Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Streptococcus agalactiae* tip pilin GBS104: a model for GBS pili assembly and host interactions

Vengadesan Krishnan,[a] Prabhat Dwivedi,[b] Brandon J. Kim,[c] Alexandra Samal,[d] Kevin Macon,[d] Xin Ma,[b] Arunima Mishra,[b] Kelly S. Doran,[c] Hung Ton-That[b] and Sthanam V. L. Narayana[d]*

[a]UNESCO Regional Centre for Biotechnology (RCB), Gurgaon 122 016, Haryana, India, [b]University of Texas Health Science Center, Houston, TX 77030, USA, [c]Department of Biology and Center for Microbial Sciences, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA, and [d]Center for Biophysical Sciences and Engineering, School of Optometry, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Correspondence e-mail: narayana@uab.edu

Received 5 November 2012
Accepted 18 February 2013

PDB References: GBS104N3, 3tvy; GBS104N3mut, 3two; GBS104N2N3, 3txa

The crystal structure of a 75 kDa central fragment of GBS104, a tip pilin from the 2063V/R strain of *Streptococcus agalactiae* (group B streptococcus; GBS), is reported. In addition, a homology model of the remaining two domains of GBS104 was built and a model of full-length GBS104 was generated by combining the homology model (the N1 and N4 domains) and the crystal structure of the 75 kDa fragment (the N2 and N3 domains). This rod-shaped GBS104 model is constructed of three IgG-like domains (the N1, N2 and N4 domains) and one vWFA-like domain (the N3 domain). The N1 and N2 domains of GBS104 are assembled with distinct and remote segments contributed by the N- and C-termini. The metal-binding site in the N3 domain of GBS104 is in the closed/low-affinity conformation. Interestingly, this domain hosts two long arms that project away from the metal-binding site. Using site-directed mutagenesis, two cysteine residues that lock the N3 domain of GBS104 into the open/high-affinity conformation were introduced. Both wild-type and disulfide-locked recombinant proteins were tested for binding to extracellular matrix proteins such as collagen, fibronectin, fibrinogen and laminin, and an increase in fibronectin binding affinity was identified for the disulfide-locked N3 domain, suggesting that induced conformational changes may play a possible role in receptor binding.

1. Introduction

The Gram-positive pathogen *Streptococcus agalactiae*, also known as group B streptococcus (GBS), is a leading cause of life-threatening infections such as pneumonia, septicemia and meningitis in neonates, as the pathogen is transferred from the mother to the fetus either through vertical transmission or by exposure during passage through the birth canal. In addition, GBS is responsible for significant morbidity and mortality in the United States, Europe and around the world (Dermer *et al.*, 2004) as the infection rates of nonpregnant, elderly and immunocompromised adults have exceeded that of neonates (Bolaños *et al.*, 2001; Edwards *et al.*, 2005; Schuchat, 1998; Sendi *et al.*, 2008; Dermer *et al.*, 2004). Therefore, a universal vaccine against multiple GBS strains (Johri *et al.*, 2006) is a necessity, especially given increased antibiotic resistance and the need for alternative targets in the development of anti-infective drugs (Heelan *et al.*, 2004; Lin *et al.*, 2000), which has resulted in a focus on GBS pilus proteins as vaccine candidates (Margarit *et al.*, 2009; Nuccitelli *et al.*, 2011).

Pathogens utilize pili for adherence to the host cells and understanding pilus assembly is an emerging area of research. The pili in Gram-negative bacteria are the best characterized and include type I pili (chaperone–usher pathway; Kline *et al.*,

2010; Proft & Baker, 2009; Sauer *et al.*, 2000), type IV pili (type II secretion system; Craig *et al.*, 2004) and curli-type pili (nucleation-dependent polymerization; Wu & Fives-Taylor, 2001). Gram-positive bacteria express type IV pili similar to those of Gram-negative organisms, but the Gram-positive type I pili are assembled by the covalent cross-linking of two or more pilin subunits catalyzed by an enzyme called sortase. The model for sortase-mediated pilus polymerization, first presented for *Corynebacterium diphtheriae*, is constituted of three distinct pilins: one major and two minor (Ton-That *et al.*, 2004; Ton-That & Schneewind, 2004).

