ϕ11, a molecule that recognizes WTA by an unknown mechanism facilitating phage adsorption. Both recently discovered proteins, TarM and Gp45, were recombinantly overproduced in *E. coli*, isolated and crystallized. Since no homologous structure was available to facilitate the usage of the measured structure factor amplitudes for molecular replacement calculations, an appropriate structure solution protocol was sought. TarM crystals were treated with iodide ions and Gp45 crystals were likewise soaked in Tantalbromide solution in order to introduce anomalous scatterers into the protein crystal lattice. Consecutively, applying a combination of single and multiple isomorphous replacement with anomalous scattering (SIRAS and MIRAS, respectively) could retrieve the respective structure factor phase information for the protein atoms. A final solvent flattening routine enabled the structure solution of TarM. Gp45 could be interpreted in a model building process only after a more complex electron density modification routine, combining histogram matching, solvent flattening and averaging protocols. These unbiased models were refined against high-resolution datasets of 2.2 Ångstroms each. The position of Gp45 in the bacteriophage baseplate was elucidated by using negative staining EM on ϕ11 particles and by superposing this x-ray structure model on the 3D reconstruction image derived therefrom. Semi-quantitative infection assays involving WTA deficient in glycosylation using TarM/TarS knockouts or mutants in strain *S. aureus* RN4220 or by blocking the cell surface specifically with recombinant Gp45 were carried out. Additionally, phages blocked with antibodies against Gp45/Gp54 were used to investigate dependencies governing impact and interaction of TarM and Gp45 with WTA in- and outside the cell.

*S. aureus* is responsible for many fatalities in clinical environments caused by infections. Treatments cannot only become tedious but are also a difficult economic factor. The fast resistance developing nature of this germ, of which many details are
Summary

not well understood, is a major obstacle to overcome lethal infections and therefore requires scientific research. The cell envelope represents a major research field and is in some cases affiliated with the emergence of resistance development. This dissertation describes a structure-based research utilizing in-vitro and in-silico laboratory methods, covering a tiny portion of this wide topic by presenting the structures of TarM and Gp45, and is to be seen as a contribution to the understanding of the complicated processes in and around the cell surface of a potent Gram-positive pathogen. The general goal of these interdisciplinary endeavours is to find sensitive treatments by implementing the vast data coming from cell envelope studies into drug design targeted against Gram-positive bacteria, specifically targeting crucial switches of the metabolism or components of the architecture itself.

III Summary (german)


List of publications

Structural and Enzymatic Analysis of TarM Glycosyltransferase from Staphylococcus aureus Reveals an Oligomeric Protein Specific for the Glycosylation of Wall Teichoic Acid

Received for publication, October 20, 2014, and in revised form, February 10, 2015 Published, JBC Papers in Press, February 19, 2015, DOI 10.1074/jbc.M114.619924

Cengiz Koç, David Gerlach, Sebastian Beck, Andreas Peschel, Guoqing Xia, and Thilo Stehle

Anionic glycolipopolymers known as wall teichoic acids (WTAs) functionalize the peptidoglycan layers of many Gram-positive bacteria. WTAs play central roles in many fundamental aspects of bacterial physiology, and they are important determinants of pathogenesis and antibiotic resistance. A number of enzymes that glycosylate WTA in Staphylococcus aureus have recently been identified. Among these is the glycosyltransferase TarM, a component of the WTA de novo biosynthesis pathway. TarM performs the synthesis of O-acetylglycosylated poly-O-phosphoribitol in the WTA structure. We have solved the crystal structure of TarM at 2.4 Å resolution, and we have also determined a structure of the enzyme in complex with its substrate UDP-GlcNAc at 2.8 Å resolution. The protein assembles into a propeller-like homotrimer in which each blade contains a GT-B-type glycosyltransferase domain with a typical Rossmann fold. The enzymatic reaction retains the stereospecificity of the endogenous center of the transferred GlcNAc moiety on the polyribitol backbone. TarM assembles into a trimer using a novel trimerization domain, here termed the HUB domain. Structure-guided mutagenesis experiments of TarM identify residues critical for enzyme activity, assign a putative role for the HUB in TarM function, and allow us to propose a likely reaction mechanism.

An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus

Received: 08 December 2015 Accepted: 28 April 2016 Published online: 23 May 2016 Scientific Reports 6, Article number: 26455 (2016) doi:10.1038/srep26455

Xuehua Li, Cengiz Koç, Petra Kühner, York-Dieter Stierhof, Bernhard Krismer, Mark C. Enright, José R. Penadès, Christine Wolz, Thilo Stehle, Christian Cambillau, Andreas Peschel, Guoqing Xia

Despite the importance of phages in driving horizontal gene transfer (HGT) among pathogenic bacteria, the underlying molecular mechanisms mediating phage adsorption to S. aureus are still unclear. Phage φ11 is a siphovirus with a high transducing efficiency. Here, we show that the tail protein Gp45 localized within the φ11 baseplate. Phage φ11 was efficiently neutralized by anti-Gp45 serum, and its adsorption to host cells was inhibited by recombinant Gp45 in a dose-dependent manner. Flow cytometry analysis demonstrated that biotin-labelled Gp45 efficiently stained the wild-type S. aureus cell but not the double knockout mutant ΔtarM/S, which lacks both α- and β-O-GlcNAc residues on its wall teichoic acids (WTAs). Additionally, adsorption assays indicate that GlcNAc residues on WTAs and O-acetyl groups at the 6-position of muramic acid residues in peptidoglycan are essential components of the φ11 receptor. The elucidation of Gp45-involved molecular interactions not only broadens our understanding of siphovirus-mediated HGT, but also lays the groundwork for the development of sensitive affinity-based diagnostics and therapeutics for S. aureus infection.

Structure of the host-recognition device of Staphylococcus aureus phage φ11

Received: 09 December 2015 Accepted: 17 May 2016 Published online: 10 June 2016 Scientific Reports 6, Article number: 27581 (2016) doi:10.1038/srep27581

Cengiz Koç, Guoqing Xia, Petra Kühner, Silvia Spinelli, Alain Roussel, Christian Cambillau, Thilo Stehle

Phages play key roles in the pathogenicity and adaptation of the human pathogen Staphylococcus aureus. However, little is known about the molecular recognition events that mediate phage adsorption to the surface of S. aureus. The lysogenic siphophage φ11 infects S. aureus SA113. It was shown previously that φ11 requires α- or β-N-acetylglucosamine (GlcNAc) moieties on cell wall teichoic acid (WTA) for adsorption. Gp45 was identified as the receptor binding protein (RBP) involved in this process and GlcNAc residues on WTA were found to be the key component of the φ11 receptor. Here we report the crystal structure of the RBP of φ11, which assembles into a large, multidomain homotrimer. Each monomer contains a five-bladed propeller domain with a cavity that could accommodate a GlcNAc moiety. An electron microscopy reconstruction of the φ11 host adhesion component, the baseplate, reveals that six RBP trimers are assembled around the baseplate core. The Gp45 and baseplate structures provide insights into the overall organization and molecular recognition process of the phage φ11 tail. This assembly is conserved among most glycan-recognizing Siphoviridae, and the RBP orientation would allow host adhesion and infection without an activation step.
V  Personal contribution

The author of this thesis was responsible for solving the crystal structures, interpreting the data and writing the manuscripts for the publications entitled “Structural and Enzymatic Analysis of TarM Glycosyltransferase from Staphylococcus aureus Reveals an Oligomeric Protein Specific for the Glycosylation of Wall Teichoic Acid” and “Structure of the host-recognition device of Staphylococcus aureus phage $\phi$11”. The author of this thesis performed experiments incorporated in the manuscript entitled “An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus”.

2 INTRODUCTION

The outbreak of incurable diseases in confined sterile places such as clinics and hospitals, and the accompanying lack on possibilities to identify treatments to even faster developing community-acquired infections was and is still one of the catalysts for a drive in bio-technological development. In the last 30 years research on the cellular envelope could be intensified due to the establishment of possibilities given by the modernization of methods applicable in laboratories on a standardized basis. Genetic tools allow for quick characterization and identification, and together with online databases give the means to analyze freshly occurring pathogens on the fly. However, the ultimate aim is rather to find solutions in real-time. Essentially, germs of the staphylococcus genus, which represent the main entity of this thesis, infect patients in care and prevent their recuperation\(^1\). Neither the origination of their resistances nor their reported evolution are fully understood. The death toll of clinically-derived diseases is hence high and rising\(^2,3\). The approach by structural research is an highly advanced analytical instrument for the molecular level and represents a first step for drug design facilitating counter-strategies to pandemics. This chapter provides an overview of the cell envelope of Gram-positive bacteria, of which quite a few apart from some Staphylococcus strains are potent pathogens and therefore medically of relevance. The target of this thesis is nested amid the canopy of this broad topic.

2.1 Gram-positive Bacteria

The phylum Firmicutes combines rod- (bacilli\(^4\)) and sphere-shaped (coccic\(^5\)) bacteria with a cell wall and represents a major branch of Gram-positive bacteria\(^1\). As a major distinction to Gram-negative bacteria, the feature of Gram-positive bacteria’s ability to adsorb and hold Gram-stain\(^6\) stands out, with only Mycoplasma\(^7\) being an exception. The most prominent Gram-positive genera are Clostridium\(^8\), Lactobacillus\(^9\), Bacillus\(^4\), Mycoplasma\(^7\) and Streptomyces\(^10\). This staining feature is possible due to the lack of an outer membrane, which in contrast Gram-negative bacteria are surrounded by, their cell wall being part of the periplasmic space. As a consequence, the latter exhibit stronger physical and antibiotic resistances when compared to the former. One interesting feature of Gram-positive bacteria are their teichoic acids\(^11-13\) in their cell walls, where Gram-negative bacteria have Lipopolysaccharides\(^14\) (LPS) instead. Taxonomic
division is also possible by the content of guanine+cytosine bases in DNA (G+C-ratio)\textsuperscript{15} of the bacterial chromosome. A low ratio, i.e. between 25 to 40 mol\% is found for the most common \textit{Firmicutes}, which share certain characteristica, whereas a high G+C content (more than 60 mol\%) indicates actinomycetes\textsuperscript{16}.

\subsection{Staphylococci}

The genus \textit{Staphylococcus} is a member of the \textit{Bacillales} order. It is non-sporing with a low G+C-ratio and consists of ca. 50 species\textsuperscript{17}. These species belong to the family of \textit{Micrococcaceae} and are catalase\textsuperscript{18} producers. The most interesting species are skin-colonizing \textit{Staphylococcus epidermidis}\textsuperscript{19} and the characteristically yellow and spherical \textit{Staphylococcus aureus} (Figure 1). \textit{Staphylococci} can be found in dust, soil, mucous membranes, on skin, i.e. particularly all places with a certain degree of moisture\textsuperscript{20}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Scanning electron microscope (SEM) micrograph of agglutinated \textit{S. aureus} (source: phil.cdc.gov, permission for reproduction courtesy of CDC/ Janice Haney Carr; Jeff Hageman, MHS)}
\end{figure}

\subsubsection{\textbeta-hemolysis and Coagulase-negative Staphylococci}

Besides certain \textit{Listeria-} and \textit{Streptococcus-species}\textsuperscript{21,22}, the also pathogenic \textit{Staphylococcus aureus} has the ability to clot blood. A test differentiating among these species is performed on so-called blood agar\textsuperscript{23}. Plated \textbeta-hemolysing species are indicated by the yellow and opaque lawn around bacterial colonies generated from lysed hemoglobins of red blood
cells (see Figure 2). In comparison, $\alpha$-hemolysis is present when red blood cells are partially broken down, releasing the greenish biliverdin, an intermediary byproduct of hemoglobin catabolism. Lastly, $\gamma$-hemolysis denotes the absence of any hemolysis reaction. On a side note, the catalase activity exhibited by *Staphylococcus* spp. is used to discriminate between the genera *Staphylococcus* and *Streptococcus*. Other than *S. aureus*, coagulase-negative *Staphylococci* (CoNS) lack blood-clotting exoenzymes such as hemolysins and virulence factors but have other means of pathogenesis. Although generally non-pathogenic, they are related to nosocomial infections in patients with implants. These usually harmless skin commensals may cause hospital-acquired infections including central nervous system shunt infection, prosthetic valve endocarditis, urinary tract infection, or endophthalmitis. The special arrangement of extracellular polysaccharides, also referred to as 'slime' or 'biofilm', is one major reason behind this, and enables adhesion of bacteria to synthetic devices.

Chronically, the presence or absence of coagulase was given as a reference to distinguish between coagulase-negative and -positive *Staphylococci*. The coagulase assay in tandem with the thermonuclease reaction and the novobiocin susceptibility resistance test however turned out to be more useful tools when differentiating up to the species level.
2.1.1.2 MRSA & MSSA

Methicillin-resistant *Staphylococcus aureus* (MRSA) is the direct progeny of penicillin-resistant strains\(^44-47\). Historically, infected patients were treated with penicillin\(^48\). An inflationary usage of penicillin led to decrease of susceptibility through the acquisition of \(\beta\)-lactamase. Methicillin, a more potent \(\beta\)-lactam, was used instead of penicillin since it was not affected by \(\beta\)-lactamase at that time\(^49-52\). Eventually, \(\beta\)-lactam-resistant MRSA emerged. The mechanism of *S. aureus* defying methicillin cell toxicity is not yet fully understood. In direct contrast to that stand the methicillin-susceptible *Staphylococcus aureus* (MSSA)\(^53\), which, from a medical point of view are less harmless than the MRSA, due to their treatability\(^54\). MRSA is the most widespread community-acquired infection. It is one of the major causes of nosocomial pneumonia, surgical site infection and blood stream infection. Due to the difficulty of its treatment, it remains a severe threat to health care systems worldwide. Antibiotics resistance in *S. aureus* is conveyed by a vast number of plasmids most likely obtained by horizontal gene transfer (HGT).


2.1.2 **Actinobacteria**

A major phylum of Gram-positive bacteria belongs to the superfamily of actinobacteria\[^{55-59}\]. More than half of the known species of this ‘high G+C’ (around 60 mol\%) branch are assumed to be inhabitants of human mucous membranes. Some are capable to dissolve even sturdy biomaterial like cellulose\[^{60}\] or chitin\[^{61}\]. Actinomycosis is a severe inflammatory illness caused by the more pathogenic species and primarily affect cattle\[^{62}\] but can also cause infectious epidemics in humans\[^{63}\].

*Mycobacterium* has its name derived from the mycolic acid\[^{64}\] present on its surface, making it highly hydrophobic and facilitating characterization by the so-called “acid-fast”\[^{65}\] staining technique. The slow-growing and non-spore-forming *M. tuberculosis* and *M. leprae* are agents responsible for tuberculosis and leprosy, respectively. *M. tuberculosis* captivates the host's macrophages to cause severe damage to peripheral tissue by repeated invasion of macrophages, which lead to malignancies of the host's immune system. Tubercules, formed by phagocytizing macrophages\[^{66,67}\] in lungs, cause lesions and lead to the characteristic symptom of tuberculosis. Eventually the parasite escapes the host through the aerial tract to infect other hosts.

*Streptomyces*\[^{68}\] are mainly soil bacteria with the highest count (ca. 500) of species among *Firmicutes*. Under favourable conditions, they grow sporophores from the hyphal tip of their spore germ tubes\[^{69,70}\]. This multinuclear mycelium\[^{71}\] makes them phenotypically reminiscent of fungi. Several species are symbionts and convey their hosts resilience against diverse pathogens, which makes *Streptomyces* important producers of essential antibiotics\[^{72-76}\]. Examples of such secondary metabolites\[^{77,78}\] are streptomycin (from *Streptomyces griseus*), chloramphenicol, nystatin (fungi-antagonist), erythromycin, tetracyclines and others.

2.1.3 **Anaerobic Bacilli**

Bacilli usually denote rod-shaped bacteria. They are mostly aerobic or facultatively anaerobic spore-formers and inhabit soil and water. The majority in the family of *Bacillaceae* are non-pathogenic. However, exceptions include *B. anthracis*\[^{79}\], *B. cereus*\[^{80}\] and *B. thuringiensis*\[^{81}\]. Moreover, genes from the latter
and from *B. amyloliquefaciens* are especially applied in the field of genetic modifications of plants because of their ability to target infected cells\textsuperscript{82-85}.

*B. subtilis* is a well-studied non-pathogenic species with a low G+C-ratio. It was first discovered in the 19\textsuperscript{th} century in hay in two forms\textsuperscript{86}. Firstly, heat-resistant endospores, which grow vegetatively, are resilient to harsh environmental conditions. Secondly, a predominant heat-prone form being the actual bacterium. *B. subtilis* is the producer of bacitracin. Its cylindrical shape is allegedly a consequence of the specially arranged “circumferential” PG layer and the actin homolog MreB it harbours\textsuperscript{87-89}.

Clostridium is one exemplary object for sporulating Gram-positive rods, which thrive under anaerobic conditions. In the medical field two species are of special interest: the causative agent of botulism, *C. botulism*, and the causer of tetanus, *C. tetanus*\textsuperscript{90}. Examples for other species of interest are *C. perfringens* and *C. difficile*\textsuperscript{91}, which are wide-spread causers of food poisoning.

Other obligate anaerobes include *Propionibacterium*, species of which cause acne, and *Bifidobacterium* which is also commonly being used to treat people with loss of intestinal flora, for example as a result of radiation- or chemotherapy and carcinoms\textsuperscript{92}. A facultative anaerobic genus, especially of interest in dairy production and generally in food production, is *Lactobacillus* spp\textsuperscript{93-95}. 


2.2 The Gram-Positive Cell Envelope and the Cell Wall

The chemical composition of the cell envelope consists of conserved macromolecules of proteoglycan nature and has glycosylated cell-surface proteins tethered to the bacterial rim. As a major part of the cell envelope, the *S. aureus* cell wall is composed of different classes of glycosylic structures, including glycosylated proteins, biofilm Polysaccharide Intercellular Adhesin (PIA or PS/A), capsular polysaccharides, cell wall peptidoglycan (PG), lipoteichoic acid (LTA) and wall teichoic acid (WTA) (see Figure 3). Moreover, specialized Gram-positive bacteria bear surface layer protein (*B. anthracis*) or mycolic acid (*Mycobacterium*) for example.

The integrity of the cell is maintained by the cell wall, a peptidoglycan scaffold. With its multigenous layer the tightly meshed cell wall is able to convey means of protection from and interaction with the environment. It serves as a protective mesh against cell adversaries. Furthermore it is flexible in a manner, which allows cell division, conveys mechanical protection, balances osmotic pressure and permits interchanging of macromolecules with the environment. Making up 90% of the dry weight of a cell, the cell wall of up to 80 nm is able to counter balance the Turgor pressure of 10-25 atm. Flagelli, which reach beyond the cell wall, are also ubiquitous components for any mobile cell.

The chemical composition of the cell wall mainly consists of N-Acetyl muramic acid (MurNAc) and N-Acetyl glucosamine (GlcNAc), β-(1-4)-linked by glycosidic bonds. The interconnection of these singular glycan strands occurs through stem-peptides and pentaglycine bridges (*Staphylococcus*) via the MurNAc carboxylate moiety to form a vast network. According to a calculated model derived from an x-ray structure of a fragment, they are believed to form pseudo-helical strands. Other studies have shown the cell wall to be compartmentalized in a low density inner wall zone, a region of up to 20 nm depending on the genera, and a high density outer wall zone, which generally represents the PG layer and its attachments.

Lipoteichoic acids (LTA) are attached to the cytoplasmic membrane and almost predominantly consist of poly-glycerolphosphate (GroP) in the bacterial kingdom. Free osmosis through the cell wall is a common process, whereas...
Gram-negative bacteria have specialized pore-proteins, called porins, and lipopolysaccharides (LPS), consisting of lipid A, core and O-Antigen, instead of LTA. Gram-negative organisms are surrounded by a secondary outer membrane, whereas the Gram-positive bacterial PG layer represents the margin of the cell and is merely covered by capsular polysaccharides and proteins. The murein layer of Gram-negative bacteria is thin, circumferential and located in the periplasmaticum. Molecularly, PGs of both types are very similar apart from minor variations. Furthermore PG functions as an exoskeleton and is a determining factor of cellular shape. The study of the cell envelope is essential for the understanding of the maintenance and development of virulence. 88, 103, 106

Especially WTA is reported to be involved in infections in wounds of skin, flesh and blood, e.g. olfactory tissue in animal models 107-110. It plays a defensive role by protecting the cell from human lysozyme degradation. Acetylation of MurNAc by OatA on approximately every second C6 in conjunction with WTA on approximately every ninth MurNAc-C6 are presumably responsible for this 111-113. It is involved in biofilm output control, which is necessary for colonization and it functions as cell receptor. The involvement of WTA in processes related to cell proliferation is crucial, although unclear.

Figure 3 Schematic representation of WTA and LTA in the Gram-positive cell envelope (source: given

Annu. Rev. Microbiol. 67:313–36
2.2.1 Glycosyltransferases and biofilm

The synthesis of each glycosidic bond in an organism is done by its own invariant glycosyltransferase (GT). Overall, GT structures can roughly be divided into two big families: GT-A and GT-B. Connecting characteristics for both include a predominant Rossmann fold, which assumes an evolutionary relationship for both GTs. These two superfamilies subsist of numerous subclasses, which define substrate and stereo specificities and targets, conservations in active sites, *in-vivo* enzyme locations, reaction classes and kinetics. An extensive collection of these data is accessible in the Carbohydrate-Active enZYmes Database (www.cazy.org).\(^{114,115}\)

An example for a GT is to be found in PIA production. PIA is referred to a polycistronic * locus in biofilm-active Staphylococci* consisting of the operon icaABCDR, encoding a GT and other auxiliary proteins. It is related to the exopolysaccharide production of CoNS on artificial prosthetic devices and represents a crucial virulence factor in pathogenesis. This slime mainly consists of poly-\(\beta\)-1,6-glucosamine (PNAG) and has an approximate molecular weight of around 30,000 kDa. The deacetylated form of PNAG, like capsular polysaccharide conjugates\(^{116}\), was also reported to be useful in vaccination\(^{117}\).

After deacetylation and sparse substitution with amino acids causing the generation of positive charging of the linear molecule, it becomes electrostatically attracted to the staphylococcal cell surface due to the paramount negative charge caused by the presence of WTA\(^{118}\). This slime production is activated upon infection by an unknown mechanism. PIA presumably assists in intercellular as well as cell-host adhesion and facilitates cell proliferation\(^{98,119}\).

2.2.2 Autolysins and Penicillin binding proteins (PBPs)

PG hydrolyzing enzymes are referred to as autolysins and are essential for PG regeneration\(^{120,121}\). Carboxypeptidases, endopeptidases, N-
acetylmuramoyl-L-alanine amidases and lytic transglycosylases are all counted to this enzyme class. Whereas carboxy- and endopeptidases cleave peptidic bonds of the stem-peptide, the lytic transglycosylases\textsuperscript{122,123} cleave the MurNac-\(\beta\)-(1,4)-GlcNAc glycosidic bond by releasing a peri-cyclic \(\alpha\)-MurNac-intermediary in distinction to lysozyme (amidase)\textsuperscript{124,125}.

The expression of the major autolysin is under control of the \textit{atl}-operon. The cleavage of the gene product results in two autonomous enzymes, N-acetylmuramoyl-L-alanine amidase and \(\beta\)-N-acetyl-glucosaminidase. The amidase is hydrolyzing the peptidoglycan bond between L-alanine of the stem-peptide and MurNac of the glycan strand, the glucosaminidase is dissolving the glycosidic \(\beta\)(1\(\rightarrow\)4) bond between GlcNAc and MurNac (glycosidase)\textsuperscript{124}. The regulatory importance of autolysin capability to degrade the cell wall becomes a crucial factor during cell wall growth\textsuperscript{126}, separation of daughter cells during cell division\textsuperscript{127} and sporulation\textsuperscript{128,129}. Pore enlargement and release of turnover products are other features of \textit{atl} activity in cell wall regulation\textsuperscript{130}.

Class A- and B-PBPs are high-molecular weight (HMW) membrane-anchored, multimodular transpeptidases, which form the peptidic bond between the stem-peptide and the “cross-bridge”, the connector between two adjacent peptidoglycan strands. Except for this transpeptidase domain, most class A-PBPs also harbour a transglycosylase domain. On the other hand, low-molecular weight (LMW) PBPs are well-characterized endopeptidases and are related to transpeptidases by catalyzing the same reaction, but in the reverse direction. Other LMW PBPs are carboxypeptidases, and together with endopeptidases these belong to class C of PBPs. Class C-PBPs have two typical sequence motifs in common, identical to those of HMW-PBPs.

PBPs\textsuperscript{131} have acquired their denomination due to their sensitivity to \(\beta\)-lactam compounds (ampicillin, carbenicillin, penicillin, etc.), which covalently bind to the serine of the conserved SXXK sequence motif of the penicillin-binding/endopeptidase or penicillin-binding/transpeptidase domain. This domain harbors a D-Ala-D-Ala binding site, an intermediary of all classes of PBP activities. The irreversible blocking of the active site by \(\beta\)-lactams via
stable protein-acyl-complexes blocks consequently cell wall biosynthesis at a crucial step\textsuperscript{132-134}.

Of the few \textit{S. aureus} PBPs, the most prominent is PBP2a. It belongs to PBP-class B1 of the HMW transpeptidases and is encoded by \textit{mecA} in the staphylococcal cassette chromosome \textit{mec} (SCC\textit{mec}) on transposon TN552\textsuperscript{135} under regulatory control of MecI and MecR1. The regulatory system and the functional arrangement of its gene products are almost identical to those of the \textit{lac} operon\textsuperscript{136}. Induced overexpression of PBP2a does not change expression levels of other PBPs in the cell, but since it has a lower sensitivity to $\beta$-lactams, it allows for replacement of the enzymes of the murein layer biosynthesis.