As observed in all MSCRAMMs (microbial surface components recognizing adhesive matrix molecules), the pilin subunits carry an N-terminal signal peptide and a C-terminal sorting signal consisting of the sorting motif LP*X*TG followed by a hydrophobic domain and a positively charged tail (Fischetti *et al.*, 1990). In addition, the major pilins of several Gram-positive pathogens also carry a conserved pilin motif (YPKN) which is essential for pilus polymerization. In the *C. diphtheriae* pili model the major pilin SpaA has a pilin motif while the minor pilin SpaC at the pilus tip does not, and the minor pilin SpaB hosts a pilin-like motif for its incorporation into the pilus fiber (Linke *et al.*, 2010; Mandlik *et al.*, 2008). Ton-That *et al.* (2004) previously suggested the requirement of a glutamate in the SpaA E box (Y*X*L*X***E**T*X*AP*X*GY), a feature that is found to be conserved in many other Gram-positive major pilins, for incorporation of SpaB into the pilus fiber (Ton-That *et al.*, 2004). However, Kang and coworkers revealed the conserved E-box Glu residue as the catalytic residue for the formation of an intramolecular isopeptide bond dictating the stability of the major pilin SpaA (Kang, Middleditch *et al.*, 2009; Kang, Paterson *et al.*, 2009). A SpaB mutant that lacks the cell-wall sorting motif but that has the pilin-like motif is incorporated into the pilus, but the resulting pili are not anchored to the cell wall (Mandlik *et al.*, 2008), suggesting that SpaB is the pilus-terminating and cell-wall-anchoring pilin. Corroborating this, the SpaB deletion mutant behaved similarly to the deletion mutant of housekeeping sortase SrtF, where the SpaA polymers were mostly secreted into the culture medium (Mandlik *et al.*, 2008). In addition, the Lys residue in the pilin-like motif of SpaB is required for SpaB–SpaA linkage (Mandlik *et al.*, 2008), similar to the requirement of the Lys residue in the pilin-like motif of the minor pilin FctB for linking to the LP*X*TG motif of the major pilus protein FctA in group A streptococcus (GAS; Linke *et al.*, 2010). Hence, the SpaB subunit can appear along the pilus shaft when the lysine of the (individual or polymerized) SpaA pilin motif is linked to the threonine of the SpaB LAFTG sorting motif by a pilus-specific sortase (Ton-That & Schneewind, 2003; Mandlik *et al.*, 2008), and also as a base for pilus assembly linked to the cell wall (Mandlik *et al.*, 2008). This sortase-mediated pilus-assembly mechanism, with some differences, is conserved in several other Gram-positive bacteria, such as GBS, *Streptococcus pyogenes* (GAS), *Streptococcus pneumoniae*, *Enterococcus faecalis* and *Actinomyces naeslundii* (Kang & Baker, 2012; Krishnan & Narayana, 2011).

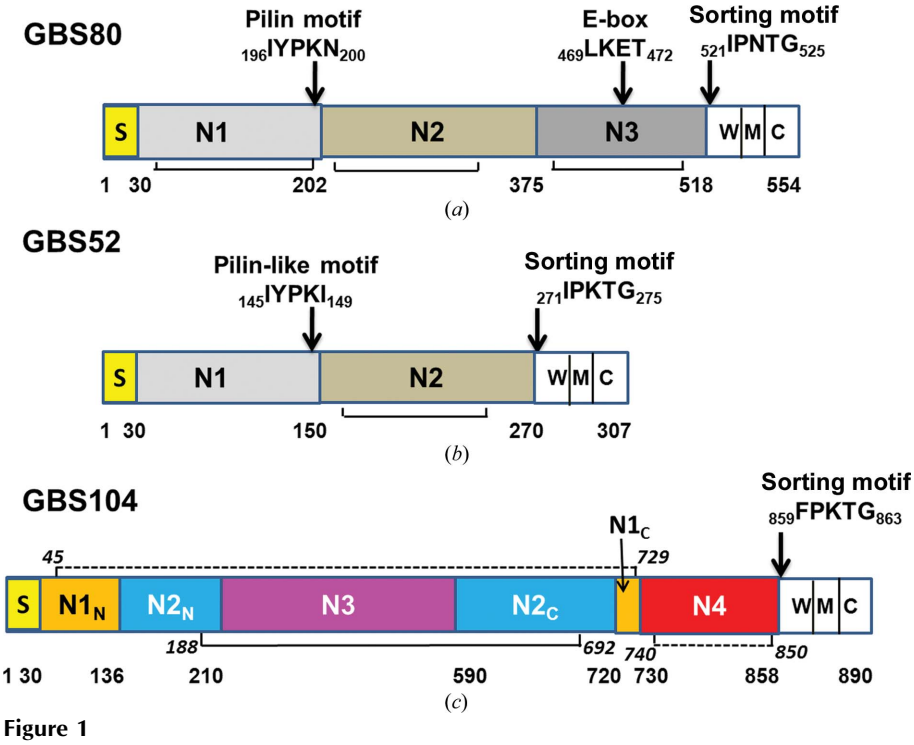

Figure 1
Schematic representation of the group B streptococcus (GBS) pilins. The GBS pilins contain a signal peptide (S) at the N-terminus and a C-terminal cell-wall sorting signal consisting of an LP*X*TG motif (W), a membrane-spanning hydrophobic domain (M) and a cytoplasmic positively charged tail (C) as observed in MSCRAMMs and other Gram-positive pilins. The structural domains between the N- and C-termini are usually made of IgG-like domains. The solid lines at the bottom denote the observed intramolecular isopeptide bonds and the broken lines represent putative isopeptide bonds. (*a*) GBS80 contains three IgG-like domains referred to as N1, N2 and N3. The pilin motif (YPKN) is present near the N1–N2 domain linker. The N3 domain contains an E-box (LK**E**T). (*b*) The minor pilin GBS52 contains two IgG-like domains: N1 and N2. The pilin-like motif IYPKI implicated in cross-linking between GBS80 and GBS52 is present at the C-terminal end of the N1 domain. (*c*) The minor pilin GBS104 contains four domains: three IgG-like domains (N1, N2 and N4) and a vWFA-like domain (N3). The N1 and N2 domains consist of two distinct segments ($N1_N$ and $N1_C$, and N2*N* and $N2_C$) present at the N- and C-termini, respectively. The N2-domain crystal structure exhibits one isopeptide bond and the modeled N1 and N4 domains are predicted to have isopeptide bonds.

(SrtC1 and SrtC2). The PI-1 island of *S. agalactiae* strain 2603V/R encodes one major pilin GBS80 (SAG0645), two minor pilins GBS52 (SAG0646) and GBS104 (SAG0649), and two pilus-specific sortases (SAG0647 and SAG0648). GBS80, GBS52 and GBS104 (Fig. 1) are homologues of the *C. diphtheriae* pilins SpaA, SpaB and SpaC, respectively. According to the current model of GBS pilus assembly (Telford *et al.*, 2006; Vengadesan *et al.*, 2011), the GBS type P-1 pilus shaft is an assembly of major pilin GBS80 arranged in a head-to-tail fashion by a pilus-specific sortase (either SAG0647 or SAG0648). This process occurs *via* linking of the Thr residue in the C-terminal IPNTG sorting motif of one GBS80 to the side chain of a conserved Lys within the pilin YPKN motif of another GBS80. Incorporation of the minor pilin GBS104 at the pilus tip is catalyzed by the sortase SAG0648 through the covalent linking of the Thr in the C-terminal FPKTG motif of GBS104 to the Lys side chain of the pilin motif YPKN of GBS80 (Fig. 1). The incorporation and location of the minor pilin GBS52 have not yet been confirmed. GBS52 is likely to be present at the base of the pili owing to a reaction catalyzed by the housekeeping sortase, similar to the mechanism suggested for the anchoring of *C. diphtheriae* pili (Mandlik *et al.*, 2008). Interestingly, immunoblotting suggests the presence of GBS1474 (an ortholog of GBS52) of strain NEM316 primarily at the pilus base and also randomly distributed along the shaft (Dramsi *et al.*, 2006).