PBP2a shares little homology with any PBPs from \textit{E. coli} or \textit{B. subtilis}, but it is similar to subclass B1 \textit{Enterococcus faecium} PBP5\textsuperscript{137}. However, striking differences are not only found in the tertiary structure of the DD-transpeptidase-domain, but PBP2a is also very unique in its active site and its N-terminal domain\textsuperscript{138}. 
2.3 Wall teichoic acid (WTA) & lipoteichoic acid (LTA)

WTAs of various species and different strains are composed of variable chemical units\(^{58,100,139,140}\). Some of these structures are related in different genera of Gram-positive bacteria, which is indicative of a common ancestry, e.g. the linkage unit\(^{141,142}\) \(\text{GroP}_2-3-\text{ManNAc-}\beta(1,4)-\text{GlcNAc-1-P}\) to peptidoglycan-MurNAc-C6 and up to 60 repetitive units covalently connected by phosphodiesters. Predominant repeating units of the WTA main chain are either 1,3-GroP or 1,5-D-ribitolphosphate (RboP) (Fig 4). Different \(B.\ subtilis\) strains have GroP, RboP or even 6-Gal-\(\alpha(1,2)\)-GroP. The genus \(Staphylococcus\) can have either GroP or RboP as WTA\(^{139}\).

LTA molecules are covalently attached to the plasma membrane. The predominant LTA type I of low-G+C Gram-positive bacteria consists of GroP repeating units linked to the cell outer membrane by the glycolipid anchor \(\text{Glc}(\beta1,6)\text{Glc}(\beta1,3)-(\text{gentiobiosyl})\text{diacyl-Gro via C6 of the non-reducing sugar. Other LTA classes are defined as follows: type II LTA has GalGal-GroP and type III has Gal-GroP repetitive units. Rare LTAs, those which do not fall under the categories described above, can be further classified according to chemical nature and degree of GroP-substitutions, glycolipid compositions and chain lengths}\(^{99,143-149}\).
2.3.1 WTA de novo biosynthesis

Biosynthesis of RboP-type WTA in *S. aureus* is carried out intracellularly by a cluster of enzymes belonging to the teichoic acid ribitol (Tar) synthesizing pathway, whereas the related cluster of enzymes for GroP-synthesis are denoted teichoic acid glycerol (Tag), i.e. LTA and WTA are generally synthesized by two independent pathways. The synthesis of the linkage unit disaccharide and attachment onto the lipid carrier undecaprenyl monophosphate, embedded in the inner leaflet of the cell membrane, is
carried out by TarA, TarB (GroP-primase), TarD and TarO, which represent the general steps of WTA-biosynthesis. After this step, WTA-synthesis becomes strain- or species-dependent, culminating in the vast variety of main chain compositions and decorations.

In concert with TarI, a cytidylyltransferase, and TarJ, an alcohol dehydrogenase, which generate the CDP-RboP supply, TarL (both primase and polymerase in S. aureus) adds additional GroP to the linkage unit and elongates the RboP backbone utilizing CDP-RboP. On a side note, some B. subtilis strains with a RboP-genotype have a separate primase (TarK), which putatively adds one primer RboP to the linkage unit, and a polymerase (TarL) for chain elongation (Fig. 5). After completion of biosynthesis, RboP is flipped to the outer leaflet of the plasma membrane via the ABC type transporter TarG/TarH and is covalently attached to the MurNAc moiety of the peptidoglycan.\textsuperscript{150-155}

2.3.2 TarM/TarS glycosylation

RboP-carrying strains may be $\alpha$-, $\beta$-glycoslated at the RboP-C4'-position or D-alaninylated at the C2'-position. Other strains may have glycosylated GroP. For example, in E. faecalis, glycosylated GroP are reported to carry an additional D-alanine on their sugars.\textsuperscript{156} Repeating units comprising GroP-glycosyl-P or GlcNAcP are also known (Staphylococcus lactis).\textsuperscript{157} ArabinolP in Agromyces cerinus\textsuperscript{158} or ErythritolP in Glycomyces tenuis\textsuperscript{159} are other known unusual repeating units.

A SaRN4220-mutant with a transposon in a gene of unknown function was discovered to have WTA with an altered phenotype. Concomitantly, it was found to be resistant to phage infection. A genetic analysis to find the cause for this effect resulted in identifying an open reading frame (ORF) transcribing a GT that was later termed $\text{tarM}$.\textsuperscript{160} On the opposite side, TarS is a GT-A type according to sequence-derived predictions and is specifically decorating WTA with $\beta$-GlcNAc\textsuperscript{161} (Fig. 5).
By now only a handful of teichoic acid biogenesis affiliated protein structures have been solved for either poly-(GroP)-type or complex-type WTA, e.g. TagF from *B. subtilis*\(^{162}\), *Streptococcus* TarI\(^{163}\) (pdb 2vsh), MnaA from *B. anthracis* (pdb 3beo)\(^{164}\), putative WTA ligase from *Streptococcus*\(^{165}\) or lipoteichoic acid polymerase LtaS\(^{166}\).

### 2.3.3 Immune evasion via WTA alanylation

After the biosynthesis of WTA and LTA has been completed and both have been brought to their respective locations in the cell envelope, the sparse D-alanylation at the 2’-position of RboP commences. The gene product DltABCD is the responsible enzyme complex for the esterification. At first, DltA esterifies D-alanine with AMP and transfers the activated substrate to DltC. Sequentially, the pantothenate cofactor of DltC forms a thioester with D-alanine, concurrently releasing AMP. From here on, the membrane-anchored DltB and membrane-spanning DltD transfer the substrate from DltC to WTA or LTA by an unknown mechanism. According to studies, DltD is thought to hydrolyze the DltC-D-alanine-bond and facilitate transfer of D-alanine. The
predicted thioesterase domain of DltD is in accordance with the research performed on hydrolysis of D-alanine on acyl-carrier-protein (ACP). Even less is known about DltB, hence it was suggested that DltB must somehow be involved in a translocation process either as D-alanine-DltC or alone\textsuperscript{118,167}.

In order to counter the host’s native immune system, Gram-positive bacteria have a sensory system, the \textit{aps-xsr} operon, which upregulates the genes necessary for D-alanylation, lysylination (\textit{mprF}) of phosphatidylglycerol of the cytoplasmic membrane and a transporter system. With the aid of this machinery the bacteria are capable of evading CAMP or lysozyme action by electrostatic repulsion through the increased positive charge on their surfaces. Alternatively, the bacterial response specifically to linear CAMPs are proteases, e.g. V8 or aureolysin in \textit{S. aureus}.
2.4 Immunogenic relevance of cell surface molecules

The recognition of the pathogen by the invaded host (mammal, plant, insect) is essential for the host's defense. For a successful countermeasure several mechanisms are employed. The distinction of the innate and the adaptive immune system of the host is important since the specialized defense triggers a response reaction according to site of infection and the tools at hand at that location. Cell surface molecules of the invader play a crucial role in this recognition process, and the type of immune reaction is dependent on the specific macromolecule.

The invasion of bacteria in mucosal membranes or the subepithelial cells triggers two kinds of innate immunogenic responses carried out by neutrophils. Neutrophils (leukocytes) like granulocytes engulf whole bacteria and dissolve them in their vacuoles and process bacterial components to the adaptive immune system. Bacteria are degraded via toxic oxygen intermediates by myeloperoxidase with NADPH in the oxygen-dependent mechanism. The oxygen-independent mechanism as a second route triggers the release of cationic antimicrobial peptides (CAMPs) upon sensing cell envelope glycopolymers via Toll-like-receptors (TLR) on dendritic cells and subsequent NF-κB translocation.

CAMPs are microbicidal positively charged small molecules able to perforate the invader's cytosolic membrane, and they are released as a response to invasion of the host cell. They are part of the innate immune system of vertebrates, fungi, plants, insects (e.g. cecropins) and bacteria. Prevalent in neutrophils of human tissue, e.g. mucous membranes and skin are defensins, a major class of human CAMPs. Other examples include cathelicidins (e.g. LL-37), which are produced in keratinocytes and polymorphonuclear neutrophils and thrombicidins. In low concentrations, however, CAMPs are able to attract leukocytes via a process called chemotaxis. CAMPs complement the transient pathogen response system.

Pathogen-associated molecular patterns (PAMPs) are microbe-related invariant macromolecules of polysaccharide nature and are readily recognized by pathogen recognition receptors (PRRs). PAMPs are exposed on microbial cell-
surfaces of Gram-positive as well as Gram-negative bacteria. For example, lipid A of LPS is recognized and transduces a proinflammatory signal in the immune response to activate against infection of Gram-negative bacteria\textsuperscript{179}. An analogous reaction is triggered when LTA is recognized, a signal referring to the infection of Gram-positive bacteria. One PAMP or one bacterium can be an agonist to several PRRs. Examples for PAMPs include PG, lipoarabinomannan, glycophosphoinositol, LTA, dsRNA, CpG DNA, flagellin, and others. PAMPs are expressed constitutively and represent conserved structures of germ-lines. These are distinct to the tightly regulated virulence factors, a response for evading the host defense system.

PRRs are highly specialized molecules of macrophages, neutrophils and dendritic cells recognizing PAMPs. They are variable membrane-spanning or membrane-anchored (e.g. CD14) protein-classes of the innate immune system\textsuperscript{180}. Some PRRs are either highly functionalized, i.e. N- or O-glycosylated, cysteine-rich, etc. or might comprise several conserved domain features, e.g. C-type lectin domains, epidermal growth factors, fibronectin types. Prominent representatives include scavenger receptors, integrins, lectins, Toll-like-receptors and others\textsuperscript{181,182}. One example for plant PRRs are PAMP-sensing LysM-domains. Utilization of PRRs is a first step in opsonization, phagocytosis, apoptosis, complement activation pathway, coagulation cascades and proinflammatory signalling. However, highly virulent pathogens like \textit{Yersinia pestis}, \textit{Salmonella enterica}, \textit{Legionella pneumophila}, \textit{Bordetella pertussis} are known to use the type III and type IV secretion systems to transmit effector molecules into macrophages in order to bypass host responses. The discrimination between pathogens and non-pathogenic commensal microflora is suppressed with anti-inflammatory cytokines by an unknown mechanism\textsuperscript{183-186}.

WTA is able to activate the complement cascade, leading to opsonophagocytosis. In the case of WTA the mannose-binding lectin (MBL) pathway is activated via WTA-sugar decorations in general, however for the classical pathway and generally for antibody-recognition exclusively $\beta$-GlcNAc is of relevance.
2.5 Mobile genetic elements (MGE)

As the name suggests, MGEs are mobile DNA-segments in prokaryotes. These gene clusters can either be inserted into a commensal plasmid or into the chromosomal genome. MGEs exhibit features that make them discernable from the specimen’s core chromosomal DNA. For example, transposases per se or site-specific recombinases are crucial factors by which they can be recognized. Deviating G+C-ratios is another such indicator. Adjacent virulence factors or resistance genes are also very common. It is being widely assumed that such adaptability and pathogenesis markers were inherited by horizontal gene transfer (hgt), although the mobility could have been lost in due course \(^{187-190}\). All MGEs, such as plasmids, temperate phages, chromosomal cassettes, pathogenicity islands or transposons are also passed on freely by vertical gene transfer (vgt).

Vgt describes the passing on of genetic information by means of replication to the progeny, hgt is encompassing all other modes of transfer of genetic information to related recipients. Three general mechanisms of hgt are known. Firstly, transformation describes the adsorption of genetic material from the environment by competent cell organisms when responsible systems for nutrient uptake are switched on. Secondly, transduction is a more specific mechanism. It occurs when temperate phages are transferring DNA upon entering the infectious state during the lytic cycle. Thirdly, conjugation is a complex mechanism by which pili are formed between a donor and an acceptor cell through which DNA is channelled. Conjugation is exclusively exploited by prokaryotes. This mechanism utilizes so-called integrative conjugative elements, which include transposases, integrases and other genes essential for conjugation. These loci can be as large as chromosomal DNA (megaplasmids) with independent origins of replication. Integrated conjugative elements are able to drag host chromosomal elements or other accessory genes during hgt. As an example, the occurrence of vancomycin-resistant \(S.\ aureus\) (VRSA) strains originates from vancomycin-resistant \(Enterococcus\) (VRE) through conjugation of the mobile genetic element TN1546. This transposon harbours the \(van\) operon and is activated upon vancomycin sensing. Gene products \(vanA\) and \(vanH\) are synthesizing a D-Ala-D-Lac intermediary with lower affinity to vancomycin,
whereas VanX is a dipeptidase specific for D-Ala-D-Ala targets. Other gene products of the van operon fulfill putative regulatory functions. Non-homologous recombination of accessory genes is an interesting phenomenon, as these incidents are currently believed to be the driving force in acquiring antibiotics resistances, making \textit{S. aureus} and related pathogens harmful entities.
2.6 Bacteriophages

The major autonomous components of a bacteriophage consist of a head, often a tail, and a tail end. The head is a proteinaceous capsid, which represents the repository of the viral DNA encoding the enzyme machinery necessary for the phage's initial preservation after invasion. The head contains the viral DNA and can either be of icosahedral symmetry or filamentous and. The tail is a hollow tube-like fiber-protein complex that is able to channel the DNA from the head to the tail tip. The tail end is generally a multimeric protein complex and has its function in attachment to the host's cell surface and permits penetration of the cell barrier for injection of the viral DNA into the cell. Some bacteriophages have additional tail fibers. The dimensions of a full bacteriophage can be up to 200x100 nm.

Taxonomic classification of bacteriophages is based on morphology, gene type, genome size and infectivity. Bacteriophages are primarily classified according to their family (tailed phages), on a second level according to order (overall shape, infection type, genome-size) and finally according to genus (specific host range, replication machinery). The genus is the most relevant classification, since it describes the least diversification. Phages can be directly determined according to their functions and mechanisms.

Phages of the Caudovirales order were originally classified according to their reaction to specific polyclonal antibodies. Hence, they were divided into eleven serogroups. With the technical development of EM-techniques, morphological differences were subsumed inside these classical serogroups. Another way to classify bacteriophages is by the lysogenic mechanism they adopt, as described below. This general classification roughly divides phages into sets of contractile tails (Myoviridae), non-contractile tails (Siphoviridae) and without tails (Podoviridae) (Fig. 6).

The parasitic nature of phages is determined by the intracellular multiplication of its own genetic code by exploiting the host's replicative biosystem. By an unknown mechanism, the phage adsorbs to the cell surface with its baseplate or tail fibers. Both baseplate and tail fibers are important determinants of host tropism. The adsorption by tail fibers is generic and weak, whereas the
attachment by the baseplate is specific and utilizes diverse cell surface components as could be shown for a series of Gram-positive bacteria, e.g. *Lactobacilli*\textsuperscript{191}, *Lactococci*\textsuperscript{192}, *Streptococci*\textsuperscript{193}. In the case of *Staphylococcus* spp and *Bacillus* spp, it could be shown that the presence of teichoic acids is crucial for virus adsorption\textsuperscript{194,195}. The cell barrier is overcome by the concerted action of each domain, ultimately leading to the injection of the viral genome and leaving the protein husk outside the cell. This mechanism is in stark contrast to the invasion mode of eukaryotic viruses, which enter cells through endocytosis. After injection of the phage DNA, the fate of the invaded host is determined either by lytic or lysogenic infection and the invasion takes its course.

![Schematic overview of the most frequent genera in the Caudovirales order of tailed phages.](source: Maniloff, Jack et al.; Taxonomy of bacterial viruses: establishment of tailed virus genera and the other Caudovirales, H. Arch. Virol. October 1998, vol. 143, no. 10, 2051-2063; copyright permission courtesy of Archives of Virology by Springer Link. Reproduction of figure with permission of Springer Link via Copyright Clearance Center.)

### 2.6.1 Lytic cycle and lysogenic cycle

The goal of the lytic cycle is to multiply and eventually destroy the host. In order to achieve this, phages hijack the transcription machinery and produce viral mRNA in the ecliptic phase, meanwhile preventing the host's gene
transcription and translation. In due process the host's chromosome is degraded. After phage assembly is completed in the late phase of the lytic cycle, lysis proteins are expressed until the host cell bursts and phages are released.

The lysogenic infection is in many cases characterized by the incorporation of the viral DNA into the host's chromosome by site-specific recombination and sustenance in a dormant state, in which the transcription is blocked by a phage-encoded and phage-specific repressor. In this state the phage is referred to as prophage and the cell as lysogen. The replication system of the host is not hindered, and the viral genetic code is passed on together with the bacterial genome to its progeny. However, upon an induction event the prophage is activated and enters the lytic mode. Induction parameters are mainly in adverse to the host's thriving. Lytic mode is generally activated by the host's exposure to mutagenics, UV-light, ionic radiation and desiccation. Responsible for the induction is the protease RecA, which cleaves the viral repressor and enables the transcription of the prophage genome.

2.6.2 The order Caudovirales

Caudovirales comprise three major families of characteristically tailed phages: Myo-, Sipho- and Podoviridae. Siphoviridal dsDNA is comprised of 50 kb with a non-contractile tail. Myoviridae also have dsDNA of up to 280 kb, but a contractile tail instead. Podoviridae have short tails and dsDNA of around 40 kb. Before DNA is packaged, the procapsid is built by scaffolding proteins and a major capsid protein during assembly. After concatamerization of DNA, it is channeled into the procapsid, an ATP-dependant step carried out by the terminase-complex consisting of the small TerS and the large TerL proteins. The tail and the baseplate are attached to complete the virion.

2.6.2.1 Siphoviridae

One example for a Siphovirus are the λ-like phages. They utilize a flexible DNA ejection system. After the cytoplasmic membrane has been punctured, the phage opens its cap, which is located between the head and the tail, in order to tunnel the DNA into the cell. From here the viral DNA circularizes by ligation at its cos (cohesive end site) sites by exploiting a host ligase. The first proteins to be expressed at the I-promoter site are the regulators
cro and cll. Generally, elevated cellular levels of the regulatory gene product cro are liable to promote lysis. Cro itself is dependant on environment variables that control its production. If the amount of cro exceeds that of cll, the phage enters its lytic pathway, if vice versa, the latent stage is enabled. In the latent stage an integrase is expressed, and the phage DNA recombines at chromosomal attB-sites regulated by cl. Either by degradation of cl and subsequent upregulation of xis and int or by entering the lytic pathway right from the beginning of infection, the circular virion is replicated. In the late stage, DNA is concatamerized, packaged, and finally the cell is being lysed.

2.6.2.2 Myoviridae

The Myovirus phage μ has a contractile tail and six fibers at its baseplate, which allow for host tropism. The modification of tail fibers occurs either by inverting the direction of genetic transcription or by mutagenesis. After binding of the phage to the cell surface, a puncturing device cuts through the outer membrane and cell wall and fuses with the inner membrane. The outer tail tube contracts, pushing the dsDNA through the inner membrane. Once inside the host, the co-invading viral protein N holds the linear DNA circularized by keeping the ends together. This diminishes host nuclease activity on the bacterial 3'-ends of the phage DNA. Before going into replication or latency mode, the DNA is integrated into the bacterial chromosome as a transposon element by MuA and MuB. The early genes MuA and MuB are responsible for DNA insertion and target site selection. MuA is a Mg-ion-binding recombinase of the DDE family and MuB is the target DNA activator. After the integration process is concluded, expression of repressors Ner facilitates latency by preventing early genes from expression or Repc controls further viral transposition yielding in up to 100 new copies of the viral transposon with the assistance of MuA and MuB. After this event, the expression of late genes leads to the formation of the structural phage proteins, DNA packaging and eventual lysis.

Coliphage P1 is another temperate bacteriophage of the myoviridial family. It has an icosahedral head and stores a dsDNA of around 90 kb size. It is distinct because of its different state of the virion in the lysogenic state. An
attP-site is lacking, therefore the linear DNA is homologously recombined at its terminally redundant sequences using the viral Cre-\textit{loxP} system, a specific recombinase and its recombinase site. The result is a prophage plasmid, which during cell division replicates via its Holliday junction and promotes segregation to daughter cells by so-called binary fission\textsuperscript{197-199}. 
2.6.3 *S. aureus* Pathogenicity Islands (SaPIs)

Pathogenicity islands represent gene loci associated with specialized virulence factors, resistance genes and enterotoxin genes, e.g. toxic shock syndrome (*tst*). Of the identified SaPIs, best known are SaPI1 and SaPIbov1. SaPIs are stably integrated into the *S. aureus* host genome at *attC*-sites with their own autonomous replicons with the whole cassettes ranging from 14-27 kbs.

SaPIs have the ability to utilize structural proteins and enzymes of phages of the *Caudovirales* order to package their DNA into transducing particles. The generation of transducing particles occurs during the lytic phase of the phage upon induction, e.g. SOS-response, superinfection of a non-lysogen SaPI by a helper phage or concomitant infection of SaPI and bacteriophage. The general act of piracy, describing the exploitation of host genomic products to suit the needs of the invading virus, is reversed. The host becomes the “pirate” and the virion a mere “helper”. The helper phages are commonly found to be recruited from the family of *Siphoviridae*, and each SaPI has a specific helper phage range. The formation of SaPI-particles follows a mechanism similar to the general assembly pathway of *Caudovirales*, although one major difference of most SaPI-particles to common phages is the smaller size its head. Responsible for this size redistribution are additional scaffolding proteins (SPs). Most SaPI gene-clusters share the so called capsid morphogenesis genes (*cpm*), which assist in the formation of the small SaPI-procapsid by replacing SPs of the general virion assembly. However, exempt from *cpm*-assisted assembly are SaPIbov2 and SaPIbov5.

All of these phage-related chromosomal islands contain common regulatory genes, among them genes for excision (*xis*) and integrases (*int*), which make them highly mobile intra- and intercellularly. In the absence of phages, SaPIs circularize at redundant termini before integrating back into the host chromosome. The SaPI is not excised and replicated, lest helper phage repressors act on *stl*. When *stl* is not repressing *str*, the process of excision turns on. The viability of SaPIs depend on this regulation, since an excised state during cell proliferation would lead to their loss. Eventually, SaPI replication is carried out at the *ori* site via Rep, a replication initiator with
helicase activity, and leads to concatamer formation. Subsequently, DNA packaging of SaPI-particles occurs analogously to the DNA packaging in phages. The pac-site of SaPI-DNA is recognized by SaPI-encoded small subunit TerS of the terminase-complex and translocated by the phage-encoded large subunit TerL inside the procapsid. When SaPI-particles are invading new hosts alone, they integrate into the new chromosome at specific sites. However, when invasion is accompanied by phages, they replicate immediately to form new particles.

2.6.4 Siphoviridial architecture

2.6.4.1 The siphoviridial tail end polymorphism
Parts of the baseplate complex of a Siphovirus are the distal tail protein (Dit), the tail associated lysin (Tal), the tape measure protein (Tmp) and the receptor binding protein (Rbp). In the structure of the lactococcus-infecting siphoviridial bacteriophage TP901-1, 18 trimers in a total of 54 RBPs are building the baseplate protein lower (BppL) unit, which is connected to the core by 6 molecules of baseplate protein upper (BppU) in a six-fold symmetrical arrangement around the baseplate core. Intermolecular contacts between the oligomeric proteins in several copies each builds up a complex of 1.8 MDa (Fig 7).
The baseplate of lactococcal p2 is another example for siphoviridial diversity. Here, 6 RBPs are located laterally on the distal tail end in a fashion comparable to the TP901-1. For the p2 baseplate, two distinct conformations are found, active and resting, which are switched on and off by the presence or absence of calcium ions. Apart from that, p2 is rather comparable to TP901-1, especially because of the presence of a baseplate. The genetic locus encoding the proteins for the baseplate is well conserved among Lactococcus spp. infecting Siphoviridae. The prophage gene cluster harboring structural segments, head and tail for instance, is located amid open reading frames, which are also responsible for lysogeny, replication, packaging and lysis. Conventionally, the baseplate is related to the tail, since not every siphophage carries a baseplate in contrast to Myoviridae, which in general do have baseplates.
The *B. subtilis Siphovirus* SPP1 has a peculiar trunk-like tail tip that is bridged via a hinge to the tail cap. The tail tip is attached to the tube cap distal to the head enabling flexibility independent of tail tube orientation and serves as an adhesion tool, recognizing *B. subtilis* YueB-ectodomain. This tail tip is larger than the usual tail spike protein of other phages, and another difference is the loss of it after adhesion, leading to conformational rearrangements of the tube cap, which facilitates the opening of the tube channel for tunneling viral DNA into the host.

Another baseplate structure is to be found in *Siphovirus* coliphage T5, a well-studied model phage of Gram-negative enterobacteria. Besides the presence of a tail tip like in SPP1, this phage also utilizes tail fibers for cell surface recognition.

### 2.6.4.2 Receptor binding proteins

Tail ends of phages of the *Siphoviridae* family are structurally multivariant determinants of selectivity. These proteinaceous molecules, which are harboured in the tube distal end define host tropism and invasiveness. For lactococcal phages, slight alterations of a given Rbp elicit shifts of selectivity to closely related species from a narrow evolutionary scope. Relatively small structural modifications of the Rbp lead to the receptor recognition shifting from dairy phage group P335 to group 936\(^{200,201}\), both of which are distinct with respect to DNA homology and phage morphology\(^{202}\), as is seen for Rbp structures of Tuc2009, TP901-1, p2 and bIL170, for instance (Fig 8). The availability of a cell surface receptor can lead to natural adaptation of Rbps to recognize additional cell surface molecules as receptors, as is reported for phage λ and its recognition of OmpF apart from its cognate receptor LamB\(^{203}\). Other examples for such occurrences are also common for other phage families\(^{204-207}\).
2.6.4.3 Teichoic acids as receptors for bacterial viruses

For invasion to happen, an irreversible recognition and successive adhesion step needs to take place on the cell surface. Formerly, LTA was believed to be the infection receptor of phages. However, phages of *Listeria spp.* and allegedly *L. lactis* are reported to recognize specific side-chains of teichoic acids as receptors\(^\text{208,209}\). *Listeria* phages temperate A118 and virulent P35 evidently recognize sugar moieties (rhamnose, GlcNAc) on poly-RboP-C2 or -C4. Whereas for A118 siphophage rhamnose is crucial, P35 requires both, GlcNAc and rhamnose, for adhesion. For lactococcal phages of groups c2, 936 and P335 it is assumed that they might be binding sugars or even glycerol units of GroP-type WTA\(^\text{200}\). For several staphylococcal phages, recognition of GlcNAc regardless of their anomeric configuration has been reported\(^\text{195}\).
2.7 Aims of Thesis

The general intention of this thesis is to describe and discuss recently identified WTA-affiliated proteins, and to highlight structural similarities and differences to other archived proteins of interest. The molecular analysis of these structures could assist in understanding WTA better. The characterization of novel domains or features specifically could identify molecular patterns to broaden our yet small knowledge of WTA needed for the development of treatments and vaccines against widely spread WTA bearing germs. Structure analysis of proteins associated with WTA biosynthesis are rare and the contribution through this thesis could serve as a puzzle piece to generate models of in vivo interdependencies of WTA.