To confirm the role of individual PI-1 products in adherence to host tissues, deletion mutants of GBS strain 2603V/R lacking the individual pilins GBS80, GBS52 and GBS104 were generated and tested for their association with lung epithelial A459 cells (Krishnan, Gaspar *et al.* 2007). Interestingly, deletion of the major pilin GBS80 did not affect bacterial adherence, but deletion of the two minor pilins caused a significant reduction in adherence. It was recognized that GBS52 and GBS104 act independently of each other in adhering to lung epithelial cells. We have previously reported the crystal structures of the minor pilin GBS52 (Krishnan, Gaspar *et al.*, 2007), the major pilin GBS80 (Vengadesan *et al.*, 2011) and two sortases: one pilus-specific (GBSSrtC1) and the other involved in housekeeping (GBSSrtA; Khare *et al.*, 2011).

```
GBS104   1  -------MKKRQKIWRGLSVTLLILSQIP--FGILVQGETQDTNQALGKVIVKKTGDNAT  51
RrgA     1  MLNRETHMKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGA  60
                ***  **       *   **      *    **  *  * ***  * **

GBS104  52  PLGKATFVLKNDNDKSETSHETVEGSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVK 111
RrgA    61  LLGDAVFELKNNTDGTTVSQRTEAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVE 120
             ** * * ***  *    *  *    *** * ***** *** * * ***   * * *

GBS104 112  VADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKENYPLVNVEGSKVGEQYKALNP 171
RrgA   121  VEKNGRTTVQGE---QVENREEALSDQYPQTGTYPDVQTPYQIIKVDGSEKNGQHKALNP 177
             *  ** *   *      * *   *  ***    * *    *    * **    * *****

GBS104 172  INGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVVVLLDNS 231
RrgA   178  N---PYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTVYEQKDKSVPLDVVILLDNS 234
                  * * ** *** *  ** ** * * ***** ***  * *    ***** *****

GBS104 232  NSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQ 291
RrgA   235  NSMSNIRNKNARRAERAGEATRSLIDKITSDSENRVALVTYASTIFDGTEFTVEKGVADK 294
             *** * *  *  **  ****   *******   ***************** ** *****

GBS104 292  NGKALNDSVSWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQF 351
RrgA   295  NGKRLNDSLFWNYDQTSFTTNTKDYSYLKLTNDKNDIVELKNKVPTEAEDHDGNRLMYQF 354
             *** ****  * *  * **  *  **** **** *   **   *  ***   * *  ***

GBS104 352  GATFTQKALMKANEILETQSSNARKKLIFHVTDGVPTMSYAINFNPY-ISTSYQNQFNSF 410
RrgA   355  GATFTQKALMKADEILTQQARQNSQKVIFHITDGVPTMSYPINFNHATFAPSYQNQLNAF 414
             ************ ***  *      * *** ********* ****    ***** * *

GBS104 411  LNKIPDRSGILQEDFIING--DDYQIVKGDGESFKLFSDRKVPVTGGTTQAAYRVPQNQL 468
RrgA   415  FSKSPNKDGILLSDFITQATSGEHTIVRGDGQSYQMFTDKTVYEKGAP--AAFPVKPEKY 472
               * *   ***  ***          ** *** *   * *  *   *    **  *

GBS104 469  SVMSNEGY-----AINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGN 523
RrgA   473  SEMKAAGYAVIGDPINGGYIWLNWRESILAYPFNSNTA------KITNHGDPTRWYYNGN 526
             * *   **       ***** * **     ***   *        *  ** **  * ***

GBS104 524  IRPKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTI 583
RrgA   527  IAPDGYDVFTVGIGINGDPGTDEATATSFMQSISSKPENYTNVTDTTKILEQLNRYFHTI 586
             * * *** ****** *****     *  ******** ****** ** **   ** ** **

GBS104 584  VEEKHSIVDGNVTDPMGEMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGI 643
RrgA   587  VTEKKSIENGTITDPMGELIDLQLGTDGRFDPADYTLTANDGSRLENGQAVGGPQNDGGL 646
             * ** **  *  ****** *  **     *   ** *  **** * ** * ***  ***

GBS104 644  LKDVTVTYDKTSQTIKINHLNLGSGQKVVLTYDVRLKDNYISNKFYNTNNRTTLSPKSEK 703
RrgA   647  LKNAKVLYDTTEKRIRVTGLYLGTDEKVTLTYNVRLNDEFVSNKFYDTNGRTTLHPKE-V 705
             **   * ** *   *    * **   ** *** *** *   ***** ** **** **

GBS104 704  EPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESLLGAKFQLQIEK-D 762
RrgA   706  EQNTVRDFPIPKIRDVRKYPEITISKEKKLGDIEFIKVNKNDK-KPLRGAVFSLQKQHPD 764
             * ** ************  *  ***  ** *  *******      * ** * **    *

GBS104 763  F---SGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQ 819
RrgA   765  YPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIV 824
                 *          * *  ***  ** * ** * * * * * **  *  ** * * *

GBS104 820  NGEVTNLKADPNANKNQIGYLEGNGKHLITNTPKRPPGVFPKTGGIGTIVYILVGSTFMI 879
RrgA   825  NGEVRDVTSIV-PQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMG 883
             ****              **   * ** *** *  *    * *****     * *   *

GBS104 880  LTICSFRRKQL 890
RrgA   884  GVLLYTRKHP- 893
                  *
```