Concretely, the structure solution and biochemical characterization of TarM of the biosynthesis pathway of RboP-type WTA and Gp45, the receptor binding protein of S. aureus infecting phages, which recognizes sugar moieties on WTA are the subject matters of this thesis.
3 RESULTS AND DISCUSSION

3.1 TarM

3.1.1 Isolation and crystallization
TarM was produced in *E. coli* and lysed by ultrasonication. After affinity chromatography and size exclusion chromatography (SEC), the fused thioredoxin protein was removed and TarM isolated.

Crystallization conditions were screened in a sparse setup, and a successful initial hit was later refined to obtain bigger crystals. For a binary complex crystal, purified TarM was incubated with UDP-GlcNAc (15 mM, 1 h, 4 °C) and a new crystallization condition was screened and refined. Crystals were vitrified in liquid nitrogen. For anomalous dispersion experiments, wild-type crystals were treated with KI (30 mM, 30 ', 16 °C), backsoaked in crystallization condition and vitrified\textsuperscript{210}. (See publication in appendix for details)

3.1.2 Diffraction data collection and model refinement
Data were collected on a PILATUS 2 M hybrid pixel detector using synchrotron beam line X06DA at the Swiss Light Source (SLS) super-bending magnet (2.9 tesla). The XDS package was used for data reduction. The crystals belong to spacegroup P6\textsubscript{3}22 with cell parameters of $a = b = 123.7$ Å and $c = 223.3$ Å. One TarM monomer (58 kDa, 493 residues) comprises the asymmetric unit (asu), which has a solvent content of 74%. Data collection of anomalous data on the beamline was performed by irradiating the crystal in predefined wedges at low electron dosage.

Alternating cycles of COOT model building and REFMAC5 or PHENIX refinement revealed additional residues, which were included in the refinement until convergence had been achieved. The final model includes residues 1–493 (PDB ID 4wac, 4wad)\textsuperscript{210}. (Data collection and refinement statistics can be found in the publication in the appendix).

3.1.3 Overall structure and domain organization of TarM
TarM is the first enzyme structure in the biogenesis pathway of poly-(RboP)-type WTA to be reported. TarM is an $\alpha$-O-N-acetylglycosyltransferase specific for RboP and a member of the RboP biosynthesis pathway, also found in other
related species and strains. A structural study of this enzyme identifies a glycosyltransferase of type B (GT-B). TarM is a homotrimeric molecule (174 kDa) with a broadened oblate shape according to combinatorial biophysical characterization via dynamic light scattering (DLS), SEC and bioinformatic calculations (PISA-server\textsuperscript{211}). The C-termini protrude from the molecular center, which is constructed by a threefold symmetry axis, so that the C-termini surround a large hollow center. Each TarM monomer consists of one domain of unknown function (DUF) and one GT-module. This DUF1975 is inserted into the GT. Three DUFs stack their 10-stranded antiparallel $\beta$-sheets against each other and thus exclusively form the assemblation interface. This stacking allows for a 30 Å wide opening at the exterior of the central hub domain, the inner side on the other hand is closed tight. The interface of this domain buries a total surface area of 777 Å$^2$ (PISA-server).

The GT-module has the typical GT-B fold of anomer-retaining enzymes as is evident from the bipartite architecture separated by a flexible linkage of ca. 10 residues. Crystal contacts are prevalent in the N-terminal protein part, however these are becoming drastically less in the C-terminal protein parts, the latter subsuming extraordinarily high B-factor values. A DALI query of this second major domain returned typical GT-4 class homologs, e.g. MshA (Z-score 33.1, pdb 3c4q) and BshA (Z-score 32.9, pdb 3mbo). The unexpectedly overall low Z-scores are presumably a result of the conformational flexibility of the GT-module due to a flexible 10-residue linker. Each GT-B crystal structure is frozen in a slightly different relative orientation of N-termini against C-termini, exacerbating a complete superposition of 2 GT-Bs. Known acceptor substrates for GT-4 class enzymes range from $1$-L-Ins-1-P (inositolphosphate) to LPS and to S-layer glycoproteins, and many others. The C-terminus is overall less well defined by electron density than the remainder of the protein, but the electron density for UDP-$\alpha$-GlcNAc is nevertheless unambiguous.

### 3.1.4 Structural and mutational analysis of the active site

The sugar-transfer in GT-4 class glycosyltransferases is typically carried out at the interface between the N-terminus and the C-terminus. Binding of UDP-GlcNAc does not lead to any remarkable structural rearrangements (r.m.s.d. value of 0.81 Å for 493 aligned residues). According to the structural model,
UDP-GlcNAc is located in a cleft formed by five loops. The anomeric center carbon C1 lies in close proximity to the carboxylate function of Glu-403. Arg-326 and Lys-331 side chains are stabilizing the phosphate moieties. The uridyl unit is held in position by several prominent loops. The ribose is anchored to the Glu-411 side chain. Specificity for uracil is generated through several polar and hydrophobic interactions, namely Tyr-382, Thr-383, and Pro-386 and the side chain of Ile-324 appears relevant here, too. In summary, the UDP-GlcNAc binding pocket is well conserved with other GT-Bs.

For a mutational analysis of the UDP-GlcNAc binding site, RN4220 was complemented with various tarM variants to be used as hosts for phage $\phi$11. EOP (efficiency of plaquing) utilizes a link between $\alpha$-O- and $\beta$-O-GlcNAc glycosylated WTA and bacteriophage adsorption to S. aureus hosts. A high EOP corresponds to a high WTA glycosylation, whereas inactive complements lead to a low degree of glycosylation.

This setup has been probed for active site variants E403A, K331S, R326S, and H249A. The integrities of the respective protein structures were verified via circular dichroism (CD) and DLS. Glu403 and Lys331 were found to be crucial for catalysis, whereas the latter two residues, Arg326 and His249, appear to be of lesser importance due to residual plaquing capacity.

### 3.1.5 DUF1975

A BLAST sequence search of protein databases finds DUF1975 only in TarM homologs of other Gram-positive bacteria. As TarM-mediated WTA glycosylation is thought to constitute a general pathway in Gram-positive bacteria with RboP-WTA, it seems likely that the DUF1975 fulfills a similar role in these related organisms. One noteworthy feature is the site DUF1975 is inserted. Here, the homologs without the insertion show a small region of undefined structure. Incidentally, this point of insertion also happens to be the dimerization site for MshA and BshA. The direct distance of two neighbouring active sites in the TarM trimer is approximately 72 Å, which corresponds to about eight or nine ribitol-units of an extended chain. Thus, a single TarM trimer could simultaneously glycosylate the same poly-RboP substrate at different locations. It is not currently known which RboP units in the long
polyribitolphosphate chain are glycosylated, but it is likely that the glycosyltransferases acting on WTA have a mechanism that enables them to move along the polyribitol chain and selectively glycosylate specific units. Single amino acid substitutions (K136S, N138Q, N180W), a double (V159Y/C164R) and a triple (V159Y/C164R/K136S) mutation near the trimer interface were studied in EOP tests. We observed a substantial decrease in TarM EOP for K136S as well as the double and the triple mutant. However, where, under *in vitro* and EOP conditions, E403A was not capable to glycosylate WTA, the case for the triple mutant was different. Interestingly, under *in vitro* conditions the triple mutant shows a stronger enzymatic capability than in the EOP assay. Unfortunately, a conclusive explanation for this phenomenon is not possible with the data at hand. Expanding on this, the CD spectra of the double and the triple mutants show an additional shoulder around 205 nm, indicating a noteworthy structural alteration.

The streptococcal GT-B GtfA (pdb 4pqg) has also a DUF1975, and that is also inserted into the GT module at a similar location. Furthermore, GtfA DUF1975 is the only structure in the DALI database with any significant structural homology to the TarM domain (Z-score 11.6, sequence identity 16%). However, GtfA is clearly monomeric, and the enzyme also does not act on WTA. DUF1975 from GtfA and TarM are very much the same module structurally, however the mode of action and the specificities for both enzymes are diverging dramatically. Here, it is almost impossible to deduce from one DUF1975 the function of the other one.
3.2 φ11 Receptor Binding Protein Gp45

3.2.1 Characterization of Gp45

Gp45 could be identified by bioinformatic analysis. The protein shares identity of about 45 % with ORF636 of PhiSLT, an annotated cell adhesion molecule. Moreover, antibodies were raised against recombinantly purified Gp45, and an immunogold-labeling was performed. Transmission electron microscopy verified the location of Gp45 and Gp54, a suspected BppU homolog, to be within the tail tube tip. The antisera exhibited a dose-dependent effect when blocking the tail tip with antisera against both Gp45 and Gp54. Neutralization of cells with recombinant Gp45 showed a comparable effect for both variations of experimental setups, plating efficiency of preincubated cells and flow cytometry analysis with biotin-labeled recombinant protein. Mutants deficient in WTA-biosynthesis were also tested, and a direct correlation could be verified between the α- or β-O-glycosylation capacity of the cells and the Rbp binding ability. It was established that both 6-O acetylation\textsuperscript{213} of PG-MurNAc or WTA integrity in general need to be fully functional for phage adsorption to occur. Furthermore, YueB homologues in \textit{S. aureus} denoted Pip (phage infection protein) were ruled out as membrane-protein receptors (see publication in appendix for details)\textsuperscript{214}.

3.2.2 HHPRED analyses of baseplate components

HHPRED analyses of φ11 baseplate components Gp43, Gp44 and Gp54 revealed similarities with components of the lactococcal phage TP901-1 baseplate. Gp43 is supposed to exist as a hexamer and represents the φ11 distal tail protein (SPP1 Dit) ring and Gp44 as a trimer forming the tail-associated lysin (Tal) N-terminus and extension, while Gp54 N-terminus (the functional equivalent of BppU N-terminus) may form a second ring\textsuperscript{215,216}. Furthermore, the N-terminal segment and the first helical bundle of φ11 RBP are structurally homologous to the N-terminal part of the phage TP901-1 or Tuc2009 Rbp trimer, a segment that connects BppU to BppL. The TP901-1 Dit hexamer was fitted together with the BppU N-terminus into the electron density map. The ring of the Dit had appropriate dimensions to fit the map above the Rbps. In contrast, the structure equivalent to BppU could not be
fitted unequivocally as the internal density is not defined sufficiently.

It is not unusual that Dit and Tal could have been conserved through evolution, a phenomenon shared by other components such as the capsids MCP, the connector - as well as the tail MTP. Only the periphery of Dit (its C-terminal domain) and the Tal extension (e.g. a C-terminal fiber) could have been adapted to specific phage infection-style requirements. In contrast, φ11 Gp45, the Rbp, does not exhibit analogy with other phage Rbps, in particular with those from lactococcal phages that also bind to saccharidic receptors (See publication in appendix for details)\(^2\)\(^1\)\(^4\).

3.2.3 Isolation and crystallization

Overproduced Gp45 was isolated according to standard procedures in a purification process involving two steps Immobilized metal affinity chromatography (IMAC) and SEC. The final concentration was adjusted using a concentration-cell.

A first round of sparse matrix condition screening had been performed using a pipetting robot to pipet 10x96 commercial conditions for protein concentrations adjusted to several values between 5 and 15 mg/mL. The successful crystallization conditions were refined by varying pH, additive and cryoprotectant amounts in small ranges. The refinement in 24-well plates using the hanging-drop variation gave bigger crystals (ca. 500x200x50 \( \mu \)M). Derivatives were prepared by soaking native crystals in crystallization solution with 2 mM Ta\(_6\)Br\(_{12}\) for up to 2 weeks before backsoaking briefly in crystallization condition and vitrification in liquid nitrogen\(^2\)\(^1\)\(^6\). (See publication in appendix for details)

3.2.4 Diffraction data collection and phasing

Data for native and derivative crystals were collected at the Swiss Light Source (SLS) on beamline X06DA (PXIII) using a PILATUS 2M hybrid pixel detector. The inverse beam diffraction data collection method was applied for derivative crystals. For the determination of peak, inflection, high-remote and low-remote wavelengths from fluorescence spectra, the program CHOOCH was adjusted to the absorption edges of Ta-L-II and Ta-L-III.

Collected data were processed with the XDS-package. Structure
Results and Discussion

determination was performed by a combination of SIRAS and non-crystallographic symmetry (NCS) averaging: Anomalous data were processed according to MAD, SAD, MIRAS and SIRAS protocols using SHARP/autoSHARP. Initial heavy atom coordinates and B-factors found with SHELXDE were reedited with the SHARP-module Sushi and refined until electron density maps showed good contrast. The outcome was then used to further refine the atomic substructure and calculate electron density maps for the protein. Resulting electron density maps were inspected for characteristic features of a protein. The outcomes of the various phasing protocols were compared and the map derived from the SIRAS protocol was taken for further improvement. The tables in the appendix section document the most relevant phasing results.

3.2.5 NCS and phase improvement

A threefold NCS was elicited from the self-rotation function via polarrfn (ccp4), giving a strong signal for rotation in reciprocal space for Eulerian angles (α=357.4, β=63.6, γ=110.6) corresponding to polar angles (ο=37.4, φ=33.4, κ=120.1). Applying GETAX (ccp4) using the above set of coordinates, the orientation of the major rotation axis and all of its locations in the asu could be obtained in another set of translation vectors. These results were combined in four rotation-translation matrices, which could be used for improving the map further.

Due to multiple domains the large dimensions of the protein made it necessary to split the molecule for averaging purposes into four domains: 2 different subdomains for the stem, the shoulder domain and the head domain. Four different masks were generated to include at least these 4 structurally different regions. To verify the accuracy of the unrefined matrices in real space, a testwise generation of correlation maps with the combined NCS-operators using maprot (ccp4) on the phasing solution was tried. By comparing the unbiased map with the correlation maps the NCS-operators could be assigned to the related domains. The refining of these NCS-matrices with IMP and a testwise averaging with AVE converged the incomplete input electron density to above 80 % of correlation after averaging, indicating a good outcome for the averaging process. The increase of electron density and
the increase of percentage correlation after each averaging round lead to visually interpretable maps, confirming the correctness of the results.

For combined density improvement, each subdomain was integrated with their respective NCS-matrices into a DM script. Starting at 5.8 Å, 80 cycles of consecutive solvent flattening, NCS-averaging, histogram matching and phase extension to a final resolution of 2.9 Å resulted in a final map that was clearly distinguishable from the unbiased calculated map. Main chains and side chains became interpretable so that a large portion of the protein model could be docked. Initial refinements were carried out with REFMAC5 and PHENIX after each step of COOT model building. Finally, refinement was pursued with autoBUSTER leading to $R_{\text{free}}/R_{\text{work}}$ values of 21.1 % and 17.5 %.

### 3.2.6 Gp45 structure (Rbp)

Although the electron density was generally of good quality, a few loops of the propeller domains (amino-acids 140-440) have very weak electron density, explaining the persistence of a few outliers in the Ramachandran plot. The Rbp of $\phi$11 assembles into an elongated homotrimer, with overall dimensions of approximately 160×120×100 Å. The structure consists of the N-terminal stem region, the central shoulder region and the C-terminal head region. A total buried surface of ~35,000 Å$^2$ is found for the trimer according to PISA (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver). Most of the buried surface area is concentrated in the stem and C-terminal regions, while the propeller domains engage in few intermolecular contacts (See publication in appendix for details).²¹⁶

#### 3.2.6.1 Stem domain

Overall, the stem is composed of 4 triple-helical bundles, connected either by $\beta$-hairpins or undefined regions. Between the first two bundles, an Fe$^{3+}$ is positioned on the 3-fold axis of the first helix bundle and octahedrally coordinated by the side chains of His42 and His50 from each of the three monomers. The HisN$\varepsilon_2$-Fe distances range from 2.19 Å to 2.32 Å. This His-Fe-octahedron is surrounded by Glu46, Arg43 and Gln54 from all chains. Similar Fe$^{3+}$ binding geometries and bond lengths are known of viral proteins, i.e. membrane-piercing spike proteins of phages P2 (PDB code...
Results and Discussion

3qr7) and ϕ92 (PDB code 3pqh), as well as in the receptor-binding domain of the long tail fiber of phage T4 (PDB code 2xgf). ϕ11 Rbp could act in a similar fashion as it was proposed for phage P2, ϕ92 and T4 spike structures by enforcing the protein to pierce the cell wall. While the first and second triple helices are collinear, a hinge is introduced into the structure, which forms a sharp angle of ~30° between the second and the third bundle. This angle allows for contacts between stem units and between the stem domain and the shoulder domain. The hinge geometry is such that the sequences of the three helices of bundles two and three remain in phase. Bundles α3 are followed by extended hairpin-structures, and helix α4 completes the stem (See publication in appendix for details).

3.2.6.2 Shoulder domain

The stem is followed by five-bladed β-propeller modules of the shoulder domain, which encompass residues 142–439 and form the midsection of the protein, and it is linked via a short helix (residues 425–432) to the C-terminal domain. Between the third helical bundle and the head domain of ϕ11 Rbp is the shoulder, which contains three five-bladed propellers in an area that is ~100 Å wide and ~40 Å thick. Each interface between two propeller domains buries a surface area of only 457 Å² from solvent, and much of this surface is buried due to a helix-helix contact at the center of the trimer axis. The plane of the propeller is not perpendicular to the 3-fold axis, but is tilted upwards by an angle of ~30°. This tilt improves access to the lower face of the propeller, and this might be linked to the function of Rbp in interacting with ligands.

Sets of four anti-parallel β-strands form each blade, and the N-terminal β-strand closes the fold by forming the final blade (blade 5) with the three C-terminal β-strands of the domain. Most similar folds are shared by enzymes that mediate degradation of carbohydrates, e.g. glutamine cyclotransferase from Zymomonas mobilis (Z=16.1, r.m.s.d.= 3.1 Å, pdb 3nol) or possess such endosialidases of coliphages K1F and phi92. Putative Rbps with β-propellers include a five-bladed propeller of the RBP-P2 head domain of phage PRD1, a Tectiviridae member infecting Gram-negative bacteria (PDB...
code 1n7u) and the C-terminal domain gp131C of the *Pseudomonas* myophage PhiKZ, forming a seven-bladed β-propeller domain (PDB code 4gbf).

Using the *Zymomonas mobilis* β-propeller structure to superpose onto the Rbp propeller domain, the active site of the enzyme revealed a pocket into which a GlcNAc molecule could be fitted. Three water molecules are indicating hydroxyl groups of a modelled GlcNAc molecule. The cavity is lined with polar residues (Gln165, Thr211, Gln330) and apolar residues Met164 and Met329, which could accommodate the modelled GlcNAc molecule (See publication in appendix for details).216

### 3.2.6.3 Head domain

The head domain contains two structurally similar domains, which are each formed by three five-stranded anti-parallel β-sheets. The two C-terminal tower domains (residues 440–541 and 542–636) form a structure of dimensions 60 × 50 × 50 Å. These two domains are very similar in structure (Cα r.m.s.d.=1.7 Å). A DALI search performed with these domains returned the highest Z-score (Z=6.1; r.m.s.d.=3.2Å) for uracil-DNA glycosylase inhibitor, a small all-β monomeric protein (PDB code 2ugi). The two structures essentially share the same anti-parallel β-sheet but differ in their oligomeric state and their surrounding structural features. The identity (9%) is not high enough to assign possible functions to the C-terminal domains.
3.2.7 Negative staining electron microscopy of the φ11 baseplate

512 images of φ11 were collected, and 778 baseplate particles were boxed. The final map has a resolution of 23 Å and accommodates six φ11 Rbp trimers in lateral position, when the hinge angle between the second and the third helix bundles are increased to a value of ~90°. The correlation is 0.845 with 95.5% of the atoms inside the map, calculated for a Rbp orientation fit with the head domain inclined towards the bottom of the baseplate. A “heads up” conformation results in a correlation of 0.826 with 74.9% of the atoms inside the map. However, it has to be taken into account that the “heads-down” conformation is not a completely non-flexible state. The angle of the stem might in fact vary from the minimum observed in the X-ray structure to larger values when the phage scans the host’s surface by moving the Rbps around the calculated average position for adhesion to the specific receptor. The remainder of the Rbp structure was left unaltered, and the modified trimers fit well in a peripheral region of the map that could accommodate the triangular shape of the shoulder domain.
4 FUTURE RESEARCH

4.1 TarM

TarM was found to be a GT module in a novel fold. A highly resolved ternary structure with the acceptor substrate would facilitate discerning a mechanism for the glycosylation of WTA and a probable involvement of the DUF1975 in this enzymatic action. It could be shown that one RboP molecule fits in the TarM structure (Simon Huber et al., Data not shown). A TarM structure of around 3.5 Å could pinpoint a density for one RboP-unit on the bottom side of the DUF1975 domain opposite to the sugar-binding cleft. This spot is located inside the cradle which is put up by the C-terminal lobes. However, a better resolved structure, probably with more than one RboP-unit, could allow for a more detailed analysis.

DUF1975 domain functionality was characterized as an oligomerization interface and proposed to be a domain able to present substrate molecules to the activated sugar ready to be transferred. According to enzyme specialization this domain could be able to bind conserved structures. Two substrates specific for DUF1975 are known: RboP-type WTA in the case of TarM and PsrP, a surface adhesin, in the case of GtfA. For a systematic comparison more DUF1975-variants are crucial to derive any sensible deductions from a bioinformatical approach. Although in the case of WTA there is generally much information, the role of WTA is not clarified and does not allow for appointing a role in the environment of the bacterial envelope to which it clearly belongs. However, the only role for DUF1975 we can report at that time is confined to the assemblation capacity of three GT modules and a putative site for binding of a RboP acceptor chain. Conserved portions of this domain in the vicinity of the identified active site region definitely indicates an acceptor site here.

4.2 Gp45

A novel fold of three distinct domains, stem shoulder and head, was found in the case of the Gp45 structure. Unexpectedly, the stem contains a bound iron. The function of this iron is unknown as its location differs from irons found in other spike proteins. It would be interesting to see if the stem structure integrity falls apart, when the iron is removed from the coordination center. This could either be achieved by using chelating agents on an isolated solute form or by using
culture media devoid by iron but rather using a non-toxic heavy metal of similar size. Both setups could be conducted by biophysical methods in a way that allows for transient processes to be inspected. The binding motif and other overall structural parameters could be studied structurally when inserting an alternative coordination center and compared to wild-type values.

The invariable TP901-1 BppU protein, seems to be present in parts in Rbp (Gp45) or in the subsequent protein (Gp54). This finding suggests an evolutionary interchange even between remote phages with different hosts. Since the role of Gp54 is not documented, it is difficult to predict any functions for it. Presumably, a subdomain of phage φ11 Gp54 might accommodate the Rbp N-terminus, comparable to TP901-1 BppU. Another isolation attempt at an in-vivo preformed or in-vitro produced complex between Gp45 and Gp54 could be taken into account. An appropriate SEC-column (e.g. Sephacryl) allowing for separation ranges around 1 MDa would be of crucial material demands in a protein purification procedure of the anticipated high molecular weight Gp45-Gp54 complex.

The molecular orientations as hinted by the electron microscopy low-resolution structure gives and the five-bladed propeller domain suggests roles for the Gp45 subdomains as exposed to the host cell surface. Although, a well-defined cavity in the size of a monosaccharide was identified, a much larger crevice surrounding this cavity could harbour a bigger molecule, of which GlcNAc might be only part. Although soaking and cocrystallization experiments with GlcNAc were not successful, modelling suggests a reasonable location for the GlcNAc binding site in the shoulder domain. However, further structural data are necessary to develop this hypothesis.

A large volume of the EM map remains to account for the Gp54 and for the Tal. When attaching the trimeric Tal N-terminal domain below the Dit hexamer, the three carbohydrate binding modules (2wao) identified by HHpred project in the direction of the tail tip. These three bulky modules should fill the electron density map in between the six RBP trimers. The functional investigation of the singular domains could be carried out analogously to the procedure as outlined before in the case of TP901-1 BppL, by connecting camelid antibody avidity and
enzymatic analysis. Since multiple nanobodies recognize several epitopes of a
given molecule, a structural analysis of nanobody-complexes in combination with
competitive enzyme assays would help in the characterization attempt.
5 REFERENCES

1 Baron, S. Medical microbiology. 4th edn, (University of Texas Medical Branch at Galveston, 1996).


<table>
<thead>
<tr>
<th>Reference</th>
</tr>
</thead>
</table>
| Reichmann, N. T. & Grundling, A. Location, synthesis and function of glycolipids and...


References


doi:10.1038/nsb858 (2002).


Comstock, L. E. & Kasper, D. L. Bacterial glycans: key mediators of diverse host


Spinelli, S. *et al.* Modular structure of the receptor binding proteins of Lactococcus lactis


215 Veesler, D. *et al.* Structure of the phage TP901-1 1.8 MDa baseplate suggests an

## 6 APPENDIX

### 6.1 Gp45 crystal data and phasing statistics tables

<table>
<thead>
<tr>
<th>DATA COLLECTION</th>
<th>crystal 1</th>
<th>SIRAS reference (crystal 2)</th>
<th>Native (crystal 3)</th>
<th>crystal 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>wavelength (Å)</td>
<td>1.25472</td>
<td>1.0000</td>
<td>1.06754</td>
<td>1.25511</td>
</tr>
<tr>
<td>absorption position</td>
<td>peak</td>
<td>-</td>
<td>-</td>
<td>inflection</td>
</tr>
<tr>
<td>space group</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>cell dimensions</td>
<td>87.65, 89.65, 93.73, 105.7</td>
<td>86.96, 89.24, 93.50, 105.3</td>
<td>87.06, 89.24, 93.50, 105.2</td>
<td>87.65, 89.24, 93.50, 105.3</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>93.50, 92.7, 105.3</td>
<td>93.50, 92.7, 105.3</td>
<td>93.50, 92.7, 105.3</td>
<td>93.50, 92.7, 105.3</td>
</tr>
<tr>
<td>a, b, γ (°)</td>
<td>93.50, 92.7</td>
<td>93.50, 92.7</td>
<td>93.50, 92.7</td>
<td>93.50, 92.7</td>
</tr>
<tr>
<td>resolution limits</td>
<td>50-3.30, 3.30</td>
<td>50-3.30, 3.30</td>
<td>50-3.30, 3.30</td>
<td>50-3.30, 3.30</td>
</tr>
<tr>
<td>R-meas (%)^b</td>
<td>12.5 (100.4)</td>
<td>8.9 (31.3)</td>
<td>7.1 (72.6)</td>
<td>9.8 (75.2)</td>
</tr>
<tr>
<td>CC^1/2</td>
<td>99.9 (87.1)</td>
<td>99.6 (94.9)</td>
<td>99.9 (77.6)</td>
<td>99.8 (77.0)</td>
</tr>
<tr>
<td>no. unique reflections</td>
<td>71,266 (5,052)</td>
<td>34,399 (2,510)</td>
<td>116,568 (8,510)</td>
<td>68,928 (4,733)</td>
</tr>
<tr>
<td>mean (I)/σd(I)</td>
<td>20.4 (2.9)</td>
<td>16.7 (4.8)</td>
<td>14.9 (2.0)</td>
<td>13.6 (2.0)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>99.3 (95.0)</td>
<td>98.6 (95.7)</td>
<td>97.8 (96.9)</td>
<td>96.4 (89.5)</td>
</tr>
<tr>
<td>multiplicity</td>
<td>13.7 (12.0)</td>
<td>3.6 (3.4)</td>
<td>3.6 (3.4)</td>
<td>3.6 (3.4)</td>
</tr>
<tr>
<td>B-wilson (Å^2)</td>
<td>74.5</td>
<td>35.6</td>
<td>44.9</td>
<td>71.8</td>
</tr>
</tbody>
</table>

**Table 1:** Data collection statistics 1 (used for structure solution)

<table>
<thead>
<tr>
<th>DATA COLLECTION</th>
<th>crystal 5</th>
<th>crystal 6</th>
<th>crystal 7</th>
<th>crystal 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>wavelength (Å)</td>
<td>1.25472</td>
<td>1.25511</td>
<td>1.25473</td>
<td>1.25511</td>
</tr>
<tr>
<td>absorption position</td>
<td>peak</td>
<td>inflection</td>
<td>peak</td>
<td>inflection</td>
</tr>
<tr>
<td>space group</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>cell dimensions</td>
<td>87.65, 89.52, 93.51, 105.6</td>
<td>87.65, 89.51, 93.50, 105.6</td>
<td>87.51, 93.50, 105.6</td>
<td>87.51, 93.50, 105.6</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>93.51, 92.8, 105.6</td>
<td>93.50, 92.8, 105.6</td>
<td>93.04, 92.7, 105.6</td>
<td>93.04, 92.7, 105.6</td>
</tr>
<tr>
<td>a, b, γ (°)</td>
<td>93.51, 92.8</td>
<td>93.50, 92.8</td>
<td>92.7, 105.6</td>
<td>92.7, 105.6</td>
</tr>
<tr>
<td>resolution limits</td>
<td>50-3.10, 3.10</td>
<td>50-3.10 (3.38-3.77)</td>
<td>50-3.68 (3.77-3.68)</td>
<td>50-3.66 (3.74-3.66)</td>
</tr>
<tr>
<td>(Å)^a</td>
<td>(3.17-3.10)</td>
<td>(3.17-3.10)</td>
<td>(3.17-3.10)</td>
<td>(3.17-3.10)</td>
</tr>
<tr>
<td>R-meas (%)^b</td>
<td>12.4 (96.3)</td>
<td>10.3 (72.9)</td>
<td>15.1 (122.9)</td>
<td>11.1 (71.6)</td>
</tr>
<tr>
<td>CC^1/2</td>
<td>99.8 (74.4)</td>
<td>99.8 (74.2)</td>
<td>99.7 (51.8)</td>
<td>99.8 (74.4)</td>
</tr>
<tr>
<td>no. unique reflections</td>
<td>83,616 (5,741)</td>
<td>82,956 (5,741)</td>
<td>49,368 (2,970)</td>
<td>50,039 (3,175)</td>
</tr>
<tr>
<td>mean (I)/σd(I)</td>
<td>13.8 (2.1)</td>
<td>11.8 (2.0)</td>
<td>11.1 (1.1)</td>
<td>12.2 (2.1)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>96.8 (89.5)</td>
<td>96.4 (90.3)</td>
<td>95.8 (77.3)</td>
<td>95.5 (82.8)</td>
</tr>
<tr>
<td>multiplicity</td>
<td>7.1 (6.3)</td>
<td>3.6 (3.3)</td>
<td>6.6 (2.8)</td>
<td>3.5 (3.2)</td>
</tr>
<tr>
<td>B-wilson (Å^2)</td>
<td>69.1</td>
<td>67.5</td>
<td>93.6</td>
<td>84.0</td>
</tr>
</tbody>
</table>

**Table 2:** Data collection statistics 2 (Ta-LIII absorption edge)
<table>
<thead>
<tr>
<th>DATA COLLECTION</th>
<th>crystal 9</th>
<th>crystal 9</th>
<th>crystal 9</th>
<th>crystal 9</th>
<th>crystal 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>wavelength (Å)</td>
<td>1.24231</td>
<td>1.24233</td>
<td>1.25476</td>
<td>1.25516</td>
<td>1.26759</td>
</tr>
<tr>
<td>absorption position</td>
<td>High remote</td>
<td>High remote chi-40</td>
<td>peak</td>
<td>inflection</td>
<td>Low remote</td>
</tr>
<tr>
<td>space group</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>cell dimensions</td>
<td>a, b, c (Å)</td>
<td>87.66, 89.77, 94.18, 105.7, 91.8</td>
<td>87.75, 89.81, 94.26, 105.1, 91.8</td>
<td>87.83, 89.90, 94.29, 105.1, 91.8</td>
<td>87.75, 89.81, 94.26, 105.1, 91.8</td>
</tr>
<tr>
<td>R-meas (%)b</td>
<td>7.9 (56.4)</td>
<td>8.8 (176.7)</td>
<td>9.3 (78.5)</td>
<td>8.9 (132.6)</td>
<td>7.7 (112.5)</td>
</tr>
<tr>
<td>CC 1/2</td>
<td>99.7 (75.6)</td>
<td>99.7 (66.3)</td>
<td>99.7 (68.6)</td>
<td>99.7 (68.8)</td>
<td>99.8 (71.8)</td>
</tr>
<tr>
<td>Nr. unique reflections</td>
<td>68,621</td>
<td>59,705</td>
<td>57,630</td>
<td>61,192</td>
<td>51,574</td>
</tr>
<tr>
<td>mean (I)/σd(I)</td>
<td>12.8 (2.0)</td>
<td>12.1 (1.2)</td>
<td>10.6 (1.7)</td>
<td>11.7 (1.3)</td>
<td>13.9 (1.3)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>86.9 (90.2)</td>
<td>82.8 (85.8)</td>
<td>83.4 (90.2)</td>
<td>84.7 (90.0)</td>
<td>81.5 (90.3)</td>
</tr>
<tr>
<td>multiplicity</td>
<td>1.8 (1.8)</td>
<td>1.8 (1.9)</td>
<td>1.8 (1.9)</td>
<td>1.8 (1.9)</td>
<td>1.8 (1.9)</td>
</tr>
<tr>
<td>B-wilson (Å²)</td>
<td>47.9</td>
<td>39.1</td>
<td>31.7</td>
<td>42.2</td>
<td>21.7</td>
</tr>
</tbody>
</table>

Table 3: Data collection statistics 3 (Ta-LIII absorption edge)
## DATA COLLECTION

<table>
<thead>
<tr>
<th></th>
<th>crystal 10</th>
<th>crystal 10</th>
<th>crystal 10</th>
<th>crystal 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>completeness (%)</td>
<td>93.2 (84.8)</td>
<td>93.7 (82.3)</td>
<td>93.7 (81.9)</td>
<td>92.7 (83.4)</td>
</tr>
<tr>
<td>multiplicity</td>
<td>1.8 (1.9)</td>
<td>1.8 (1.9)</td>
<td>1.8 (1.9)</td>
<td>1.8 (1.7)</td>
</tr>
<tr>
<td>B-wilson (Å²)</td>
<td>91.1</td>
<td>92.6</td>
<td>95.1</td>
<td>101.5</td>
</tr>
</tbody>
</table>

*Table 4: Data collection statistics 4 (Ta-LII absorption edge)*

<table>
<thead>
<tr>
<th>Site</th>
<th>Occupancy</th>
<th>B-factor (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 1</td>
<td>0.21</td>
<td>162.2</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 2</td>
<td>0.20</td>
<td>90.1</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 3</td>
<td>0.36</td>
<td>33.6</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 4</td>
<td>0.18</td>
<td>143.2</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 5</td>
<td>0.29</td>
<td>187.3</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 6</td>
<td>0.12</td>
<td>190.3</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 7</td>
<td>0.39</td>
<td>152.1</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 8</td>
<td>0.20</td>
<td>138.2</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 9</td>
<td>0.10</td>
<td>185.7</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 10</td>
<td>0.05</td>
<td>64.1</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 11</td>
<td>0.28</td>
<td>109.1</td>
</tr>
<tr>
<td>Spherical Cluster Ta₆Br₁₂:Ta 12</td>
<td>0.20</td>
<td>114.3</td>
</tr>
</tbody>
</table>

*Table 5: Refinement parameters of atomic substructure*
### Table 6: Lack of closure for the reference native dataset determined by SIRAS protocol divided by total variance for crystal 2

<table>
<thead>
<tr>
<th>BIN</th>
<th>$D_{\text{min}}$</th>
<th>$D_{\text{max}}$</th>
<th>$N_{\text{acen}}$</th>
<th>$N_{\text{cen}}$</th>
<th>LOC$_{\text{acen}}^c$</th>
<th>LOC$_{\text{cen}}^c$</th>
<th>$N_{\text{acen}}$</th>
<th>$N_{\text{cen}}$</th>
<th>LOC$_{\text{acen}}^c$</th>
<th>LOC$_{\text{cen}}^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48.79</td>
<td>14.16</td>
<td>436</td>
<td>0</td>
<td>2.966</td>
<td>0.000</td>
<td>431</td>
<td>0</td>
<td>1.605</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>14.16</td>
<td>10.23</td>
<td>735</td>
<td>0</td>
<td>1.247</td>
<td>0.000</td>
<td>735</td>
<td>0</td>
<td>1.101</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>10.23</td>
<td>8.41</td>
<td>935</td>
<td>0</td>
<td>1.091</td>
<td>0.000</td>
<td>927</td>
<td>0</td>
<td>1.039</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>8.41</td>
<td>7.31</td>
<td>1111</td>
<td>0</td>
<td>0.989</td>
<td>0.000</td>
<td>1102</td>
<td>0</td>
<td>1.020</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>7.31</td>
<td>6.56</td>
<td>1253</td>
<td>0</td>
<td>0.966</td>
<td>0.000</td>
<td>1248</td>
<td>0</td>
<td>1.102</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>6.56</td>
<td>5.99</td>
<td>1387</td>
<td>0</td>
<td>0.985</td>
<td>0.000</td>
<td>1382</td>
<td>0</td>
<td>1.011</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>5.99</td>
<td>5.55</td>
<td>1499</td>
<td>0</td>
<td>1.020</td>
<td>0.000</td>
<td>1489</td>
<td>0</td>
<td>0.994</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>5.55</td>
<td>5.20</td>
<td>1608</td>
<td>0</td>
<td>1.037</td>
<td>0.000</td>
<td>1604</td>
<td>0</td>
<td>0.981</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>5.20</td>
<td>4.91</td>
<td>1717</td>
<td>0</td>
<td>1.011</td>
<td>0.000</td>
<td>1712</td>
<td>0</td>
<td>0.999</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>4.91</td>
<td>4.66</td>
<td>1818</td>
<td>0</td>
<td>1.005</td>
<td>0.000</td>
<td>1813</td>
<td>0</td>
<td>1.074</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>4.66</td>
<td>4.44</td>
<td>1912</td>
<td>0</td>
<td>1.041</td>
<td>0.000</td>
<td>1906</td>
<td>0</td>
<td>1.100</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>4.44</td>
<td>4.25</td>
<td>1986</td>
<td>0</td>
<td>1.050</td>
<td>0.000</td>
<td>1984</td>
<td>0</td>
<td>1.200</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>4.25</td>
<td>4.09</td>
<td>2078</td>
<td>0</td>
<td>1.014</td>
<td>0.000</td>
<td>2076</td>
<td>0</td>
<td>1.182</td>
<td>0.000</td>
</tr>
<tr>
<td>14</td>
<td>4.09</td>
<td>3.94</td>
<td>2150</td>
<td>0</td>
<td>0.993</td>
<td>0.000</td>
<td>2148</td>
<td>0</td>
<td>1.119</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>3.94</td>
<td>3.81</td>
<td>2255</td>
<td>0</td>
<td>1.015</td>
<td>0.000</td>
<td>2252</td>
<td>0</td>
<td>1.085</td>
<td>0.000</td>
</tr>
<tr>
<td>16</td>
<td>3.81</td>
<td>3.69</td>
<td>2288</td>
<td>0</td>
<td>0.966</td>
<td>0.000</td>
<td>2287</td>
<td>0</td>
<td>1.018</td>
<td>0.000</td>
</tr>
<tr>
<td>17</td>
<td>3.69</td>
<td>3.58</td>
<td>2384</td>
<td>0</td>
<td>0.985</td>
<td>0.000</td>
<td>2382</td>
<td>0</td>
<td>0.975</td>
<td>0.000</td>
</tr>
<tr>
<td>18</td>
<td>3.58</td>
<td>3.48</td>
<td>2464</td>
<td>0</td>
<td>0.898</td>
<td>0.000</td>
<td>2462</td>
<td>0</td>
<td>0.881</td>
<td>0.000</td>
</tr>
<tr>
<td>19</td>
<td>3.48</td>
<td>3.39</td>
<td>2528</td>
<td>0</td>
<td>0.864</td>
<td>0.000</td>
<td>2528</td>
<td>0</td>
<td>0.849</td>
<td>0.000</td>
</tr>
<tr>
<td>20</td>
<td>3.39</td>
<td>3.30</td>
<td>2556</td>
<td>0</td>
<td>0.807</td>
<td>0.000</td>
<td>2553</td>
<td>0</td>
<td>0.811</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall</td>
<td>35100</td>
<td>0</td>
<td>1.032</td>
<td>0.000</td>
<td></td>
<td></td>
<td>35021</td>
<td>0</td>
<td>1.033</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Table 7: Lack of closure (LOC) for the derivative crystal determined by SIRAS protocol for crystal 1

<table>
<thead>
<tr>
<th>BIN</th>
<th>$D_{\text{min}}$</th>
<th>$D_{\text{max}}$</th>
<th>$N_{\text{acen}}$</th>
<th>$N_{\text{cen}}$</th>
<th>PP$_{\text{acen}}$</th>
<th>RC$_{\text{acen}}^d$</th>
<th>PP/RC$_{\text{acen}}$</th>
<th>$N_{\text{acen}}$</th>
<th>PP$_{\text{acen}}^f$</th>
<th>RC$_{\text{acen}}^e$</th>
<th>PP/RC$_{\text{acen}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48.79</td>
<td>14.16</td>
<td>430</td>
<td>1.183</td>
<td>1.183</td>
<td>0.608</td>
<td>0.000</td>
<td>431</td>
<td>5.116</td>
<td>0.292</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>14.16</td>
<td>10.23</td>
<td>733</td>
<td>1.949</td>
<td>1.949</td>
<td>0.523</td>
<td>0.000</td>
<td>735</td>
<td>6.759</td>
<td>0.232</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>10.23</td>
<td>8.41</td>
<td>925</td>
<td>1.723</td>
<td>1.723</td>
<td>0.566</td>
<td>0.000</td>
<td>927</td>
<td>5.747</td>
<td>0.269</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>8.41</td>
<td>7.31</td>
<td>1092</td>
<td>1.600</td>
<td>1.600</td>
<td>0.662</td>
<td>0.000</td>
<td>1102</td>
<td>4.909</td>
<td>0.306</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>7.31</td>
<td>6.56</td>
<td>1243</td>
<td>1.278</td>
<td>1.278</td>
<td>0.759</td>
<td>0.000</td>
<td>1248</td>
<td>3.177</td>
<td>0.412</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>6.56</td>
<td>5.99</td>
<td>1372</td>
<td>0.930</td>
<td>0.930</td>
<td>0.849</td>
<td>0.000</td>
<td>1382</td>
<td>2.999</td>
<td>0.432</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>PP</td>
<td>RC</td>
<td>PHO</td>
<td>PPH</td>
<td>C</td>
<td>PC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5.99</td>
<td>5.55</td>
<td>1488</td>
<td>0.656</td>
<td>0.892</td>
<td>0.000</td>
<td>1489</td>
<td>2.291</td>
<td>0.518</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5.55</td>
<td>5.20</td>
<td>1597</td>
<td>0.411</td>
<td>0.932</td>
<td>0.000</td>
<td>1604</td>
<td>1.651</td>
<td>0.665</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5.20</td>
<td>4.91</td>
<td>1702</td>
<td>0.249</td>
<td>0.969</td>
<td>0.000</td>
<td>1712</td>
<td>1.089</td>
<td>0.787</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4.91</td>
<td>4.66</td>
<td>1803</td>
<td>0.155</td>
<td>0.991</td>
<td>0.000</td>
<td>1813</td>
<td>0.623</td>
<td>0.897</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4.66</td>
<td>4.44</td>
<td>1893</td>
<td>0.089</td>
<td>0.978</td>
<td>0.000</td>
<td>1906</td>
<td>0.390</td>
<td>0.956</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4.44</td>
<td>4.25</td>
<td>1963</td>
<td>0.044</td>
<td>0.980</td>
<td>0.000</td>
<td>1984</td>
<td>0.163</td>
<td>0.988</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4.25</td>
<td>4.09</td>
<td>2049</td>
<td>0.013</td>
<td>0.969</td>
<td>0.000</td>
<td>2076</td>
<td>0.043</td>
<td>0.999</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4.09</td>
<td>3.94</td>
<td>2133</td>
<td>0.022</td>
<td>0.971</td>
<td>0.000</td>
<td>2148</td>
<td>0.058</td>
<td>0.999</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3.94</td>
<td>3.81</td>
<td>2223</td>
<td>0.045</td>
<td>0.959</td>
<td>0.000</td>
<td>2252</td>
<td>0.099</td>
<td>0.997</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3.81</td>
<td>3.69</td>
<td>2264</td>
<td>0.068</td>
<td>0.954</td>
<td>0.000</td>
<td>2287</td>
<td>0.122</td>
<td>0.996</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>3.69</td>
<td>3.58</td>
<td>2348</td>
<td>0.082</td>
<td>0.936</td>
<td>0.000</td>
<td>2382</td>
<td>0.135</td>
<td>0.996</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3.58</td>
<td>3.48</td>
<td>2436</td>
<td>0.103</td>
<td>0.940</td>
<td>0.000</td>
<td>2462</td>
<td>0.147</td>
<td>0.997</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3.48</td>
<td>3.39</td>
<td>2488</td>
<td>0.114</td>
<td>0.929</td>
<td>0.000</td>
<td>2528</td>
<td>0.144</td>
<td>0.998</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.39</td>
<td>3.30</td>
<td>2517</td>
<td>0.126</td>
<td>0.945</td>
<td>0.000</td>
<td>2553</td>
<td>0.143</td>
<td>0.998</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>34699</td>
<td>0.693</td>
<td>0.834</td>
<td>0.000</td>
<td>35021</td>
<td>1.402</td>
<td>0.744</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 Phasing power (PP) and R-Cullis (RC) for the derivative crystal determined by SIRAS protocol for crystal 1

---

*Appendix*

7 5.99 5.55 1488 0.656 0.892 0.000 1489 2.291 0.518 0.000
8 5.55 5.20 1597 0.411 0.932 0.000 1604 1.651 0.665 0.000
9 5.20 4.91 1702 0.249 0.969 0.000 1712 1.089 0.787 0.000
10 4.91 4.66 1803 0.155 0.991 0.000 1813 0.623 0.897 0.000
11 4.66 4.44 1893 0.089 0.978 0.000 1906 0.390 0.956 0.000
12 4.44 4.25 1963 0.044 0.980 0.000 1984 0.163 0.988 0.000
13 4.25 4.09 2049 0.013 0.969 0.000 2076 0.043 0.999 0.000
14 4.09 3.94 2133 0.022 0.971 0.000 2148 0.058 0.999 0.000
15 3.94 3.81 2223 0.045 0.959 0.000 2252 0.099 0.997 0.000
16 3.81 3.69 2264 0.068 0.954 0.000 2287 0.122 0.996 0.000
17 3.69 3.58 2348 0.082 0.936 0.000 2382 0.135 0.996 0.000
18 3.58 3.48 2436 0.103 0.940 0.000 2462 0.147 0.997 0.000
19 3.48 3.39 2488 0.114 0.929 0.000 2528 0.144 0.998 0.000
20 3.39 3.30 2517 0.126 0.945 0.000 2553 0.143 0.998 0.000

Overall 34699 0.693 0.834 0.000 35021 1.402 0.744 0.000

---

\[ R_{\text{meas}} = \frac{\sum_{hkl} |n_{hkl} I_i(hkl) - n_{hkl} I_i(\text{obs})|}{\sum_{hkl} n_{hkl} I_i(hkl)}, \text{with } I_i(hkl) = \sum_{n_{hkl}} l_i(hkl) \text{ and } I_i(hkl) \text{ the intensity of a unique reflection.} \]

\[ L O C = \sum_{hkl} |F_P(\text{obs}) + F_H(\text{calc}) - F_{PH}(\text{obs})| \]

\[ R_{\text{Cullis \text{ iso}}} = \frac{\sum_{hkl} P(\phi) |L O C|}{\sum_{hkl} |F_{PH}(\text{obs}) - F_P(\text{obs})|} \]

\[ R_{\text{Cullis \text{ ano}}} = \frac{\sum_{hkl} P(\phi) |\Delta \text{ano (obs)} - \Delta \text{ano (calc)}|}{\sum_{hkl} |\Delta \text{ano (obs)}|} \]

\[ P P = \frac{|F_{PH}|}{\sigma(\sum_{hkl} P(\phi) |L O C|)} \]
6.2 Publications
Anionic glycopolymers known as wall teichoic acids (WTAs) functionalize the peptidoglycan layers of many Gram-positive bacteria. WTAs play central roles in many fundamental aspects of bacterial physiology, and they are important determinants of pathogenesis and antibiotic resistance. A number of enzymes that glycosylate WTA in *Staphylococcus aureus* have recently been identified. Among these is the glycosyltransferase TarM, a component of the WTA *de novo* biosynthesis pathway. TarM performs the synthesis of α-O-N-acetylglycosylated poly-5′-phosphoribitol in the WTA structure. We have solved the crystal structure of TarM at 2.4 Å resolution, and we have also determined a structure of the enzyme in complex with its substrate UDP-GlcNAc at 2.8 Å resolution. The protein assembles into a propeller-like homotrimer in which each blade contains a GT-B-type glycosyltransferase domain with a typical Rossmann fold. The enzymatic reaction retains the stereochemistry of the transferred GlcNAc moiety on the polyribitol backbone. TarM assembles into a trimer using a novel trimerization domain, here termed the HUB domain. Structure-guided mutagenesis experiments of TarM identify residues critical for enzyme activity, assign a putative role for the HUB in TarM function, and allow us to propose a likely reaction mechanism.

*Staphylococcus aureus* is a leading cause of nosocomial pneumonia, surgical site infections, and blood stream infections.

---

**Background:** TarM catalyzes the addition of α-GlcNAc to 4′-polyribitol-phosphate of wall teichoic acid (WTA) in *S. aureus*.

**Results:** Structural analysis shows that TarM is a homotrimeric propeller-like glycosyltransferase.

**Conclusion:** Enzyme processivity is linked to a novel domain that generates the trimer.

**Significance:** Our structure-function analysis helps define the biosynthetic pathway leading to WTA glycosylation in *S. aureus*.

The bacterium remains a severe threat to human health, in part due to the continued emergence of strains that are resistant to existing antibiotics (1). To survive, *S. aureus* relies heavily on virulence and adaptability to its environment. The *S. aureus* cell envelope structure is highly complex, and this complexity is central to the survival and adaptability of the organism. Major components of the cell envelope are glycosylated structures (2, 3), including glycoproteins, polysaccharide intracellular adhesin, capsular polysaccharides, peptidoglycan, lipoteichoic acid, and wall teichoic acid (WTA) (4). The unique ability of methicillin-resistant *S. aureus* (MRSA) to develop resistance to β-lactams as well as other antibiotics (4) is in part due to the structure and composition of specific cell wall components (5–7). The role of WTAs in these processes is complex and not well understood at the molecular level. WTAs serve to protect the cell from degradation through lysozyme (8) or from the action of cationic antimicrobial oligopeptides (9). However, WTAs also assist in staphylococcal adhesion and colonization (10, 11). Furthermore, they play a critical role in cell division and biofilm formation (12). The chemical structure of WTA varies substantially among Gram-positive bacteria (13), and this variability represents one strategy that allows these organisms to adapt to the environment or react to host defense systems (3).

Most of the *S. aureus* strains produce poly-ribitol-phosphate (RboP)-type WTA, which is composed of ~40 RboP units that are connected by 1,5-phosphodiester bonds. Some of the C4 hydroxyl groups of the WTA RboP unit are either substituted with α-O- or β-O-GlcNAc, whereas the C2 hydroxyls sometimes carry a D-alanine (Fig. 1).