Figure 2
Primary sequence alignment of the tip pilins GBS104 and *S. pneumoniae* RrgA. Sequence segments representing individual domains are colored (red for the N1 domain, blue for the N2 domain, magenta for the N3 domain and green for the N4 domain) and residues that are identical in the two proteins are marked with an asterisk below the alignment.

The structures of several pilins of Gram-positive pathogens, including the major pilins Spy0128 of *S. pyogenes* (Kang *et al.*, 2007), SpaA of *C. diphtheriae* (Kang, Paterson *et al.*, 2009), BcpA of *Bacillus cereus* (Budzik *et al.*, 2009), RrgB of *S. pneumoniae* (Spraggon *et al.*, 2010), FimA of *A. naeslundii* (Mishra *et al.*, 2011) and the minor pilins RrgA of *S. pneumoniae* (Izoré *et al.*, 2010) and FctB of *B. cereus* (Linke *et al.*, 2010), are also available. The pilins usually consist of one to four domains, mostly IgG-like (Krishnan & Narayana, 2011), and are stabilized by intramolecular isopeptide bonds (Kang *et al.*, 2007; Kang, Middleditch *et al.*, 2009). Interestingly, the structure of the tip pilin RrgA of *S. pneumoniae*, which is homologous to GBS104 (51% sequence identity; Fig. 2) revealed the presence of one vWFA-like domain (von Willebrand factor A-like domain; approximately 200 residues in length and mediating various protein–protein interactions) along with three IgG-like domains (Izoré *et al.*, 2010).

In this report, we present the crystal structures of two GBS104 fragments consisting of the central N3 domain (GBS104N3) and of the N2 and N3 domains (GBS104N2N3). We identified the core GBS104N3 fragment to be composed of a vWFA-like domain similar to that observed in RrgA of *S. pneumoniae* (Izoré *et al.*, 2010) and some eukaryotic protein families such as integrins and complement proteins (Krishnan, Xu *et al.*, 2007; Whittaker & Hynes, 2002). The vWFA-like domains are adhesive in function, specifically through a metal-ion-dependent adhesion site (MIDAS). The C-terminal α7 helix and the MIDAS of vWFA-like domains (also known as I domains in integrins) can adopt two conformations: (i) a closed/low-affinity state and (ii) an open/high-affinity state (Zhang, Liu *et al.*, 2009; Luo *et al.*, 2004, 2007). Conformational changes involving an allosteric coupling of the MIDAS configuration and a C-terminal α7-helix positional shift constitute an essential mechanism for the regulation of the ligand affinity of these adhesive proteins. By locking the I domain into the open/high-affinity or closed/low-affinity conformations with disulfides, Luo and coworkers were able to explore both the open and closed conformations of β3 integrin and the role of structural differences in their ligand affinity (Luo *et al.*, 2004). Since the vWFA-like N3 domain of the GBS104N3 and GBS104N2N3 crystal structures is observed in the closed/low-affinity conformation, we generated recombinant disulfide-locked GBS104 N3 domain (GBS104N3mut), determined its crystal structure and investigated the effect of the open MIDAS conformation on the affinity for extracellular matrix (ECM) proteins such as fibrinogen, fibronectin, collagen and laminin.

We built a homology model of full-length GBS104, which is an elongated spear-like structure with the adhesive N3 domain at the tip and the N4 domain at the other end. The N3, N2 and modeled N1 domains of GBS104 exhibit structural differences from the corresponding domains of the homologous *S. pneumoniae* minor pilin RrgA (Izoré *et al.*, 2010). Significantly, GBS104 and RrgA carry a reverse RGD motif, a molecular switch that is implicated in integrin–ligand recognition (Corti & Curnis, 2011), on two different faces of their respective N2 domains.