Biosynthesis of WTA in *S. aureus* is carried out by a cluster of enzymes belonging to the teichoic acid ribitol (Tar) synthesizing pathway, many of which have only been recently character-

---

© 2015 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
ized. The polyribitol backbone is covalently attached to the N-acetylmuramic acid moiety of the peptidoglycan via a disaccharide (ManNAc→β(1,4)-GlcNAc→1-P) linkage unit followed by two units of glycerol phosphates (14) as shown in Fig. 1. In concert with TarA, TarB, TarI, TarJ, and TarL, the main chain is synthesized on the lipid carrier undecaprenyl monophosphate (C55P), which is embedded in the inner leaflet of the cell membrane. After the completion of glycosylation, the main chain of WTA is flipped to the outer leaflet of the plasma membrane via the ABC-type transporter TarG/TarH (13, 15).

The regulated addition of alanines at the ribitol 2′-position by the d-alanyltansferase, one gene product of the dltABCD (16) gene cluster, as a final modification counterbalances the predominant negative charge of the linkage phosphate groups and results in WTA becoming zwitterionic. The evolvement of host-pathogen interaction is thought to have led to the increase of positive charges in the bacterial cell wall to circumvent the action of cationic antimicrobial peptides (17, 18).

The enzymes TarM and TarS decorate the WTA backbone with α-GlcNAc and β-GlcNAc, respectively (6, 19). The β-GlcNAcylation of RboP is critical for the resistance of S. aureus (MRSA) to β-lactams (6). Furthermore β-GlcNAc residues on WTA are recognized by the mannose-binding lectin, leading to complement activation pathway of the human innate immune system as well as by antibodies in the adaptive immune system (20). The role of the α-GlcNAcylation is not yet known.

To define the mechanism of RboP glycosylation, we have performed a structure-function analysis of the glycosyltransferase TarM, a 171-kDa protein. Sequence analysis and database research predicted one domain of TarM to belong to the GT-B superfamily of glycosyltransferases (21), whereas the second domain was assigned DUF1975 (domain of unknown function) according to the Pfam database (22). The crystal structure of TarM reveals a propeller-like trimer, with the three GT-B domains arranged as blades around a central hub formed by the three DUF1975 domains. Accordingly, we suggest the name HUB for DUF1975. The structure analysis of TarM bound to its substrate UDP-GlcNAc identifies the active site, defines essential contacts with this ligand, and suggests a plausible reaction mechanism. As TarM is the first known enzyme structure in the biogenesis pathway of poly-RboP WTA, our work sheds light on an essential aspect of S. aureus glycosylation and provides an initial framework for investigating parameters that dictate glycosylation of WTA in bacteria.

**EXPERIMENTAL PROCEDURES**

**Strains and Media—**S. aureus strains were cultured in BM media (1% (w/v) Tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.1% (w/v) glucose, 0.1% (w/v) K2HPO4, pH 7.2). *Escherichia coli* strains were cultivated in LB media (1% (w/v) casein hydrolysate peptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1% (w/v) glucose, 1% (w/v) K2HPO4, pH 7.2).

**Cloning and Expression of tarM and Mutant tarM—**Wild-type *tarM* (SACOL 1043) was subcloned as reported previously (19). QuikChange (Stratagene) was used to introduce point mutations into the glycosyltransferase active site in either pRB474-tarM or pBAD-TOPO-102/202-tarM (EcoRI/BamHI, Amp or Kan) as template. pRB474 shuttle vectors containing *tarM* variants (wt or mutant) were transformed into *S. aureus* RN4220 mutant ΔtarMΔtarS for determining the efficiency of plaquing (EOP). Thus *tarM* and *tarM* mutants were fused to a hexahistidine tag at the N terminus and subcloned into the pBAD vector for recombinant expression in *E. coli* strain Top10 (Table 1).

**Expression—**Single colonies of *E. coli* transformants containing *tarM* variants were grown on antibiotics containing LB agar (1.5% (w/v) agar-agar in LB-medium). They were inoculated into 2 ml of LB medium and grown overnight at 37 °C. For large scale protein production, bacterial culture was induced at the log phase (*A560* = 0.5–1.0) with l-arabinose at a final concentration of 0.001% (w/v) at 20 °C for 12–20 h before harvesting by centrifugation at 7,900 × g for 13 min. After washing once with buffer A (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA), the cells were resuspended with buffer B (100 mM triethanolamine, pH 8.5, 500 mM LiCl, 5 mM EDTA, 1 mM DTT) for storage at −80 °C or for purification as described below.

**Purification—**Cells were lysed by ultrasonication (Digital Sonifer, Branson). After centrifugation at 38,000 × g for 55 min, the supernatant containing recombinant TarM was collected and used as the crude TarM preparation after dialyzing against...
buffer C (50 mM triethanolamine, pH 8.5, 500 mM LiCl, 25 mM imidazole, 1 mM DTT). The crude preparation was applied onto a His-Trap-FF nickel-chelate affinity column (GE Healthcare, 5 ml), and the column was subsequently washed with 30 column volumes of buffer C followed by another washing step with 10 column volumes of buffer C containing 10 mM imidazole. Pure TarM was eluted using a gradient ranging from 36 to 400 column volumes of buffer C in buffer C. The pure sample was then concentrated to 1 mg/ml (Sartorius vivaspin 20 PES, 50,000 molecular weight cut-off), dialyzed against buffer D (50 mM triethanolamine, pH 8.5, 250 mM LiCl, 5 mM EDTA, 1 mM DTT), and subjected to 2 successive treatments with enterokinase (Ekmax, 20 °C, 15–20 h, 1.5 units/mg; Life Technologies). Aggregated proteins and excess enterokinase were removed by gel filtration (Superdex 200). A final concentration step yielded highly monodisperse pure protein that was then used for crystallization and biophysical characterization. Purity and homogeneity of TarM were assessed by SDS-PAGE as well as dynamic light scattering (DLS).

**Structure of TarM**

Initial small crystals grew as trap-ezoïd-shaped plates (diameter ~ 50 μm) in 600-nl drops containing 300 nl of TarM protein solution in buffer D and 300 nl of crystallizing buffer E (100 mM imidazole, pH 8.0, 200 mM Ca(OAc)₂, 20% (w/v) PEG-1000) in 96-well plates using the sitting-drop vapor diffusion method. Refinement yielded the final crystallization condition (100 mM imidazole, pH 7.7–8.2, 18–21% (w/v) PEG-1000, 0.1–0.25 mM Ca(OAc)₂) which produced large, single crystals (diameter ~ 250 μm). After soaking crystals in cryoprotection solution (buffer D:buffer E 1:1, 10% (v/v) 4s-2-methyl-2,4-pentanediol) they were directly frozen in liquid nitrogen. Data were collected on a PILATUS 2 M detector at beamline X06DA of the Swiss Light Source. Data processing using XDS (23) yielded the same space group as the crystals of unbound TarM, with slightly altered cell parameters of a = b = 122 Å and c = 212 Å. The unbiased whole native structure solution was used as a molecular replacement (30, 32) input model for phasing the new data. After one refinement run of the phased structure model, the UDP-GlcNAc moiety was clearly visible in the unbiased electron density maps, and thus the ligand was incorporated into the model using the refmac library (33) in COOT (28). TLS refinement utilizing REFMAC5 and PHENIX yielded the final model for the binary complex (PDB ID 4WAD). Data collection and refinement statistics are given in Table 2. Figures were generated with PyMOL (34).

**Circular Dichroism**—CD measurements were performed on a JASCO J-720 spectropolarimeter with purified TarM samples at 1 mg/ml in buffer D or buffer G (10 mM Na₂HPO₄, 0.01 mM NaH₂PO₄, pH 8.5, 200 mM NaF). Data were recorded and evaluated using Zeta Software (Malvern).

**Plaques Efficiency of φ11**—To analyze the in vivo activity of TarM and its variants, the plaquing efficiency of bacteriophage φ11 was determined by plating φ11 on S. aureus mutant strain RN4220ΔtarMΔtarS complemented with empty plasmid (pRB474), a plasmid encoding wt TarM (pRB474-tarM), or plasmids encoding TarM variants (see Table 1). To determine the plaquing efficiency, 100 μl of φ11 lysate with ~1000 plaque-forming units (pfu) was mixed well with 100 μl of bacteria culture containing ~4 × 10⁷ colony forming units. After incubation at 25 °C for 10 min, the infection mixture was mixed well with 5 ml of soft agar and then poured onto BM plates containing 10 μg/ml chloramphenicol. The plates stood at 37 °C overnight (16–24 h) and following up, the pfu was enumerated. The plaquing efficiency of φ11 on tarM-complemented RN4220-ΔtarMΔtarS was set to 100%.

**WTA Glycosyltransferase Activity Assay**—The colorimetric assay was prepared according to Mulder’s procedure (35) with slight modifications. 1.5 μg of recombinant TarM variants (in 20 mM Tris, pH 8.0, 10 mM MgCl₂) were incubated with UDP-GlcNAc (2 mM) and non-glycosylated WTA (25 μM) that was isolated from RN4220 ΔΔtarS (36) and a reaction mixture consisting of phosphonoctylpyruvate and NADH (0.2 mM each). The release of UDP by TarM was
assayed through the coupled conversion of NADH to NAD⁺ (340 nm, 40 min, 25 °C) by pyruvate kinase and lactate dehydrogenase (2 units each), leading to the decrease of absorbance.

**RESULTS**

**Overall Structure and Domain Organization of TarM—**
TarM assembles into a symmetric, propeller-like homotrimer, with three blades projecting from the central hub (Fig. 2, A and...
B). The three blades project at angles of ~120 degrees from the hub (Fig. 2A), giving the propeller a cradle-like appearance, with a large cavity at its center. Each TarM monomer can be divided into two regions (Fig. 2, C and D); the glycosyltransferase (GT) domain forms the blade, which can be further subdivided into an N-terminal domain (Gt-N, residues 1–80, 202–309) and a C-terminal domain (Gt-C, residues 310–493). The trimer is assembled by three copies of a domain (residues 81–201) that was originally annotated as a domain of unknown function (DUF1975) and that is inserted into Gt-N. This domain features a 10-stranded antiparallel β-sheet composed of strands β4 through β13, with one face of the sheet covered by a single α-helix (α4). Given its function in TarM trimerization, we refer to this domain as the HUB domain. The Gt-N and HUB domains are well ordered and exhibit low overall temperature factors (B-factors). In contrast, large portions of Gt-C display higher mobility and elevated B-factors, probably as a result of the larger surface-exposed area of this domain, the paucity of its interactions in the crystal lattice, and its flexible linkage to Gt-N. The Gt-N domain is positioned atop the HUB, whereas Gt-C projects away from this assembly at an angle of ~40 degrees. Each TarM monomer, therefore, has a bent, hook-like conformation, giving rise to the cradle-like structure of the trimer (Fig. 2).

Together, subdomains Gt-N and Gt-C form the glycosyltransferase unit, referred to as GT-B. Gt-N consists of a parallel seven-stranded β-sheet (strands β1-β3 and β14-β17) connected by eight α-helices (helices α1-α3, α5-α8 and α16), whereas Gt-C contains a central six-stranded parallel β-sheet (strands β18-β23) and seven flanking α-helices (helices α9-α15) (Fig. 2D). The Gt-N and Gt-C domains are linked by ~10 solvent-exposed residues that connect strands β17 and β18. A DALI (37) query of the Gt-N/Gt-C-unit returned several hits for structural homologs in the GT-B glycosyltransferase superfamily. The glycosyltransferases MshA from Corynebacterium glutamicum (Z-score 33.1, r.m.s.d. 2.9 Å, 334 aligned residues, 18% sequence identity, PDB ID 3C4Q (38)) and BshA from Bacillus anthracis (Z-score 32.9, r.m.s.d. 3.2 Å, 336 aligned residues, sequence identity 18%, PDB 3MBO (39)) yielded the highest scores, directly followed by the streptococcal enzyme GtfA (Z-score 32.8, r.m.s.d. 2.8 Å, 361 aligned residues, sequence identity 23%, PDB 4PQG (40)). Other homologs...
gous proteins include PimB (Z-score of 30.6, r.m.s.d. of 3.0 Å, 331 aligned residues, 14% sequence identity, PDB 3OKA (41)) and chlorovirus NY-2A gene product B736L (Z-score of 28.7, r.m.s.d. of 3.4 Å, 333 aligned residues, 14% sequence identity, PDB 3OY7 (42)). Table 3 lists the conservation of the binding pocket.

The Oligomeric State of TarM—The asymmetric unit of the crystals contains one TarM monomer (58 kDa, 493 residues) that assembles into the trimeric structure shown in Fig. 2 through a crystallographic three-fold symmetry operator. Trimer contacts exclusively involve the HUB domains, which form a funnel-like arrangement that is ~30 Å wide at one end. The other end of the funnel is almost closed as a result of three closely approaching Val-159 side chains (Fig. 3A). Close-up view showing the interface between two monomers. Three crystallographically related lysines (Lys-136) point with their side chains into the cavity. Other interfacing residues of relevance, determined by PISA, are represented as sticks. The TarM trimer interface is composed of 20 residues (amino acids 98, 100–104, 110, 112–119, 136–139, and 159) in one monomer (orange) and 28 residues (amino acids 85, 87, 88, 91, 136, 141, 142, 144, 154, 156–159, 164, 165, 167, 169, 174, 175, 178, 180, 190, 192–194, and 197–199) in the other monomer (blue). Residues contributing to the interface are concentrated in loops β4', β8', and β9–β13.

**Table 3**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3C4Q, MshA</td>
<td>12</td>
<td>Thr-12</td>
<td>Thr-13</td>
<td>Thr-12</td>
<td>Thr-13</td>
<td>Thr-14</td>
<td>Thr-15</td>
<td>Thr-16</td>
<td>Thr-17</td>
<td>Thr-18</td>
<td>Thr-19</td>
<td>Thr-20</td>
<td>Thr-21</td>
<td>Thr-22</td>
<td>Thr-23</td>
<td>Thr-24</td>
<td>Thr-25</td>
<td>Thr-26</td>
<td>Thr-27</td>
<td>Thr-28</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** Close-up view of the HUB domain-generated trimer interface. A, view along the trimer axis showing three crystallographically related valines (Val-159) closing the HUB on the side opposite to the HUB maw. The color code is the same as in Fig. 2. B, close-up view showing the interface between two monomers. Three crystallographically related lysines (Lys-136) point with their side chains into the cavity. Other interfacing residues of relevance, determined by PISA, are represented as sticks. The TarM trimer interface is composed of 20 residues (amino acids 98, 100–104, 110, 112–119, 136–139, and 159) in one monomer (orange) and 28 residues (amino acids 85, 87, 88, 91, 136, 141, 142, 144, 154, 156–159, 164, 165, 167, 169, 174, 175, 178, 180, 190, 192–194, and 197–199) in the other monomer (blue). Residues contributing to the interface are concentrated in loops β4', β8', and β9–β13.
which evaluates the physiologic relevance of crystallographic interfaces, classifies this interface as significant for complexation (CSS/H11005), in contrast to all other contacts of TarM subunits in the crystals. To examine whether the trimer also exists in solution, we performed size-exclusion chromatography and dynamic light scattering with purified TarM. Both experiments provide evidence for a trimeric state of the enzyme in solution. TarM elutes as a single peak in gel filtration, with a hydodynamic diameter of 14 nm corresponding to a molecular mass of 300 kDa. Although this value is higher than that calculated for the trimer (174 kDa), the protein deviates significantly from a globular shape (Fig. 2A) and would, therefore, be expected to elute at a higher apparent molecular weight. A calculated Perin-factor of 1.5 suggests a molecular shape deviant from a spherical protein (44). In accordance with this, the molecular shape derived from the structural data is reminiscent of an oblate rather than a sphere.

Architecture of the Active Site—Glycosyltransferases of the GT-B class typically bind their substrates at the interface between Gt-N and Gt-C. To characterize the ligand binding site of TarM, we solved the structure of the enzyme bound to its substrate UDP-GlcNAc through incubation of soluble TarM with UDP-GlcNAc and subsequent cocrystallization of the complex. The overall structures of unbound and UDP-GlcNAc-bound TarM are highly similar (r.m.s.d. value of 0.81 Å for 493 aligned residues, Fig. 4A), and thus binding of UDP-GlcNAc does not lead to any larger structural rearrangements. Unambiguous electron density in the active site cleft allowed us to build the UDP-GlcNAc substrate and assign contacts (Fig. 4B). UDP-GlcNAc is located in a cleft formed by five loops (loops 1–5, Fig. 4C).

The GlcNAc moiety rests in a shallow pocket formed by residues in loops 1, 2, and 5, with one face of the sugar ring buried and the other exposed to solvent. The N-acetyl group faces into a small, hydrophobic pocket formed by Met-18 (loop 1), His-249 (loop 2), and Leu-407 (loop 5). The carbonyl oxygen as well as the C3 and C4 hydroxyl groups form hydrogen bonds to loop residues, whereas the anomeric center carbon C1 lies in close proximity to the carboxylate function of Glu-403. The negative charge of the pyrophosphate unit is negated by salt bridges to the Arg-326 and Lys-331 side chains. One of the phosphates also forms contacts with the backbone amide of Gly-17, whereas the other is hydrogen-bonded to Ser-408. Finally, the uridyl unit lies in a narrow pocket that is lined by residues from loops 1 and 3 and closed at the rear end by loop 4. The ribose faces toward the Glu-411 side chain. Specificity for uracil is generated through several polar and hydrophobic interactions with loop 4 residues Tyr-382, Thr-383, and Pro-386 as well as the side chain of Ile-324 within loop 3.

EOP—EOP was utilized to assess in vivo functionality of several TarM variants. Enzymatic activity of wild-type and mutant TarM was assayed using an established semiquantitative method that is based on a link between glycosylated WTA and bacteriophage adsorption to S. aureus hosts (45, 46). Recently we showed sugar residues on WTAs served as the receptor of siphophage such as φ11. The laboratory strain RN4220 lacks all

FIGURE 4. The ligand-binding site of one TarM chain. A, superposition of free and UDP-GlcNAc-bound structures in ribbon representation. The UDP-GlcNAc-bound TarM structure is shown in red, and unbound TarM is depicted in black. 2F0 − Fc, map at 1.0 σ contour-level of UDP-GlcNAc in the active site of TarM, seen from the face-on side. 
B, representation of the bound sugar surrounded by five prominent and well conserved loops (loop-1 (residues 9–18), loop-2 (residues 248–251), loop-3 (residues 324–332), loop-4 (residues 380–386), and loop-5 (residues 401–406)). Dashed lines indicate putative contacts with distances ranging from 2.2 to 4.5 Å. A distance cutoff of 4.5 Å was used to show hydrogen bonds and salt bridges.
resistance mechanisms; hence, phage plaquing efficiency on strains derived from RN4220 indicates the abundance of GlcNAc residues on WTA, reflecting the \textit{in vivo} activity of a WTA glycosyltransferase (19, 45). We also showed that a double mutant RN4220 \textit{tarM}\textit{tarS}, which lacks both \textit{\alpha}-O- and \textit{\beta}-O-GlcNAc, did not only produce any GlcNAc on WTA but was resistant to \textit{\phi}11 infection (6). In this study we complemented this mutant with various \textit{tarM} variants, and the resulting complemented strains were used as hosts for plating \textit{\phi}11 (Table 1). The efficiency of \textit{\phi}11 plaquing reflects the level of WTA glycosylation in those \textit{tarM} variant complemented strains. Cells expressing wild-type TarM, therefore, show the highest EOP, whereas cells lacking TarM activity do not show any plaque forming capacity. Falsifying concentration effects were ruled out by performing EOP experiments for each mutant at threshold titers from the same freshly prepared phage cultures.

\textbf{Structure-guided Mutagenesis of Active Site Residues—}\To obtain insight into the catalytic mechanism and assess the validity of the observed interactions, several of the amino acids that lie in close proximity to the bound UDP-GlcNAc were mutated, and the enzymatic activities of the mutated proteins were analyzed in each case ("Experimental Procedures"). We specifically generated mutants E403A, K331S, R326S, and H249A, all of which probe interactions with substrate (Table 1). To confirm that the mutated proteins are still folded, each protein was purified and subjected to circular dichroism (CD) spectroscopy experiments and DLS analysis (Fig. 5, \textit{A} and \textit{D}). These data show that all mutants are structurally intact and have secondary structure elements that are indistinguishable from those of WT TarM.

Residue Glu-403 is clearly among the most important residues for catalysis. Its mutation to alanine essentially renders TarM inactive as it was not able to produce almost any observable spots on the bacterial lawn and generated no detectable output in the EOP-measurement (Fig. 5). The Glu-403 carboxyl group is thus essential for catalysis. Likewise, the mutation of Lys-331 to serine diminished all transferase activity in the EOP measurement. This mutation was aimed at removing a contact with the pyrophosphate group of UDP-GlcNAc as well as removing a potentially stabilizing interaction with Glu-403, as Lys-331 lies in close proximity to Glu-403, and the two residues could form a salt bridge during catalysis. Our results show that Lys-331 plays an essential role in substrate binding and/or catalysis. Residue Glu-411, which lies near the ribose of UDP, is also highly conserved. Its mutation to alanine also leads to severely reduced enzymatic activity, probably because the Glu-411 side chain is an integral part of the UDP-GlcNAc-binding site. Mutations of Arg-326 to serine and His-249 to alanine led to 20 and 30\%, respectively, remaining WT activity. This suggests that both residues are important contact points that are, however, not essential for the reaction to proceed.

After the phospholysis reaction, the activated GlcNAc oriented on Glu-403 has its anomic carbon pointed to the gap between Gt-N and Gt-C, where the activated acceptor (polyribitol-phosphate) must be located for the chemical reaction of glycosylation to occur. Unfortunately we lack a structure of TarM bound to WTA fragments, which would shed light on the exact structure of the sugar-transfer transition state.

\textbf{Physiologic Role of the HUB and the Trimer—}\To obtain insight into the putative function of the novel HUB domain, we selected a small number of residues in this domain for site-directed mutagenesis (Table 1). The rationale of these experiments was to subtly alter HUB regions mediating trimerization and to test the impact of these mutants on enzyme turnover efficiency. We generated single amino acid substitutions (K136S, N138Q, N180W) as well as a double (V159Y/C164R) and a triple (V159Y/C164R/K136S) mutation near the trimer interface. We observed a substantial decrease in TarM EOP (Fig. 5A) for K136S as well as the double and the triple mutant. We next selected the triple mutant and tested, alongside the wild type and E403A, its enzyme activity under \textit{in vitro} conditions. Although E403A substantially decreased the EOP outcome and the enzymatic activity in the same order (Fig. 5E), indicating fully impaired glycosylation of WTA, we could not observe a comparable outcome for the triple mutant. Thus, the triple mutant produces different results \textit{in vitro} and in the EOP assay. To rationalize this, we hypothesize that the triple mutant may lead to a subtle alteration or destabilization of the TarM trimer structure. Such a subtle change might not affect the catalytic activity of the enzyme in solution, but it might elicit a more severe effect in a physiologic setting. In support of this hypothesis, the CD spectra of the double and the triple mutants (Fig. 5B) show an additional shoulder around 205 nm, indicating a small alteration of secondary structure elements in the HUB domain. According to DLS and size exclusion chromatography analysis of recombinant TarM-variants, the variations in molecular dimensions are at best marginal (Fig. 5), indicating that the putative alteration is small.

\textbf{DISCUSSION—}\We have determined the first structure of an enzyme in the biogenesis pathway of poly-RboP WTA, and we have characterized the ligand binding site of this enzyme. Our work sheds light on an essential aspect of \textit{S. aureus} glycosylation and can be used as a template for understanding similar reactions in related organisms.

Glycosyltransferases can be classified into two groups that either retain the stereochemistry of the donor anomic bond (\textit{\alpha} \rightarrow \alpha) or that invert this bond during the transfer reaction (\textit{\alpha} \rightarrow \beta). A common feature of GT-4 class enzymes is that they retain glycosyltransferases, and combined with previous biochemical data our structural analysis suggests that TarM is also a retaining glycosyltransferase that employs an \textit{\alpha}-1-like mechanism in accordance with the widely acknowledged mechanism for a typical GT-4 class enzyme. The most salient structural features are shared by TarM and closely related GT-4 class enzymes MshA and BshA, and these latter enzymes can, therefore, serve as a useful basis for comparison.

The reaction mechanism for this class of enzymes has been established for MshA (38) and others (47–49). The acceptor substrates of GT-4 enzymes range from small molecules such as inositol phosphate to lipopolysaccharides and to \textit{S}-layer glycoproteins. Although the resolution of UDP-GlcNAc-bound TarM is only 2.8 Å and although the Gt-C domain is more
mobile and less well defined by electron density than the remainder of the protein, the electron density for UDP-GlcNAc is nevertheless unambiguous and allows placement of the ligand into the structure in the conformation shown in Fig. 4.

A critical difference to other GT-4 class enzymes such as MshA and BshA is that TarM has a HUB domain that is inserted into the Gt-N domain between helices 4 and 5 (Fig. 2D) and that folds into a long antiparallel β-sheet. The point of insertion of the HUB domain into Gt-N also happens to be the dimerization site for MshA and BshA. The HUB domain gives rise to a unique trimeric, propeller-like assembly of three glycosyltransferase domains. Given the proximity of the three-fold symmetry axis to the active sites, the HUB-generated trimer may also participate in interactions with WTA and assist with catalysis. Our mutational analysis clearly implicates the HUB domain in this process. It is interesting that a BLAST (50) sequence search of protein databases only finds HUB-like sequences in TarM homologs of other Gram-positive bacteria (NCBI# WP_029331270.1, identity 53%, similarity 71%; NCBI# WP_014124998.1, identity 43%, similarity 65%; NCBI# WP_025702814.1, identity 27%, similarity 50%; NCBI# WP_003756742.1, identity 33%, similarity 55%). To analyze the level of conservation of residues in TarM and its homologs, we generated a

FIGURE 5. Plating assay & CD-spectra of active site variants. A, EOP assay for functional analysis of TarM point mutations. The histogram shows the effects of key amino acids of TarM on phage susceptibility, as probed by complementation of RN4220 ΔtarMtarS with TarM carrying specific point mutations. Approximately 1000 pfu of phage 11 were mixed with 100 μl of bacteria suspension of optical density 0.4. After a brief incubation, soft agar was applied, and mixture was poured onto agar plates followed by overnight incubation at 37 °C. pfu was counted, and EOP of ΔtarMΔtarS complemented with wild-type tarM was designated as 100%. Mutant-tarM complements are indicated in relation as the mean of four experiments ± S.D. Statistically significant differences of mutant TarM from wild-type TarM complementation were calculated by paired two-tailed Student’s t test: ns, not significant, p > 0.05; *, p < 0.05; **, p < 0.01; ***, p < 0.001. B, overlaid CD spectra of recombinant TarM active site mutants H249A, R326S, K331S, and E403A and HUB domain mutants K136S, V159Y/C164R, and K136S/V159Y/C164R. Recombinant enzymes were purified according to the same protocol used for the wild-type enzyme (see “Experimental Procedures”). CD measurements were performed at concentrations ranging from 0.5 to 0.8 mg/ml in buffer G. C, overlay of size exclusion chromatography (SD200 16/60, GE Healthcare) elution profiles from the last step of recombinant protein purification with buffer D at 4 °C. All TarM variants elute in a volume range with an average peak point corresponding to a molecular size estimated to be 300 kDa according to the size calibration proteins shown on top. D, overlaid DLS spectra of recombinant wild type, E403A, and K136S/V159Y/C164R TarM at concentrations of 0.5–1 mg/ml performed at 20 °C. E, relative in vitro activity of TarM and selected TarM variants. Activity of wild-type TarM was set to 100%. Values are given as the mean of three experiments (n = 3) ± S.D. The reactions were carried out in the presence of 2 mM UDP-GlcNAc and 25 μM WTA at room temperature. The reaction was followed via a coupled enzymatic assay with non-saturating amounts of TarM variants.
sequence alignment (not shown) and colored the TarM surface according to the level of conservation (Fig. 6). As expected, the active site region and the UDP-GlcNAc binding site are rather conserved (red in Fig. 6). Interestingly, surface-exposed portions of the HUB domain that lie adjacent to the active site region are also well conserved, and because there is no obvious structural reason for this conservation, we predict that these regions might be involved in the binding of the second substrate, the RboP acceptor chain. As TarM-mediated WTA glycosylation is thought to constitute a general pathway in Gram-positive bacteria with RboP-WTA (19), it seems likely that the glycosyltransferases acting on WTA have a structural reason for this conservation, thereby facilitating the glycosylation of complex glycopolymers. The direct distance of two neighboring active sites in the TarM trimer is 72 Å, which corresponds to about eight or nine ribitol units in an extended chain. Thus, a single TarM trimer could simultaneously glycosylate the same poly-RboP substrate at different locations. It is not currently known which RboP units in the long polyribitol-phosphate chain are glycosylated, but it is likely that the glycosyltransferases acting on WTA have a mechanism that enables them to move along the polyribitol chain and selectively glycosylate specific units. The result of such a glycosylation pattern is for example relevant for the selectivity of pattern recognition receptors. Future studies of TarM in complex with WTA components should help reveal the molecular details of this process.

A recent structural analysis of the streptococcal glycosyltransferase GtfA has identified a novel domain that is very similar in structure to the HUB domain (40) and that is also inserted into a glycosyltransferase subdomain at a similar location. The GtfA domain is the only structure in the DALI database with any significant structural homology to the TarM HUB (Z-score 11.6, r.m.s.d. 2.1 Å, 103 aligned residues, sequence identity 16%, PDB 4PQG). However, GtfA is clearly monomeric, and the enzyme also does not act on WTA. A structural alignment shows that the novel GtfA domain is unlikely to form a similar trimeric arrangement due to an insertion sequence (PVDNK) that extends the turn connecting strands β9 and β10 (Fig. 7). This loop is much shorter in TarM, allowing trimer formation, and the tip of the loop moreover carries Val-159, which makes direct contacts to the two other Val-159 residues in the trimer and thus stabilizes the trimeric arrangement. Consistent with this, a mutation of Val-159 that would disrupt the trimer affects the ability of TarM to process WTA.

By now only a handful of teichoic acid biogenesis-affiliated protein structures have been solved for either polyglycerol phosphate-type or complex-type WTA, e.g. TagF from Bacillus subtilis (51), Streptococcus TarI (spr1149, PDB 2VSH) (52), MnaA (BA5590) from B. anthracis (PDB 3BEO) (53), putative WTA ligase from Streptococcus (54), or lipoteichoic acid polymerase LtaS (55). TarM is the first enzyme structure in the biogenesis pathway of poly(RboP)-type WTA to be reported.

Acknowledgments—X-ray data were collected at beam line X06DA of the Swiss Light Source (Villigen, Switzerland), and we are grateful to the beam line staff, particularly Vincent Oliferic and Tomizaki Takashi, for assistance. We also thank Nicolas Binder and Simon Huber for the purification of recombinant TarM variants used for CD spectroscopy.
REFERENCES

9884 JOURNAL OF BIOLOGICAL CHEMISTRY

VOLUME 290 • NUMBER 15 • APRIL 10, 2015

1. Baron, S. (1996) Medical Microbiology, 4th Ed., University of Texas Medical Branch at Galveston, Galveston, TX


Structural and Enzymatic Analysis of TarM Glycosyltransferase from Staphylococcus aureus Reveals an Oligomeric Protein Specific for the Glycosylation of Wall Teichoic Acid
Cengiz Koç, David Gerlach, Sebastian Beck, Andreas Peschel, Guoqing Xia and Thilo Stehle

J. Biol. Chem. 2015, 290:9874-9885. doi: 10.1074/jbc.M114.619924 originally published online February 19, 2015

Access the most updated version of this article at doi: 10.1074/jbc.M114.619924

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 53 references, 27 of which can be accessed free at http://www.jbc.org/content/290/15/9874.full.html#ref-list-1
An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus

Xuehua Li1, Cengiz Koç2, Petra Kühner1, York-Dieter Stierhof3, Bernhard Krismer1, Mark C. Enright4, José R. Penadés5, Christiane Wolz1, Thilo Stehle2,6,7, Christian Cambillau8, Andreas Peschel1,7 & Guoqing Xia1,7,9

Despite the importance of phages in driving horizontal gene transfer (HGT) among pathogenic bacteria, the underlying molecular mechanisms mediating phage adsorption to *S. aureus* are still unclear. Phage φ11 is a siphovirus with a high transducing efficiency. Here, we show that the tail protein Gp45 localized within the φ11 baseplate. Phage φ11 was efficiently neutralized by anti-Gp45 serum, and its adsorption to host cells was inhibited by recombinant Gp45 in a dose-dependent manner. Flow cytometry analysis demonstrated that biotin-labelled Gp45 efficiently stained the wild-type *S. aureus* cell but not the double knockout mutant ΔtarM/S, which lacks both α- and β-O-GlcNAc residues on its wall teichoic acids (WTAs). Additionally, adsorption assays indicate that GlcNAc residues on WTAs and O-acetyl groups at the 6-position of muramic acid residues in peptidoglycan are essential components of the φ11 receptor. The elucidation of Gp45-involved molecular interactions not only broadens our understanding of siphovirus-mediated HGT, but also lays the groundwork for the development of sensitive affinity-based diagnostics and therapeutics for *S. aureus* infection.

Recently, there has been a renewed interest in phage-bacteria interactions because phages have not only profound influence on the biology of bacterial pathogens1–3 but also promising applications in the detection of pathogens, the biocontrol of bacterial food contamination4, and the treatment of bacterial infections4.

Phages infecting gram-positive bacteria need to adsorb and penetrate a cell envelope with a thick peptido-glycan meshwork. The mechanism of phage adsorption and genome translocation across the gram-positive cell envelope remains largely unknown for many phages, with the exception of a few dairy phages infecting *Lactobacillus*, *Lactococcus*, or *Streptococcus* spp. Genome comparison of several dairy phages with different host ranges enabled the identification of their receptor binding proteins (RBPs), which are essential for phage adsorption and virulence. The first RBP recognizing a gram-positive cell envelope was identified from phage Dt1 infecting *Streptococcus thermophilus*5. Recently, the structures of RBPs from several lactococcal phages were solved6. These RBPs are generally homotrimeric and are composed of three modular structures, which encompass the N-terminal shoulder domain for connection to the virion, a β-helical linker or the neck domain, and the C-terminal head domain bearing the receptor binding site for host recognition6.

Wall teichoic acids (WTAs) are phosphate-rich anionic glycopolymers covalently linked to the peptidoglycan in gram-positive bacteria. The two common types of WTA are either poly-1,3 glycerol-phosphate (GroP) or poly-1,5 ribitol-phosphate (RboP). The main chains of both types of WTAs can be further substituted with sugar residues and alanyl groups7. Previous studies on *Bacillus* phage SPP1 revealed that adsorption of this phage to its...
host cell initially depends on the reversible binding to WTAs, which accelerates the subsequent irreversible binding to membrane receptor YueB. Interestingly, incubation of the purified SPP1 virions with recombinant YueB leads to phage DNA release in vitro, indicating that the binding of this protein is the trigger for DNA injection.

*Staphylococcus aureus* is a gram-positive pathogen that causes not only superficial skin infections but also severe, deep tissue infections such as endocarditis, osteomyelitis, septic arthritis, and bacteraemia. It is very well known that phages or mainly siphoviruses play vital roles in the virulence, adaptation, and evolution of *S. aureus*1–3. However, it remains unclear how siphoviruses recognize *S. aureus* and what ligand-receptor interactions mediate phage adsorption to the cell surface of *S. aureus*.

Among all *S. aureus* phages, φ11 is probably one of the best-studied siphoviruses due to its high transduction efficiency and broad application in transducing genetic markers among *S. aureus* strains. Recently, there has been a growing interest in studying the function of φ11 as a helper phage mediating the horizontal gene transfer (HGT) of *S. aureus* pathogenicity islands (SaPIs)19. We have shown that staphylococcal siphoviruses use α-O-GlcNAc modified WTA as a receptor11 and that WTA structures govern phage-mediated horizontal transfer of SaPIs among major bacterial pathogens12. Although many structural proteins of φ11 have been reported13–15, its receptor binding protein (RBP) has yet to be identified. Here we report the identification and characterization of the φ11 RBP and the major components of its receptor in the cell wall of *S. aureus*. These data not only provide novel insight into phage-host recognition at the staphylococcal cell surface, but also establish a molecular basis to develop novel diagnostics and therapeutic treatments of *S. aureus* infection.

**Results**

**Sequence analysis of the putative baseplate proteins of φ11.** In staphylococcal siphovirus genomes, the genes coding for tail proteins are usually located downstream of the gene of the tape measure protein (TMP) and upstream of the lysis module1–15. Among the genes localized between *tmp* (gp42) and the lysis module, gp43, gp44, gp45 and gp46 (Fig. 1) were previously shown to be essential for phage φ11 infectivity13,16. Of note, gp54 was not initially annotated in the genome of φ1117, but it was later identified as an open reading frame localized between gp45 and gp4615. To advance an understanding of the putative functions of the proteins encoded by these four essential genes, HHpred18 (Homology detection and structure prediction by HMM–HMM comparison) analysis was carried out for each protein in addition to BlastP analysis at NCBI (http://goo.gl/DE9BkO).

The HHpred analysis identified Gp43 with 100% probability as a distal tail protein (Dit) because it is similar to the Dit protein (PDB 2×8K) in the baseplate of the siphophage SPP1, which infects *Bacillus subtilis*19 (Fig. 1). The N-terminal regions of Dit proteins form a hexameric ring and are very conserved among phages20, although their C-terminal peripheral domains may differ considerably21.

BlastP search revealed that Gp44 possesses an endopeptidase domain at its N-terminus (1–350 residues) and an SGNH/GDSL hydrolase domain at its C-terminus (406–633 residues). Of note, the SGNH hydrolase represents a diverse family of lipases and esterases, but the enzyme activity of Gp44 is yet to be characterized experimentally. Further sequence analysis by HHpred revealed that the N-terminal domain of Gp44 aligns well with the tail associated lysin (Tal) of bacteriophage MU (PDB 1WRU)22, and its C-terminal domain exhibits striking similarity to a carbohydrate esterase (PDB 2WAO) from *Clostridium thermocellum* (Fig. 1). Tal proteins are structurally

---

**Figure 1.** HHpred analysis of the four structural gene products following the *tmp* (gp42) of φ11. The genes gp42, gp43, gp44, gp45, gp54 and gp46 are represented by arrows. The tail proteins encoded by these genes are indicated above the arrows. The structural homologues of these tail proteins are presented in the boxes beneath the corresponding genes. The PDB identifiers and ribbon structures (α-helices in blue, β-strands in violet) are shown for the structural homologues. The starting and ending amino acid residues of the regions, which could well align with these structural homologues are indicated above the boxes. The similarity probability (%) returned by HHpred is indicated to the right of the PDB identifier. The PDB entries shown here include 2×8K, Bacillus phage SPP1 baseplate Dit protein; 1WRU, Tail associated lysin (Tal) of bacteriophage MU; 2WAO, carbohydrate esterase of *Clostridium thermocellum*; 3NOL, glutaminyl cyclase of *Zymomonas mobilis*; and 4DIV, BppU of Lactococcus phage TP901-1.
similar to Gp27, a baseplate component of the puncturing device of phage T4\(^{23}\). Notably, the gene \(tal\) is always localized directly downstream of the gene \(dit\) in siphophage genomes. In the \(\phi 11\) tail module, \(gp44\) exists directly downstream of \(gp43\) (\(dit\)). Hence, both sequence homology and conserved genome localization suggest that \(gp44\) encodes a Tal protein. Recently, it was shown that phage mutants deficient in \(gp43\) (\(dit\)), or \(gp44\) (\(Tal\)) were defective in tails, suggesting that these two baseplate proteins are required for tail formation\(^{46}\). Furthermore, it was shown that the tail protein Gp49 possesses peptidoglycan hydrolase activity but is dispensable for \(\phi 11\) infectivity\(^{46,24}\). These facts suggest that \(\phi 11\) may have two virion-associated peptidoglycan hydrolases, Tal and Gp49, but the activity of Tal needs to be verified by further experiments.

BlastP search with Gp45 as a query returned a hit of ORF636, which shares 44% identity with Gp45 and is localized at the tail tip of phage phiSLT, a serogroup A phage of \(S.\) \(aureus\). Of note, the tail protein ORF636 was characterized as an adhesion protein essential for phiSLT adsorption and infectivity\(^{25}\). HHpred analysis revealed that the central part of Gp45, covering amino acid residues 160–420, shares high similarity with 5-bladed propeller proteins (Fig. 1), for example the glutaminyl cyclase of \(Zymomonas mobilis\) (PDB 3NOL). The segment upstream was predicted to be \(\alpha\)-helical by Jpred\(^{26}\), while the segment downstream was predicted to form \(\beta\)-strands.

Just downstream of \(gp45\), \(gp54\) most likely encodes an upper baseplate protein (BppU)\(^{37}\). The N-terminus of Gp54 (amino-acids 1–195) displays high similarity to a large part of the BppU\(^{37}\), which attaches the RBP to the central baseplate core in lactococcal phage TP901-1 (Fig. 1). In TP901-1, BppU assembles as a trimer. Its N-terminus (amino-acids 1–120) is a stand-alone domain, while amino-acids 121–193 assemble as a triple \(\alpha\)-helix bundle. This structure is followed by a trimeric all-beta domain (~100 residues), to which the N-terminus of RBP is plugged in\(^{27}\). Thus, the C-terminus of BppU and the RBP exhibit strong shape complementarity in phage TP901-1. However, the C-terminus of Gp54 possesses a domain of unknown function, which includes ~400 amino acid residues, and is much larger than that of BppU in TP901-1.

Taken together, the HHpred analyses revealed that Gp43 (\(Dit\)), Gp44 (\(Tal\)), Gp45 (ORF636-like protein), and Gp54 (BppU) very likely constitute the baseplate of \(\phi 11\). Moreover, the central part of the \(\phi 11\) baseplate gathering \(Dit\), \(Tal\), and the N-terminus of BppU, forming the dsDNA passage, is similar to that of other phages\(^{26}\), whereas the role of the tail proteins, Gp45 and Gp54, most likely located at the periphery of the baseplate is elucidated below.

**Localization of \(\phi 11\) at the baseplate of \(\phi 11\).** Baseplate proteins or tail fibre proteins play critical roles in phage adsorption, the first step of phage replication cycle\(^6\). Previously it was shown that the two putative tail proteins Gp45 and Gp54 were essential for phage infectivity\(^{13}\). To demonstrate that both Gp54 and Gp45 are localized at the tail tip, both Gp54 and Gp45 antisera were raised and used for immunogold labelling of \(\phi 11\). Electron micrographs of negatively stained phage samples indicate that Gp45 and Gp54 are clearly localized at the tail baseplate of \(\phi 11\) (Fig. 2a,b), whereas immunogold labelling of mutant phages deficient in \(gp45\) or \(gp54\) resulted in negligible background labelling (Fig. 2c).

**Neutralization of \(\phi 11\) infection with anti-Gp45 or anti-Gp54 serum.** As both Gp45 and Gp54 are baseplate proteins, their roles in phage adsorption and infection were analysed. Phage \(\phi 11\) virions were pre-incubated with increasing concentrations of antisera before plating on the host. Notably, pre-immune sera exhibited hardly any inhibitory effects on phage plating efficiency (data not shown), whereas both anti-Gp45 and anti-Gp54 serum decreases the plating efficiency of \(\phi 11\) in a dose-dependent manner (Fig. 3a,b), which clearly suggests that these sera can specifically neutralize \(\phi 11\) infectivity. It is most likely that masking of Gp45 or Gp54 with antiserum prevents their access to the phage receptor in the cell wall, hence blocks the phage adsorption and leads to neutralization of \(\phi 11\).

**Gp45 binds to the cell wall with \(\alpha\)- or \(\beta\)-O-GlcNAc modified WTAs.** To investigate the molecular interaction of \(\phi 11\) with its cognate receptor on the host cell surface, recombinant Gp45 was expressed and purified (Supplementary Fig. S1). Pre-incubation of host cells with increasing concentrations of recombinant Gp45 led to dose-dependent inhibition of \(\phi 11\) adsorption (Fig. 4a).

We recently demonstrated that \(S.\) \(aureus\) siphoviruses use \(\alpha\)-O-GlcNAc modified WTAs as their adsorption receptor\(^{11}\). To examine whether Gp45 binds to WTAs, \(S.\) \(aureus\) wild-type strain RN4220 and mutants with altered WTAs were stained with biotin-labelled recombinant Gp45 and subsequently analysed by flow cytometry. In contrast to the well-stained wild-type \(S.\) \(aureus\) with glycosylated WTA, the mutants \(\Delta tarM/S\), which lacks \(\alpha\)-O- and \(\beta\)-O-GlcNAc residues on WTA, or \(\Delta tugO\), which is deficient in WTA, demonstrated drastically decreased background staining (Fig. 4b). These results indicate that Gp45 binds to the cell wall with \(\alpha\)- or \(\beta\)-O-GlcNAc modified WTAs. Unfortunately, recombinant Gp45 purified from \(E.\) \(coli\) was found to be susceptible to degradation and was therefore not suitable for flow cytometry analysis.

**The major components of the \(\phi 11\) receptor in the cell wall of \(S.\) \(aureus\).** Previous studies have shown that the entire cell wall of \(S.\) \(aureus\) could inactivate \(S.\) \(aureus\) phages, while the isolated WTAs could not\(^{28,29}\). Additionally, treating the cell wall preparations with either muramidase or amidase or using deacetylated cell walls destroyed the phage inactivation capacity of these preparations\(^{30,31}\). These observations suggested that peptidoglycan may participate in phage adsorption directly or indirectly by providing rigid support for WTAs. These data prompted us to re-examine phage adsorption with an extended set of \(S.\) \(aureus\) cell wall mutants. In particular, we aimed to investigate how phages interact with WTAs, and how peptidoglycan structures affect their adsorption.

Adsorption assays were carried out using isogenic mutants with altered WTAs as hosts. As shown in Fig. 5a, \(\phi 11\) virions were able to adsorb to either the \(\Delta tarM\) mutant with only \(\beta\)-GlcNAc residues on WTA or the \(\Delta tarS\)
mutant with only α-GlcNAc residues on WTA with efficiency comparable to that of wild-type cells. In contrast, phage adsorption was significantly impaired when the ΔtagO mutant, devoid of WTAs, or the double mutant ΔtarM/S, deficient in both α- and β-GlcNAc residues on WTAs, were used as hosts (Fig. 5a). Consistent with these findings, the cell wall preparation from the wild-type strain dose-dependently inactivated phage, with full inactivation reached at a concentration of 240 nmol phosphate per reaction, whereas the cell wall preparation from double mutant ΔtarM/S exhibited significantly less inhibitory effect on plating efficiency at a similar concentration (Fig. 5b). Collectively, these observations demonstrate that GlcNAc residues on WTAs are essential for phage adsorption regardless of their anomeric configurations.

*S. aureus* cell wall preparations with deacetylated peptidoglycan fail to inactivate phage 52A 31, which is also a serogroup B phage like φ11. To examine if peptidoglycan acetylation is involved in φ11 adsorption, the oatA mutant 32 deficient in 6-O acetylation of muramic acid residues in peptidoglycan was used as a host for the adsorption assay. As shown in Fig. 5c, phage adsorption efficiency decreased to 50% when compared to the wild-type adsorption, suggesting that peptidoglycan acetylation favours φ11 adsorption.

**Figure 2. Immunogold labelling of tail proteins Gp45 and Gp54.** (a,b) Transmission electron microscopy (TEM) images of negatively stained φ11 after immunogold labelling with anti-Gp45 serum (a) and anti-Gp54 serum (b), respectively. (c) TEM images of mutant phages Δgp45 and Δgp54. (Left), mutant phage Δgp45 labelled with anti-Gp45 serum; (Right), mutant phage Δgp54 labelled with anti-Gp54 serum. Insets show enlarged views of the boxed areas.
Pip homologues in *S. aureus* do not play a role in φ11 adsorption. Some phages require a membrane-embedded protein receptor for irreversible binding before the translocation of the phage genome into the host cell. Previous studies on phage-resistant mutants derived from *L. lactis* identified the phage infection protein (Pip) as the membrane receptor for lactococcal phage c23. YueB, the Pip homologue in *B. subtilis*, was also identified as the membrane receptor for siphophage SPP13. Using the amino acid sequence of YueB or Pip as a probe, two homologues with conserved membrane topology and 40% similarity to YueB were identified from the *S. aureus* genome and designated as Pip1 (SAV2643) and Pip2 (SAV0283), respectively. To determine if these two membrane proteins are involved in φ11 adsorption, knockout mutants deficient in pip1, pip2, or both were generated. Interestingly, φ11 plates well on pip mutants, and no decrease in φ11 adsorption efficiency was observed when these mutants were used as a host (Fig. 5d), suggesting that Pip homologues in *S. aureus* are not involved in phage φ11 adsorption.

**Discussion**

Research on *S. aureus* phages has a very long history that can be traced back to the early studies of bacteriophages. Since the discovery of bacteriophages, many *S. aureus* phages have been isolated, and these were classified into...
Before molecular techniques became available, *S. aureus* phages had been widely used for typing *S. aureus*. It was known for a long time that many *S. aureus* phages carry virulence genes and are required for *S. aureus* virulence and adaptation. Despite comprehensive studies on phage genomes and the role of *S. aureus* phages in horizontal transfer of resistance and virulence genes among clones and species, the molecular interactions mediating phage adsorption to the staphylococcal cell surface remain poorly understood.

The mechanism underlying *S. aureus* phage adsorption has often been assumed to be similar to that of phages infecting gram-negative bacteria. However, as gram-positive bacteria have a very different cell wall structure compared to that of gram-negative bacteria, phages infecting gram-positive bacteria may employ adsorption mechanisms different from those infecting gram-negative bacteria. Accounting for over 50% of the cell wall mass, WTAs are the most abundant surface molecules in the cell wall of bacteria belonging to the order *Bacillales*, which includes genera such as *Bacillus*, *Listeria* and *Staphylococcus*. Hence, it is most likely that phages infecting bacteria of these genera need to interact with WTAs for successful adsorption.
In this study, we demonstrated that GlcNAc residues on WTAs are essential for φ11 adsorption regardless of their anomeric configurations. We also found that 6-O-acetylation of muramic acid residues in peptidoglycan is involved in φ11 adsorption. We showed that Gp45 and Gp54 are two baseplate proteins critical for φ11 infection, as both antisera can neutralize φ11 infection dose-dependently. Recombinant Gp45 inhibits φ11 adsorption in a dose-dependent manner and binds to glycosylated WTAs, demonstrating that Gp45 is the RBP of φ11. Unfortunately, recombinant Gp54 purified from E. coli was not stable and hence unsuitable for cell wall binding studies, and its functions could not be tested.

Staphylococcal pathogenicity islands (SaPIs) have an intimate relationship with temperate staphylococcal phages. Phages can induce the SaPI cycle, which allows the SaPIs to be efficiently encapsidated into special small phage heads commensurate with their size. Previous mutational analyses of the genes present in the morphogenesis cluster of φ11 demonstrated that the Gp45 was essential for both the phage infectivity and transduction of its cognate SaPI11. Of note, Δgp45 seemed to lose its baseplates and failed to plate on S. aureus. Surprisingly, Δgp54 was still able to transduce SaPIs, although with a 100-fold reduction in transduction efficiency when compared with wild-type φ11. These results highlight that Gp45 is essential for the recognition process, while the presence of the Gp54 significantly increases the binding affinity between the phage and its receptor. As the N-terminus of Gp54 was predicted to be similar to that of BppU, which maintains the attachment of RBP to the baseplate core in TP901-1, it is tempting to speculate that Gp54 plays an important role in anchoring RBP in the baseplate.

Previously, a tail protein ORF636 from a serogroup A phage phiSLT was characterized as an adhesion protein that binds to poly-glycerolphosphate (GroP) chain of lipoteichoic acids (LTAs). Notably, ORF636 shares high homology with Gp45 (62% similarity) and the ORF636 sequence exists in all known serogroup A phages infecting S. aureus. However, it was shown that all tested serogroup A phages inflicting S. aureus. Despite its homology with Gp45, ORF636 was not essential for phage infection. The tight binding of glycerol and glycerolphosphate for the RBP suggested that LTAs could act as receptors for lactococcal phages, however the structure of LTAs is well conserved and thought to be too simple to explain the different host specificities of various lactococcal phages. Recently, by mutational analysis, it was demonstrated that cell wall polysaccharide (CWPS) is the host cell surface receptor of tested lactococcus phages of different groups and that differences between the CWPS structures play a crucial role in determining phage host range.

It is noteworthy that many phages need a protein receptor for adsorption, for example, FhuA, OmpA, OmpC, LamB for E. coli phages, GamR, YueB for Bacillus phages and Pip for Lactococcus phage. Interestingly, all these protein receptors are non-essential and many of them were identified by transposon mutagenesis. However, by screening a mutant library of S. aureus we were unable to isolate φ11-resistant mutants, which carry transposon insertions in genes encoding membrane proteins. It is now generally acknowledged that carbohydrate-recognizing phages possess a broad baseplate structure with multiple receptor binding sites. Conversely, phages with stubby ends or tail fibres, including the lactococcal c2 phages and the Bacillus phage SPP1, may recognize protein receptors on the cell surface. The crystal structure of Gp45 was solved and it was found that Gp45 forms six trimeric baseplates in the baseplate of φ11 and that each monomer of Gp45 contains a five-bladed propeller domain with a cavity that could accommodate a GlcNAc moiety (Koc et al., unpublished data). Hence, the presence of 18 receptor binding sites in the baseplate of φ11 suggests that its receptors are saccharides but not proteins.

Accounting for over 50% of the cell wall mass, WTAs are considered to be the most abundant surface molecules in S. aureus and have been implicated in various critical processes and interactions such as staphylococcal cell division, biofilm formation, β-lactam resistance, and staphylococcal pathogenesis. Due to the in-homogeneity of WTA, its analysis has proven to be very challenging. Unlike research carried out on DNA, RNA or protein, methods available for studying WTA function are very limited. Despite technical limitations, a few WTA-interacting proteins such as PmtA, WTA antibody, MBL, and SREC-1 have recently been identified. Here, we report Gp45 as a new WTA-interacting protein. Our results may eventually provide new tools for detecting the subdomain structures in the cell wall of S. aureus. Additionally, this study establishes a solid basis for the development of sensitive affinity-based infection diagnostics and therapeutics for MRSA infection.

Materials and Methods

Bacterial strains and growth conditions. S. aureus strains used in these studies are listed in Table 1. Bacteria were grown at 37°C in BM broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl, 0.1% K2HPO4, 0.1% glucose) under agitation.

Construction of S. aureus mutants. The deletion mutants Δpip1, Δpip2, and Δpip1/2 were constructed by allelic exchange. For knockout plasmid construction, the primers listed in Table S1 were designed. For deletion of pip1, flanking regions were amplified with primer pairs pip1-F1-up/pip1-F1-down and pip1-F2-up/pip1-F2-down. Purified PCR products were ligated into pSall/Nhel and Nhel/EcoRI respectively, and subsequently ligated into the Sall/EcoRI-digested knockout vector pBAD63. The resulting plasmid was used for allelic exchange. For the construction of the pip2 deletion mutant, a similar approach was pursued. The flanking regions of pip2 were amplified with primer pairs pip2-F1-up/pip2-F1-down and pip2-F2-up/pip2-F2-down, digested with XbaI and ligated. Afterwards this marker-less knockout cassette was subcloned into pKOR-1, and the resulting plasmid was used for mutant construction via allelic exchange.

Overexpression and purification of the recombinant Gp45 and Gp54. Both gp45 and gp54 were amplified by PCR from S. aureus strain SA113, which is a φ11 lysogen. The primers used for the PCR reaction are

Accounting for over 50% of the cell wall mass, WTAs are considered to be the most abundant surface molecules in S. aureus and have been implicated in various critical processes and interactions such as staphylococcal cell division, biofilm formation, β-lactam resistance, and staphylococcal pathogenesis. Due to the in-homogeneity of WTA, its analysis has proven to be very challenging. Unlike research carried out on DNA, RNA or protein, methods available for studying WTA function are very limited. Despite technical limitations, a few WTA-interacting proteins such as PmtA, WTA antibody, MBL, and SREC-1 have recently been identified. Here, we report Gp45 as a new WTA-interacting protein. Our results may eventually provide new tools for detecting the subdomain structures in the cell wall of S. aureus. Additionally, this study establishes a solid basis for the development of sensitive affinity-based infection diagnostics and therapeutics for MRSA infection.

Materials and Methods

Bacterial strains and growth conditions. S. aureus strains used in these studies are listed in Table 1. Bacteria were grown at 37°C in BM broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl, 0.1% K2HPO4, 0.1% glucose) under agitation.

Construction of S. aureus mutants. The deletion mutants Δpip1, Δpip2, and Δpip1/2 were constructed by allelic exchange. For knockout plasmid construction, the primers listed in Table S1 were designed. For deletion of pip1, flanking regions were amplified with primer pairs pip1-F1-up/pip1-F1-down and pip1-F2-up/pip1-F2-down. Purified PCR products were ligated into pSall/Nhel and Nhel/EcoRI respectively, and subsequently ligated into the Sall/EcoRI-digested knockout vector pBAD63. The resulting plasmid was used for allelic exchange. For the construction of the pip2 deletion mutant, a similar approach was pursued. The flanking regions of pip2 were amplified with primer pairs pip2-F1-up/pip2-F1-down and pip2-F2-up/pip2-F2-down, digested with XbaI and ligated. Afterwards this marker-less knockout cassette was subcloned into pKOR-1, and the resulting plasmid was used for mutant construction via allelic exchange.

Overexpression and purification of the recombinant Gp45 and Gp54. Both gp45 and gp54 were amplified by PCR from S. aureus strain SA113, which is a φ11 lysogen. The primers used for the PCR reaction are
listed in Table S1 in the Supplementary Information. The amplified gp45 or gp54 genes were subcloned into the expression vector PET28a between the NheI and XhoI sites. The resulting plasmids were transformed into E. coli BL21 for overexpression of Gp45 or Gp54. Both proteins were fused to a hexa-histidine-tag at the N-terminus to facilitate purification. After IPTG induction of the host cells, recombinant Gp45 was extracted and purified according to the procedure described previously49. Briefly, cells were lysed via ultrasonication (Digital Sonifier, Branson). After centrifugation at 38,000 × g for 55 min, cell debris was removed, and the supernatant containing recombinant Gp45 protein was loaded on a 5 mL Ni-NTA-column (GE Healthcare). Fractions containing Gp45 were pooled and concentrated to 1 mg/mL using Vivaspin 20 centrifugal concentrators with a molecular size cut-off of 50,000 (Sartorius, Göttingen, Germany). The concentrated sample was then loaded on a size-exclusion chromatography column SD200 pre-equilibrated with SEC-buffer containing 25 mM HEPES, 150 mM NaCl, 1 mM DTT. Fractions containing Gp45 were pooled and concentrated as pure Gp45 preparations. The purity and folding of the recombinant Gp45 were assessed with SDS-PAGE, Circular dichroism (CD) spectroscopy and dynamic light scattering (DLS). Gp54 was purified by the same procedure as for Gp45.

**Preparation of cell wall from S. aureus strains.** The cell wall was extracted according to the procedure described previously50. Briefly, S. aureus overnight cultures were harvested by centrifugation at 5000 × g for 10 minutes. The cells were washed with 20 mM NH4Ac buffer (pH 4.8) to remove the intact cells. The supernatant was collected and used as a crude extract of cell wall and mixed well with 5 mM MgSO4, 40 U/mL DNase and 80 U/mL RNase at final concentrations before overnight incubation at 37 °C. Next, to remove any cell membrane contamination, SDS was added to a final concentration of 2%, followed by ultrasoundication for 15 min. After heating at 65 °C for one hour, the cell wall preparations were washed six times with 20 mM NH4Ac buffer by centrifugation at 13,000 × g. Finally, the cell wall preparations were re-suspended in distilled water and quantified by measuring the amount of inorganic phosphate using the QuantiChrom™ Phosphate Assay Kit (BioAssay Systems, USA) as described previously50.

**Table 1. Bacterial strains used in this study.**

<table>
<thead>
<tr>
<th>Bacterial strain</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL21</td>
<td>E. coli BL21, host of inducible recombinant protein expression</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>RN4220</td>
<td>S. aureus strain deficient in restriction, capsule, or prophage.</td>
<td>51</td>
</tr>
<tr>
<td>∆tagO</td>
<td>RN4220, ∆tagO</td>
<td>11</td>
</tr>
<tr>
<td>∆tarM</td>
<td>RN4220, ∆tarM</td>
<td>52</td>
</tr>
<tr>
<td>∆tarS</td>
<td>RN4220, ∆tarS</td>
<td>52</td>
</tr>
<tr>
<td>∆tarM/S</td>
<td>RN4220, ∆tarM, ∆tarS</td>
<td>52</td>
</tr>
<tr>
<td>∆pip1</td>
<td>RN4220, ∆pip1</td>
<td>This study</td>
</tr>
<tr>
<td>∆pip2</td>
<td>RN4220, ∆pip2</td>
<td>This study</td>
</tr>
<tr>
<td>∆pip1/2</td>
<td>RN4220, ∆pip1, ∆pip2</td>
<td>This study</td>
</tr>
<tr>
<td>SA113</td>
<td>Derivative of S. aureus strain NCTC8325 harboring prophages ø11, ø12, and ø13</td>
<td>53</td>
</tr>
<tr>
<td>∆oatA</td>
<td>SA113, ∆oatA</td>
<td>32</td>
</tr>
</tbody>
</table>

**Bacteriophage experiments.** Using the double layer soft agar method, ø11 was propagated with the indicator strain, S. aureus strain RN4220, as a host. Phage plating efficiencies were determined to investigate the plating efficiencies of Gp45, Gp54 anti-sera and cell wall preparations on the inactivation of ø11. In brief, 100 μL of ø11 (3 × 106 PFU/mL) was mixed with 100 μL of cell wall preparations or antisera of certain concentrations and incubated at 37 °C for 10 min. Samples pre-incubated without any cell wall preparations or sera served as controls. Next, the mixtures were diluted before plating on the indicator strain (S. aureus strain RN4220) using double agar overlay methods. After overnight incubation at 37 °C, the plaques were enumerated. The efficiency of plating was calculated relative to that of plating of strain RN4220 without any cell wall preparations or sera served as controls. Next, the mixtures were diluted before plating on the indicator strain (S. aureus strain RN4220) using double agar overlay methods. After overnight incubation at 37 °C, the plaques were enumerated. The efficiency of plating was calculated relative to that of plating of strain RN4220 without any cell wall preparations or sera served as controls.

Adsorption assays were performed according to the procedure described previously11. Briefly, 200 μL of S. aureus wild-type or mutant cells containing 8 × 105 CFU were mixed with 100 μL of ø11 containing 3 × 109 PFU and incubated at 37 °C for 15 min. The bound phases were separated from the free phases by centrifugation at 13,000 × g for 5 min. Adsorption was calculated by determining the number of PFU of the unbound phase in the supernatant and subtracting it from the total number of input PFU. Adsorption efficiency was expressed relative to the adsorption of wild-type strain RN4220. Each adsorption assay was repeated at least three times. To study the inhibition of adsorption by Gp45, cells were pre-incubated with the purified recombinant Gp45 of indicated concentrations for 15 min before adding phages to the host cells.

**Purification of ø11 and electron microscopy methods.** Phage ø11 lysate was centrifuged at 73000 × g, 4 °C for two hours (Beckman Optima XL-80K). The resulting pellet was re-suspended in 500 μL of TMN buffer.
containing 10 mM Tris-HCl, pH 7.5, 10 mM MgSO₄, 500 mM NaCl. The sample was then mixed well with 55% CsCl in TMN-buffer to give a final concentration of 42% CsCl and subjected to ultracentrifugation at 245,000 × g, 15°C for 20 hours (Beckman). The visible phage band on the CsCl gradient was collected and sequentially dialyzed for two hours each in a D-Tube Dialyzer Mini (Novagen®, Merck Millipore, Darmstadt, Germany) against decreasing concentrations of NaCl in TMN buffer (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 4 M NaCl) until the NaCl concentration after each round of dialysis was at 4 M, 2 M, 1 M and 10 mM NaCl, respectively.

For immunogold labelling, purified phage samples were adsorbed to glow discharged, piaoform and carbon-coated grids. The grids were then blocked with 0.2% gelatin in phosphate-buffered saline for 10 min followed by incubation with rabbit anti-Gp45 or rabbit anti-Gp45 serum, which were diluted in blocking buffer at 1:20 and 1:100, respectively. Polyclonal rabbit antisera were raised against purified recombinant Gp45 or Gp54 using a custom antigen antibody service, Speedy 28-Day polyclonal program from Eurogentec (Brussels, Belgium). After blocking at room temperature for 60 min, the grids were washed six times with blocking buffer for a total time of 15 min before incubation with goat anti-rabbit IgG coupled with 12 nm gold colloids (Dianova, Hamburg), which was diluted with blocking buffer at 1:30. After incubation at room temperature for 60 min, the grids were washed three times with blocking buffer for 10 min and three times with phosphate-buffered saline for 10 min, followed by washing four times with double-distilled water for 2 min. Finally, the grids were negatively stained with 1% (w/v) aqueous uranyl acetate before examination with a JEM-1400Plus transmission electron microscope (JEOL, Japan)

Flow cytometry analysis. Flow cytometry was carried out to evaluate the binding of recombinant Gp45 to the S. aureus cell surface. Purified recombinant Gp45 was labelled with biotin using the EZ-Link™ NHS-Biotin kit (Thermo Fisher Scientific). Biotin-labelled Gp45 was then incubated with S. aureus wild-type or mutant cells for 30 min with shaking at room temperature. Cells were washed and stained with strep-Alu488 (Invitrogen) for one hour at 4°C. Finally, cells were fixed for flow cytometry analysis.

Statistical analysis. Results are expressed as the means ± standard deviations from at least three independent experiments. Statistical analysis was performed using GraphPad Prism (GraphPad Software, Inc., La Jolla, USA, Version 5.04). Statistically significant differences were calculated with two-tailed Student’s t-test or one-way ANOVA with Bonferroni’s post-test as indicated.

References
Author Contributions

X.L. and G.X. designed this study; X.L., C.K., P.K., Y.S. and B.K. performed the experiments; X.L., Y.S., M.E., J.P., C.W., T.S., C.C., A.P. and G.X. analysed the data; C.C., M.E. and G.X. wrote the manuscript. Every author reviewed the manuscript prior to submission.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.
How to cite this article: Li, X. et al. An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus. Sci. Rep. 6, 26455; doi: 10.1038/srep26455 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
An essential role for the baseplate protein Gp45 in phage adsorption to *Staphylococcus aureus*

Supplementary information:

Xuehua Li¹, Cengiz Koç², Petra Kühner¹, York-Dieter Stierhof³, Bernhard Krismer¹, Mark C Enright⁴, José R Penadés⁵, Christiane Wolz¹, Thilo Stehle²,⁶,⁸, Christian Cambillau⁷, Andreas Peschel¹,⁸, Guoqing Xia¹,⁸,⁹,*

1. Interfaculty Institute of Microbiology and Infection Medicine, University of Tübingen, 72076 Tübingen Germany,
2. Interfaculty Institute of Biochemistry, University of Tübingen, 72076 Tübingen, Germany
3. Center for Plant Molecular Biology, University of Tübingen, 72076 Tübingen, Germany
4. School of Healthcare Sciences, Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, United Kingdom
5. Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom
6. Vanderbilt University, School of Medicine, Nashville, TN 37232, USA
7. Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, UMR 6098, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France
8. German Center for Infection Research (DZIF), partner site Tübingen, Germany
9. Institute of Inflammation & Repair, Faculty of Medical and Human Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom

* Correspondence to Dr. Guoqing Xia, e-mail: guoqing.xia@manchester.ac.uk
Supplementary Table S1: Primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp45-up</td>
<td>CACGGCTAGCATGAGTAATAAAACTAATTACAGATTTAAGTAG</td>
</tr>
<tr>
<td>gp45-dn</td>
<td>CAGTCTCGAGTTATCAGCACCACCTTTCCCTTGAAT</td>
</tr>
<tr>
<td>gp54-up</td>
<td>CACGGCTAGCATGGTAGATATATTTCCTCTCAGAAAGATGA</td>
</tr>
<tr>
<td>gp54-dn</td>
<td>CAGTCTCGAGTTATTTCCACCCCCATAATTCTGTTATAGT</td>
</tr>
<tr>
<td>pip1-F1-up</td>
<td>CTAGGATCTGTTCGACTAATGATGTTG</td>
</tr>
<tr>
<td>pip1-F1-dn</td>
<td>GGAAGCTAGCTAGTCGATGACATTACCTATC</td>
</tr>
<tr>
<td>pip1-F2-up</td>
<td>CCAGAGTTCGCTAGCGAATAAAAATT</td>
</tr>
<tr>
<td>pip1-F2-dn</td>
<td>CAGAATTCACATTTTTAAAATACTTCCAGCG</td>
</tr>
<tr>
<td>pip2-F1-up</td>
<td>GGGGACAAATTGGTACAAAAAAGCAGGCTGCCCCGCGGTGTCTTTATTCTACTTTC</td>
</tr>
<tr>
<td>pip2-F1-dn</td>
<td>GTCATCTAGAATCTTTGGGATCCCTTTTCGATGTCTCC</td>
</tr>
<tr>
<td>pip2-F2-up</td>
<td>GTCATCTAGAATCTTTTAATGCGATGGGTACTTCCGATGTCGACTCTGGGC</td>
</tr>
<tr>
<td>pip2-F2-dn</td>
<td>GGGGACCACTTTTTGTAAGAAAGCTGGGATGAAACTGTCACGTAGTTCAGTCTTATGCGATGTC</td>
</tr>
</tbody>
</table>

Supplementary Figure S1: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the purified Gp45 (a) and Gp54 (b) preparation from *E. coli*. The gel was stained with Coomassie brilliant blue R-250.
Structure of the host-recognition device of *Staphylococcus aureus* phage φ11

Cengiz Koç1, Guoqing Xia2,3,4, Petra Kühner2, Silvia Spinelli5,6, Alain Roussel5,6, Christian Cambillau5,6 & Thilo Stehle1,3,7

Phages play key roles in the pathogenicity and adaptation of the human pathogen *Staphylococcus aureus*. However, little is known about the molecular recognition events that mediate phage adsorption to the surface of *S. aureus*. The lysogenic siphophage φ11 infects *S. aureus* SA113. It was shown previously that φ11 requires α- or β-N-acetylglucosamine (GlcNAc) moieties on cell wall teichoic acid (WTA) for adsorption. Gp45 was identified as the receptor binding protein (RBP) involved in this process and GlcNAc residues on WTA were found to be the key component of the φ11 receptor. Here we report the crystal structure of the RBP of φ11, which assembles into a large, multidomain homotrimer. Each monomer contains a five-bladed propeller domain with a cavity that could accommodate a GlcNAc moiety. An electron microscopy reconstruction of the φ11 host adhesion component, the baseplate, reveals that six RBP trimers are assembled around the baseplate core. The Gp45 and baseplate structures provide insights into the overall organization and molecular recognition process of the phage φ11 tail. This assembly is conserved among most glycan-recognizing *Siphoviridae*, and the RBP orientation would allow host adhesion and infection without an activation step.

*Staphylococcus aureus* is a Gram-positive bacterium that causes a wide range of infections. It is a leading cause of bacteremia, infective endocarditis, as well as osteoarticular, skin and soft tissue, pleuropulmonary and device related infections1. Methicillin-resistant *S. aureus* (MRSA) remains a severe global problem threatening the health care system as resistance restricts treatment options to a few drugs of last resort.

All *S. aureus* genomes sequenced to date contain one or several prophages2,3. Most *S. aureus* phages can be integrated into the bacterial chromosome or exist as extra-chromosomal elements. It is known that many of these phages encode a large variety of *S. aureus* virulence or fitness factors that allow the bacterium to escape the host immune system. Among all mobile genetic elements in *S. aureus*, phages are probably most efficient in mediating horizontal gene transfer of virulence or resistance genes between strains, and across species or even genus. Therefore, phages play important roles in staphylococcal pathogenicity and adaptation of *S. aureus* to different hostile environments2,3.

The large number of staphylococcal phages sequenced to date display an extensive mosaicism in their gene structure, which is a consequence of gene shuffling among different phages that can infect staphylococcal species. The resulting mosaic gene organization is consistent with a modular evolution involving exchanges of genome modules by horizontal transfer and genetic recombination. The genetic exchanges of modules can involve single genes, protein domains, groups of genes, or even functional modules3.

Although phages are the most abundant and diversified biological entity on earth, each phage can only infect a limited number of bacterial strains. This specific phage-host interaction is determined, in part, by the protein recognition device located at the tip of the phage tail, which engages a receptor at the bacterial cell surface. Since...
bacterial cell wall polysaccharides or glycopolymers project from the cell surface and are thus easily accessible, they are the most common molecules targeted by bacteriophages.

The cell wall of *S. aureus* typically contains poly-ribitol phosphate type wall teichoic acid (WTA), which is modified with D-alanine and N-acetyl-glucosamine (GlcNAc). *S. aureus* φ11 is often used as model to study horizontal gene transfer of virulence genes. Recently it was shown that φ11 requires GlcNAc residues on WTA for adsorption, *Gp45 of φ11* was identified and characterized as the receptor-binding protein (RBP) of φ11. Furthermore, it was shown that φ11 was unable to bind to the cell wall in the absence of WTA–GlcNAc, identifying glycosylated WTA as the receptor.

Phages adopt a two-fold strategy for host adhesion. They first deploy adhesion modules on fibers or on capsid or tail that recognize the host cell wall glycans structures in a reversible way: this allows cell wall scanning in search for the final, specific receptor, to which they bind irreversibly. Attachment to membrane protein often requires a unique and strong attachment of the phage's tail tip, as observed in phage T5. In contrast, the loose affinity observed between saccharides and proteins requires the presence of several attachment sites provided by a multimeric RBP carrying device, the baseplate. In order to provide a foundation for understanding the initial recognition mechanism of phage φ11 and its receptor in the cell wall of *S. aureus*, we embarked on structural analyses of the Gp45 and baseplate of φ11 using X-ray crystallography and electron microscopy. The RBP structure reveals a trimer with a complex fold that can be divided, from N- to C-terminus, into a "stem", a "platform" and a "tower". The stem is formed by a long, severely bent triple helix, a "platform" formed by a repetitive all-β coiled coil that features three interruptions: the first and third interruptions are both β-hairpin structures and the second is a short disordered region. A putative "hinge"-like feature is located between the second and the third interruption. The stem is followed by a "platform" of three β-propellers, and the protein terminates with a "tower" formed by a repetitive all-β domain. Platform and tower are interconnected by a fifth short triple helix buried inside of the protein on the molecule's longest three-fold axis. An unusual iron is located at the C-terminal end of the first coiled coil and may play a role in mediating flexibility or conformational rearrangements within the helical domain. Six copies of the trimer assemble around the baseplate core. This hexameric organization is commonly observed in lactococcal *Siphoviridae*, and it is compatible with host adhesion and infection in the absence of an activation step.

**Results**

**Structure determination.** The *gpa45* gene was cloned into the pET28 vector (Novagen) for overexpression in *E. coli* as described elsewhere. Briefly, the protein was expressed with an N-terminal hexa-histidine-tag, and purified by nickel-affinity chromatography and size exclusion chromatography as a trimer. Structure determination was performed with a TaBr4 derivative using single isomorphous replacement with anomalous scattering (SIRAS) and exploiting the non-crystallographic symmetry (NCS) present in the crystals. Initial refinement with PHENIX was followed by several runs with autoBUSTER, and alternating refinement and model building cycles resulted in excellent Rfree and Rwork values of 21.1% and 17.5%, respectively, for the final model (Table 1). Although the map is generally of good quality, a few loops of the propeller domains have very weak electron density, explaining the persistence of a small number of outliers in the Ramachandran plot (0.5%). Of the remaining residues, 94.5% are located in regions of preferred conformation and 5% in regions that are classified as allowed.

**Overall structure of φ11 RBP.** The RBP of φ11 assembles into an elongated homotrimer, with overall dimensions of approximately 160 × 120 × 100 Å (Fig. 1A). The structure can be divided into an N-terminal “stem” region that forms a triple-helical bundle (Figs 1B and S1), a central “platform” region composed of three β-propeller domains (Figs 1D and S1) and a C-terminal “tower” region (Figs 1E and S1). Overall, the stem contains three non-helical interruptions. The first of these occurs between residues 46 and 67 and contains a bound iron as well as a β-hairpin that faces away from the bundle axis (Figs 1C and 2). The second and the third interruptions are located between residues 81 and 107, which introduce a sharp kink into the stem (the "hinge") and thus break the shared three-fold symmetry of α1 and α2 (Figs 1C and 3). Helix α3 (residues 88–97), a short triple helical coil located in the hinge, has an independent rotation axis not aligned to the remainder of the molecule (Fig. 3). This helical bundle is followed by five-bladed β-propeller modules of the platform, which encompass residues 142–439 and form the midsection of the protein (Fig. 1). This “platform” is linked via a short helix (residues 425–432) to the C-terminal “tower”. The latter contains two structurally similar domains (residues 440–541 and 542–636), which are each formed by three five-stranded anti-parallel β-sheets, one from each monomer, that are covered on their surface-exposed side by loops and one short α-helix each (Fig. 1).

Each φ11 RBP monomer forms extensive contacts with the two others monomers in the trimer. For each contact, 2 × 5,800 Å2 (11,600 Å2) are buried in the interaction as calculated by PISA. This results in a total buried surface of ~35,000 Å2 for the trimer. Most of the buried surface area is concentrated in the stem and C-terminal regions, while the propeller domains engage in few intermolecular contacts (Fig. S2).

**The stem structure.** The stem comprises three separate triple-helical bundles, which are composed of helices α1, α2 and α3/α4, respectively (Fig. 1B). The helices pack tightly together in each of the bundles, and almost every residue of each monomer is in contact with a residue of one of the two other monomers (Fig. S2) through central hydrophobic contacts or lateral hydrogen or ionic bonds. The trimeric ensemble comprising the extended N-terminus and helix α1 can be superposed onto the 30 first residues of phage TP901-1 RBP (PDB code 3U6X) with a r.m.s.d. value of 1.8 Å for 90 Cα atoms (Fig. 1B). The remaining helical bundles most closely resemble those found in phage TP901-1 Baseplate protein Upper (BppU) and other viral trimeric helix bundles.

A strong electron density feature suggesting the presence of a metal ion was observed at the junction between the first two bundles. Using an RBP crystal and extended X-ray absorption fine structure (EXAFS) spectroscopy,
The identity of this ion was determined to be iron, which probably exists in its oxidized form Fe$^{3+}$ (analyzed at SOLEIL beamline PX1) (Fig. S3). The Fe$^{3+}$ ion is positioned along the 3-fold axis of the first helix bundle and coordinated by the side chains of His42 and His50 from each of the three monomers (Fig. 1C). This gives rise to a near-perfect octahedral coordination, in which the HisN$\varepsilon_2$-Fe distances range from 2.19 Å to 2.32 Å. His42X is in close hydrogen-bond distance to an acid/base pair, which forms a second shell around the His-Fe-octahedron. The side-chain functional groups of Glu46X and Arg43Z allow for an arrangement in which the deprotonated glutamate-carboxylate is oriented to His42X-N$\varepsilon_1$ (2.66–2.81 Å), forcing His42X-N$\varepsilon_2$ to point towards the Fe-center. A similar tautomerization effect might occur for the diametrically opposed His50Z, as it is in close distance to Gln54Z (3.24–3.27 Å), forcing His50Z-N$\varepsilon_2$ to coordinate to Fe$^{3+}$, resulting in an intertwined chelate-complex comprised of all three protein chains (see Fig. 2). It is noteworthy that the His42-N$\varepsilon_2$-Fe$^{3+}$ distances are comparable (2.295 ± 0.025 Å) but significantly longer than the His50-N$\varepsilon_2$-Fe$^{3+}$ distances (2.21 ± 0.02 Å). Such a Fe$^{3+}$ binding geometry has previously been observed in the membrane-piercing spike proteins of phages P2 (PDB code 3QR7) and φ92 (PDB code 3PQH)$^{24}$, as well as in the receptor-binding domain of the long tail fiber of phage T4 (PDB code 2XGF)$^{25}$. In the three reported cases, the N$\varepsilon_2$-Fe distances were 2.20 ± 0.01, 2.23 ± 0.01 and 2.31 ± 0.06, respectively, compared to an average of 2.25 ± 0.15 Å and 2.16 ± 0.13 Å, respectively$^{26}$. However, all of these values are larger than those observed for distances between a heme Fe$^{3+}$ and the N$\varepsilon_2$ of histidines coordinating it axially in myoglobin (2.00–2.11 Å)$^{27}$. Indeed, iron ions often absorb in visible light wavelength ranges, giving rise to a red color for hemes and a brownish color for Fe-S clusters. However, the φ11 RBP and the related phage proteins discussed above are all colourless in solution$^{24}$. It is worth noting that the Fe$^{3+}$ binding regions in the phage P2, φ92 and T4 spike structures involve histidines at the apex of an intertwined triple β-helix. It has been proposed that this ion binding structure might strengthen the puncturing device of phages that pierce the cell wall$^{24}$. In φ11 RBP, the His42-X$_{α1}$/His50 motif lies at a junction between two helical bundles. We therefore suggest that it serves a different role, perhaps by helping to stabilize the bundles that undergo a sharp turn at the hinge. If the Fe lock would not be in place, the structure of the α1/α2 segment would likely not be maintained as a rigid unit.

While the first and second helical bundles are collinear, the hinge introduces a sharp angle of ~30° between the second and the third bundle. This angle is the smallest possible since the second and third bundles are in contact at

<table>
<thead>
<tr>
<th></th>
<th>Derivative Ta6Br12</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>87.65, 89.60, 93.73</td>
<td>87.06, 89.08, 93.3</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>92.7, 105.7, 117.9</td>
<td>93.0, 105.2, 117.6</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50–3.30 (3.37–3.30)</td>
<td>44.4–2.20 (2.38–2.20)</td>
</tr>
<tr>
<td>R$_{free}$ (%)</td>
<td>12.5 (100.4)</td>
<td>6.6 (57.5)</td>
</tr>
<tr>
<td>CC (1/2) (%)</td>
<td>99.9 (87.1)</td>
<td>99.8 (83.5)</td>
</tr>
<tr>
<td>I/σI</td>
<td>20.4 (2.9)</td>
<td>12.6 (2.4)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3 (95.0)</td>
<td>97.1 (97.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>13.7 (12.0)</td>
<td>3.2 (3.2)</td>
</tr>
<tr>
<td>Phasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sites</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Anomalous phasing power</td>
<td>1.4 (0.143)</td>
<td></td>
</tr>
<tr>
<td>Figure of merit - acentric</td>
<td>0.26 (0.05)</td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>44.4–2.20 (2.26–2.20)</td>
<td></td>
</tr>
<tr>
<td>No. reflections</td>
<td>116,564 (8497)</td>
<td></td>
</tr>
<tr>
<td>R$<em>{free}$/R$</em>{free}$ (%)</td>
<td>17.5/21.1 (21.7/25.1)</td>
<td></td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>15324</td>
<td></td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>1446</td>
<td></td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>51.1</td>
<td></td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>33.6</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>54.2</td>
<td></td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.12</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. X-ray data collection, phasing and refinement statistics of Gp45.
The hinge geometry is such that the sequences of the three helices of bundles two and three remain in phase (Fig. 3). The second bundle terminates with Met80, and the following sequences in the three monomers adopt a coil structure that abuts helices $\alpha_3$ where the three sequences are already in phase (see Asp94, Fig. 3). Helices $\alpha_3$ are followed by extended hairpin-structures and helices $\alpha_4$ forming the final bundle.

The five-bladed propeller platform and the two C-terminal tower domains. The C-terminal end of the third helical bundle abuts the three five-bladed propeller domains that form the platform of $\phi 11$ RBP (Fig. 1D). The three propellers are all equidistant to each other and to the molecule's main NCS-axis. This whole platform domain occupies a space that is ~100 Å wide and ~40 Å thick. Contacts between the three propeller domains are sparse, as each interface between two propeller domains buries a surface area of only 457 Å$^2$ from solvent, and much of this surface is buried due to a helix-helix contact at the center of the trimer axis (Fig. S2). The plane of the propeller is not perpendicular to the 3-fold axis, but is tilted upwards (as represented in Fig. 1) by an angle of ~30°. This tilt improves access to the lower face of the propeller, and this might be linked to the function of RBP in interacting with ligands (see below).

As in other propeller structures, sets of four anti-parallel $\beta$-strands form each blade, and the N-terminal $\beta$-strand closes the fold by forming the final blade (blade 5) with the three C-terminal $\beta$-strands of the domain (Fig. 1D). A DALI search$^{28}$ with the $\phi 11$ RBP propeller returned many significant hits above a Z-score of 15, and with r.m.s.d. values ranging from 3.1 to 4.0 Å. Most of the identified proteins are enzymes that mediate the degradation of carbohydrates. The highest score ($Z = 16.1$, r.m.s.d. = 3.1 Å), however, was obtained for the
enzyme glutamine cyclotransferase from *Zymomonas mobilis* (PDB code 3NOL)\(^2\). To our knowledge, only two other examples of β-propellers in putative RBPs have been reported. For one, a distorted five-bladed propeller has been identified as the head domain of the RBP-P2 protein of phage PRD1, a Tectiviridae member infecting Gram-negative bacteria (PDB code 1N7U)\(^3\). The second example is the C-terminal domain gp131C of the *Pseudomonas* myophage PhiKZ, forming a seven-bladed β-propeller domain (PDB code 4GBF)\(^3\), and its position

---

**Figure 2. The Fe\(^{3+}\) binding motif.** (A) Final 2Fo-Fc electron density map of the Fe\(^{3+}\) binding motif contoured at 2.0 σ. The salmon-coloured sphere represents the octahedrally coordinated iron center in the histidine-rich stem region of helix-bundle α1. A second shell of residues, responsible for maintaining the active tautomeric states of the histidines, are represented with their side chains as sticks. Dashed lines indicate hydrogen-bonds and polypeptide chains are coloured blue for chain A, magenta for chain B and green for chain C. (B) Schematic representation of the octahedral motif, with the distances between His-Nε2 atoms from residues His 42/His 50 and the Fe\(^{3+}\) ion.

**Figure 3. The hinge domain.** (A) Ribbon representation of the trimeric hinge domain, encompassing amino acids Met80 to Gln108. Equivalent residues in the three monomers have been joined, and the sequence phase is conserved at this position. The direction of each chain is indicated by arrows. (B) 90° rotated view.
at the periphery of the baseplate has led to speculation that the propeller might act as the receptor-binding domain or as a cell-degrading enzyme. Neither of these hypotheses, however, have been confirmed experimentally.

When the *Zymomonas mobilis* glutamine cyclotransferase structure was superimposed onto the RBP propeller domain, the active site of the enzyme overlaid a deep crevice located within the RBP lower face (Fig. 4). Modelling indicates that a cavity in this crevice has the correct size to accommodate a GlcNAc molecule, the cell wall teichoic acid (WTA) component specifically recognized by phage φ11. Three water molecules occupy this cavity (Fig. S4A), which can be nicely replaced with hydroxyl groups of a modelled GlcNAc molecule. The cavity is lined with polar residues (Gln165, Thr211, Gln330), which could serve to establish hydrogen bonds with the modelled GlcNAc molecule (Fig. S4B). Apolar residues Met164 and Met329 complete the walls of the cavity.

The two C-terminal tower domains form a structure of dimensions $60 \times 50 \times 50$ Å. These two domains are very similar in structure, which is confirmed by a superposition that yields a low r.m.s.d. value of 1.7 Å for their Cα atoms (Fig. 1E,F). A DALI search performed with these domains returned only lower Z-scores, with the highest of these ($Z = 6.1; r.m.s.d. = 3.2$ Å) for uracil-DNA glycosylase inhibitor, a small all-β monomeric protein (PDB code 2UGI). The two structures essentially share the same anti-parallel β-sheet but differ in their oligomeric state and their surrounding structural features, and the identity (9%) is not high enough to assign possible functions to the C-terminal domains.

**Negative staining electron microscopy structure of the φ11 baseplate.** To define the topology of the φ11 baseplate and allow location of the RBP, we conducted electron microscopy analysis using negative staining of the virion. This approach has been successful in other cases. We collected 512 images of the phage, and boxed 778 baseplate particles (see experimental procedures section). The final map has a resolution of 23 Å (determined using the 0.5 FSC criterion) and allowed us to unambiguously place six φ11 RBP trimers (Fig. 5A,B). To optimize this fit, we modified the hinge angle between the second and the third helix bundles from a value of ~30° to ~90°. The correlation is 0.845 with 95.5% of the atoms inside the map, calculated for a RBP orientation fit with the tower domain inclined towards the bottom of the baseplate. Compared to this, an orientation in which the tower domain would be “heads up”, reminiscent of the resting state of p2 baseplate, only resulted in a correlation of 0.826 with 74.9% of the atoms inside the map. However, it has to be taken into account that the “heads-down” conformation is not a completely non-flexible state. The angle of the stem might in fact vary from the minimum observed in the X-ray structure to larger values when the phage scans the host’s surface by moving the RBPs around the calculated average position for adhesion to the specific receptor. Such movements have been observed for several phages, such as phage T7.

The remainder of the RBP structure was left unaltered, and the modified trimers fit well in a peripheral region of the map that could accommodate the triangular shape of the platform domain’s platform. In order to explain the remaining density of the baseplate, we performed HHpred analyses of φ11 proteins Gp43, Gp44 and Gp54, which are the most likely candidates for baseplate components. This analysis revealed similarities with components of the lactococcal phage TP901-1 baseplate, suggesting that the central part of the φ11 baseplate is organized similarly to that of TP901-1. Based on this analysis, Gp43 is predicted to exist as a hexamer and form the distal tail protein (Dit) ring and Gp44 as a trimer forming the tail-associated lysin (Tal) N-terminus and extension, while Gp54 N-terminus (the functional equivalent of BppU N-terminus) may form a second
ring. Furthermore, the N-terminal segment and the first helical bundle of φ11 RBP are structurally homologous to the N-terminal part of the phage TP901-1 RBP trimer, a structural domain that anchors the RBP into the BppU C-terminus. We therefore also attempted to fit the phage TP901-1 Dit hexamer together with the BppU N-terminus (amino acids 1–160) into our electron density map. The ring of the Dit had appropriate dimensions to fit the map above the RBPs (Fig. 5C,D). In contrast, the structure equivalent to BppU could not be fitted unequivocally as the internal density is not defined sufficiently. A large volume of the EM map remains to account for the Gp54 and for the Tal (Fig. 5E). When attaching the trimeric Tal N-terminal domain below the Dit hexamer, the three carbohydrate binding modules (2WAO) identified by HHpred project in the direction of the tail tip. These three bulky modules should fill the electron density map in between the six RBP trimers.

Discussion

We have solved the crystal structure of φ11 RBP and located this protein in the tail spike of the assembled phage using electron microscopy. Our analysis defines the domain organization of RBP, which can be divided into a stem region, a platform domain and a tower-like C-terminal structure composed of two nearly identical domains. Interestingly, the stem displays a severely bent, hook-like conformation that may undergo a conformational change as the protein can only be fitted into the electron density of the tail spike in a less bent arrangement. Unexpectedly, the stem also contains a bound iron. The function of this iron is unknown as its location differs from irons found in other spike proteins. The platform region is formed by three propeller domains and likely harbours the binding site for the substrate GlcNAc. Although soaking and cocrystallization experiments with GlcNAc were not successful, modelling suggests a reasonable location for the GlcNAc binding site in the platform region. Of note, the propeller fold was identified in the endosialidases of several phages. These enzymes cleave polysialic acid at the surface of their host in order to obtain access to the capsular cell wall. For example, coliphages K1F and phi92 possess such endosialidases, which also exhibit trimeric propeller domains attached to a stem.

Comparison with the phage TP901-1 tail spike assembly allows us to also assign a putative location of the Gp43 and Gp54 proteins of φ11. Gp43 likely forms the hexameric Dit, while the N-terminus of Gp54 resembles the first 160 residues of TP901-1 BppU. The remainder of the electron density is likely occupied by the rest of the large Gp54 and by Gp44, the Tal protein. Interestingly, the N-terminal folds of Dit and Tal are found in a wide range of phages, including Myoviridae infecting Gram-negative bacteria (T4 or Mu) Siphoviridae from Gram-negative (T5) or Gram-positive (SPP1) bacteria, lactococcal phages, or even mycobacteria (Araucaria). It is worth noting that the Tal protein is also found in the type VI secretion system machinery. This observation suggests that the block formed by Dit and Tal could have been conserved through evolution, a phenomenon shared by other components such as the capsids MCP, the connector, as well as the tail MTP.
Only the periphery of Dit (its C-terminal domain) and the Tal extension (e.g. a C-terminal fiber) could have been adapted to specific phage infection-style requirements\(^5\). In contrast, 11 Gp45, the RBP, does not exhibit analogy with other phage RBPs, in particular with those from lactococcal phages that also bind to saccharidic receptors. Lactococcal phages p2\(^5\), TP901-1\(^1\), Tuc2009\(^1\), bIL170\(^1\) and 1358\(^4\) all possess a trimeric receptor recognition head sharing a bono fide or a modified jelly-rol motif. The rest of their RBPs share common motifs in the neck or in the N-terminal domain (or stem). In Gp45, only the first 30 amino-acids of the stem resemble those of phages TP901-1 or Tuc2009.

Although the phage TP901-1 BppU protein does not seem to share such an extensive evolution coverage, the presence of a large part of it either in the RBP (Gp45) or in the subsequent protein (Gp54) was quite surprising. This finding suggests that phages might not capture only widespread elements in the protein domains repertoire, but also less diffused components, even between remote phages with different hosts. In the present case, the role of Gp54 is not documented and difficult to predict. We think it likely that in phage 11 Gp54, a large C-terminal domain might also accommodate the RBP N-terminus, but with different structural features compared to TP901-1 BppU, because of its much increased size. The electron microscopy low-resolution structure gives hints of the putative receptor binding sites, located below the five-bladed propeller domain. This arrangement allows for a correct orientation to capture the GlcNAC of the cell wall teichoic acids. Although, a well-defined cavity in the size of a monosaccharide exists, a much larger crevice surrounds this cavity, suggesting that other WTA components might complement the interaction. However, further structural data are necessary to develop this hypothesis.

**Experimental Procedures**

**Overexpression, purification and crystallization of Gp45.** Gp45 was produced and purified as described elsewhere\(^5\). Briefly, after induction with IPTG the protein was purified to homogeneity using nickel-affinity chromatography and size exclusion chromatography. The purified protein carries a hexa-histidine tag at its N-terminus. Two similar crystallization solutions (0.1 M bicine/Trizma base pH 8.5, 10% w/v PEG 8000, 20% v/v ethylene glycol, 0.12 M monosaccharide-mix\(^5)\) yielded initial crystals (50 × 20 × 5 μm) of triangular shape that grew to bouquets at 16 °C over 1 week. Reproduction of the crystals in 5 μL hanging drops lead to bigger crystals (500 × 200 × 50μm) that were used for X-ray structural analysis.

**Phasing, construction, refinement.** Data for native and derivative crystals were collected at the Swiss Light Source (SLS) on beamline X06DA (PXIII) using a PILATUS 2M hybrid pixel detector. For the determination of peak, inflection, high-remote and low-remote wavelengths from fluorescence spectra, the program CHOOCH was adjusted to the absorption edges of Ta-L-II (1.11325 Å peak) and Ta-L-III (1.25476 Å peak). Data were processed with the XDS package\(^6\). The crystals belong to spacegroup P1 and have unit-cell dimensions: a = 87.06 Å, b = 89.01 Å, c = 93.98 Å, α = 110.6°, β = 105.7° and γ = 117.9°. A Ta3Br12 derivative (a = 87.65 Å, b = 89.60 Å, c = 93.73 Å, α = 92.7°, β = 105.7° and γ = 117.9°) was prepared by soaking native crystals in crystallization solution + 2 mM Ta3Br12 for up to 2 weeks before backsoaking in crystallization condition and vitrification in liquid nitrogen. Anomalous data were processed according to MAD, SAD, MIRAS and SIRAS protocols with the SHARP-module Sushi and were refined until electron density maps showed good contrast. The outcomes of the various phasing protocols were compared, and the map derived from the SIRAS protocol was selected for further improvement.

A threefold NCS was elicited from the self-rotation function via polarrfn (ccp4\(^5\)), giving a strong signal for rotation in reciprocal space for eulerian angles (α = 357.4°, β = 63.6°, γ = 110.6°) corresponding to polar angles (α = 37.4°, β = 33.4°, γ = 120.1). Using this self-rotation solution, GETAX\(^6\) was able to find a set of translation vectors for the asymmetric unit in real space.

Due to the size of the multi-domain protein, it was split in various parts for further processing: two for the stem (before and after the 'hinge'), the platform domain and the C-terminal tower domain. Molecular masks\(^4\) were created for each part and by generating correlation maps of them separately, a set of NCS operators could be assigned to each of them. The NCS-matrices were refined with IMP and averaging with AVE\(^5\)–\(^7\) converged the respective domains to about 80–90% of correlation. Each subdomain was integrated with respective NCS-matrices into a DM script for a combined density modification\(^5,\)\(^6\). Starting at 5.8 Å, 80 cycles of consecutive solvent flattening, NCS-averaging, histogram matching and phase extension to a final resolution of 2.9 Å resulted in an interpretable map, which was clearly distinguishable from the unbiased calculated map. Initial refinement was carried out with REFMAC\(^5,\)\(^6\) and PHENIX\(^2\), and after each step model building was done in COOT\(^2\). The final rounds of refinement were performed with autoBUSTER\(^1\), leading to Rfree/Rwork values of 21.1 and 17.5% (Table 1). A portion of the electron density map is shown in Fig. 3. Structural images were generated using pymol\(^5\).

**Negative staining electron microscopy.** Phages were purified as previously described\(^3\). Purified 11 phage (5 μL, 10\(^9\) pfu) was applied to glow-discharged carbon-coated grids and left to adsorb for one min. Sample excess was blotted off and the grids were stained with 10 μL of 1% uranyl acetate for 30 sec. Micrographs (512) were recorded on a 2Kx2K FEI Eagle CCD camera using a Tecnai Spirit electron microscope operated at 120 kV and a magnification of 48,500 (resulting in a pixel size of 4.83 Å/pixel) (Fig. S5A). The three-dimensional reconstruction was produced using a single particle procedure and the XMIPP software package\(^5\). Particles defined around the baseplate (778) were manually picked and subjected to maximum likelihood (ML) classification and alignment implemented in Xmipp\(^7\) imposing a 6-fold symmetry. The initial volume was determined using a random sample consensus (RANSAC) approach\(^9\) with 5 2D classes. The resolution of the final volume was estimated at 23 Å using the Fourier Shell Correlation (FSC) 0.50 criterion (Fig. SSB).
Molecular fitting and structure visualization. Molecular graphics and analyses were performed with the UCSF Chimera package (Resource for Biocomputing, Visualization, and Informatics at UC-San Francisco)\(^7\). The model/EM map fitting was performed by the option “fit in map” of the “volume” register. The Dit fitting resulted in a correlation coefficient of 0.85 with 94% of the atoms in the map volume. The correlation coefficient calculated for six RBPs, with an orientation of the tower domains inclined towards the bottom of the baseplate, is 0.845 with 95.5% of the atoms inside the map.

Data deposition. X-ray structures and structure factors have been deposited with the Protein Data Bank (PDB, www.rcsb.org) under accession code SEVF. The EM map of the baseplate reconstruction has been deposited at the Electron Microscopy Data Bank (EMDB, emdatabank.org).

References

RepoRts to thank its beamline support team for assistance. This work was supported in part by the French Infrastructure Switzerland for provision of synchrotron radiation beamtime at beamline PXIII (x06da) of the SLS and would like to thank its beamline support team for assistance. This work was supported in part by the French Infrastructure.
for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01. TS, GX and CK were supported by Collaborative Research Centers 766 and TR34 from the Deutsche Forschungsgemeinschaft.

Author Contributions
C.K. solved crystal structure, interpreted data and wrote manuscript. G.X. supervised experiments and wrote manuscript. P.K. produced and purified the phage particles. S.S. collected EM data. A.R. supervised experiments. C.C. and T.S. supervised the work and wrote the manuscript. Every author reviewed the manuscript prior to submission.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Koç, C. et al. Structure of the host-recognition device of Staphylococcus aureus phage φ11. Sci. Rep. 6, 27581; doi: 10.1038/srep27581 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
Structure of the host-recognition device of *Staphylococcus aureus* phage φ11

Supplementary information

Cengiz Koç¹, Guoqing Xia²,³,⁴, Petra Kühner², Silvia Spinelli⁵,⁶, Alain Roussel⁵,⁶, Christian Cambillau⁵,⁶ * and Thilo Stehle¹,³,⁷ *

1. Interfaculty Institute of Biochemistry, University of Tübingen, 72076 Tübingen, Germany
2. Interfaculty Institute of Microbiology and Infection Medicine, University of Tübingen, 72076 Tübingen, Germany
3. German Center for Infection Research (DZIF), partner site Tübingen, Germany
4. Institute of Inflammation & Repair, Faculty of Medical and Human Sciences, University of Manchester, Manchester, United Kingdom
5. Architecture et Fonction des Macromolécules Biologiques, UMR 7257 CNRS, 13288 Marseille Cedex 09, France.
6. Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, UMR 6098, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France
7. Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

*Correspondence to:
Thilo Stehle, E-mail: thilo.stehle@uni-tuebingen.de
or Christian Cambillau, E-mail: ccambillau@gmail.com
Supplementary Figure S1. Primary and secondary structure of one Gp45 protomer. A/ Colour-coded sequence of Gp45 monomer: “Stem” in blue, “platform” in green, “tower1” in light-red and “tower2” in dark-red. Highlighted in coloured boxes are special features: the iron-binding-site in dark red, the interruptions in the stem in black, the putative GlcNAc binding motif in pink and the disordered loop in yellow. B/ Topology of one Gp45 protomer partitioned in a colour-code according to Figure S1A.
Supplementary Figure S2. Buried surface area along the Gp45 chain. The water accessible surface area of one monomer, buried by the two other monomers, has been calculated by PISA for each residue. Note the extensive interaction of the trimers along the stem (amino acids 1-140), while practically no interaction is observed between the β-propeller domains.
Supplementary Figure S3. EXAFS spectrum of φ11 RBP crystal. X-ray fluorescence spectra obtained upon illumination of a φ11 RBP crystal by 12.65 keV incident X-rays. The only detectable fluorescent line emission corresponds to the iron K-α energy at 6.40 keV (highlighted in light blue). The large peak on the right corresponds to the diffused scattering of incident X-rays.
Supplementary Figure S4. Putative GlcNAc binding-site. A/ Hydroxyl groups O3, O4 and O6 of a GlcNAc substrate can be overlayed with W901, W1275 and W1267, respectively, defining a specific binding mode for a glucose residue. B/ The overlay shown in A facilitates the modelling of GlcNAc, a major moiety of staphylococcal peptidoglycan, in a preformed cavity. The peptide-oxygen of M164 and the sidechain of Q165 of blade 2 are in appropriate proximity of O6 and O1, respectively, allowing for hydrogen bonds. The second half of GlcNAc is accommodated by M329 and Q330 of blade 5, allowing for hydrogen bridges to the nitrogen of the N-acetyl moiety, O4 and O6, respectively. This arrangement keeps GlcNAc stably in position.
Supplementary Figure S5. Electron microscopy of phage φ11. A/ An EM picture of the whole phage φ11. B/ The Fourier shell correlation curve plotted against 1/resolution. The FSC 0.5 cut-off yields a resolution of 23 Å.