2. Materials and methods

2.1. Expression and purification of recombinant proteins

The expressed recombinant GBS104 protein (residues 29–854; rGBS104$_{29–854}$) containing an N-terminal His tag but lacking the N-terminal signal peptide and C-terminal sorting motif was soluble, but was not produced in sufficient amounts for crystallization attempts. However, the expression of a construct consisting of residues 212–863 (rGBS104$_{212–863}$) resulted in a yield suitable for crystallization. rGBS104$_{212–863}$ was cloned into a pMCSG7 vector with an N-terminal TEV-cleavable His$_6$ tag and was expressed in *Escherichia coli* BL21(DE3) cells (Novagen). Cells were grown in 1 l Luria–Bertani (LB) broth containing 100 µg ml^{-1} ampicillin at 310 K until the OD$_{600}$ reached 0.7. The culture temperature was adjusted to 295 K, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 m*M* and the culture was allowed to grow overnight. Selenomethionyl (SeMet) rGBS104$_{212–863}$ was expressed in *E. coli* BL21(DE3) cells, which were grown in 1 l M9 SeMet growth medium (Medicilon Inc.) containing 100 µg ml^{-1} ampicillin at 310 K until the OD$_{600}$ reached 1.2. After decreasing the temperature to 289 K, IPTG was added to a final concentration of 1 m*M* and the culture was allowed to grow overnight. Both native and SeMet rGBS104$_{212–863}$ were purified using similar protocols. The bacterial cells were harvested by centrifugation (2568*g* for 20 min at 277 K) using a Beckman Allegra 6R centrifuge and resuspended in lysis buffer consisting of 50 m*M* Tris buffer pH 8.0, 300 m*M* NaCl, 1 m*M* phenylmethylsulfonyl fluoride (PMSF), 10% glycerol. The cells were then lysed by ultrasonication. The cell lysate was centrifuged (18 384*g* for 30 min at 277 K) and the supernatant was loaded onto a column containing 8 ml Ni–NTA Superflow Agarose resin (Qiagen). The column was washed with ten column volumes of lysis buffer and ten column volumes of buffer *A* (50 m*M* Tris buffer pH 8.0, 100 m*M* NaCl, 0.5 m*M* PMSF) prior to protein loading. The protein was eluted with a linear gradient of 0–100% buffer *B* (50 m*M* Tris pH 8.0, 100 m*M* NaCl, 0.5 m*M* PMSF, 300 m*M* imidazole). Fractions that contained rGBS104$_{212–863}$ (as identified by SDS–PAGE) were pooled and dialyzed against a buffer consisting of 50 m*M* Tris pH 8.0, 100 m*M* NaCl, 1 m*M* EDTA, 1 m*M* dithiothreitol (DTT). The His tag was removed using recombinant His-tagged TEV protease at a ratio of 1:100 at 277 K for 14 h. The mixture was dialyzed against 50 m*M* Tris buffer pH 8.0, 100 m*M* NaCl for 5 h at 277 K. The protein was reloaded onto an Ni–NTA column and was collected in the flowthrough fraction. After concentration, the protein was purified by size-exclusion chromatography on a Superdex 200 column (GE Healthcare) in 20 m*M* Tris buffer pH 8.0, 100 m*M* NaCl. The fractions containing rGBS104$_{212–863}$ were pooled and concentrated to 10 mg ml^{-1} by ultrafiltration using a 30 kDa membrane (Amicon).

Two Cys residues were substituted for Thr and Lys at positions 564 and 571 of the vWFA-like N3 domain in the rGBS104$_{212–581}$ construct using the QuikChange Mutagenesis Kit (Stratagene, La Jolla, California, USA). The expression

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost resolution shell.

	GBS104N3 (native)	GBS104N3 (SeMet)	GBS104N3mut (Cys mutant)	GBS104N2N3 (native)
Beamline	In-house	NE-CAT 24-ID	In-house	NE-CAT 24-ID
Detector	R-AXIS IV IP	ADSC 315 CCD	R-AXIS IV IP	ADSC 315 CCD
Wavelength (Å)	1.5418	0.97918	1.5418	0.97949
Space group	$P2_12_12$	$P2_12_12$	$P1$	$P3_121$
Unit-cell parameters				
a (Å)	74.8	75.4	53.1	124.6
b (Å)	101.3	101.1	77.3	124.6
c (Å)	52.7	52.7	96.4	117.8
α (°)	90.0	90.0	74.1	90.0
β (°)	90.0	90.0	87.3	90.0
γ (°)	90.0	90.0	90.0	120.0
Resolution (Å)	35.1–2.0 (2.07–2.00)	40.0–2.0 (2.07–2.00)	46.3–2.0 (2.07–2.00)	39.8–2.6 (2.67–2.60)
$R_{\rm merge}$† (%)	11.9 (44.8)	10.5 (66.5)	7.0 (29.0)	9.6 (67.0)
Average multiplicity	5.8 (5.2)	4.2 (3.5)	3.6 (3.7)	7.3 (5.9)
Unique reflections	27094	27855	88454	32251
Completeness (%)	97.7 (94.4)	100.0 (100.0)	89.0 (80.2)	100.0 (99.4)
χ^2	1.0 (1.2)	1.3 (1.2)	1.0 (1.1)	1.2 (0.7)
$\langle I/\sigma(I)\rangle$	9.3 (3.1)	15.6 (2.1)	11.5 (4.2)	21.9 (2.0)
No. of molecules in the asymmetric unit	1	1	4	1
$R_{\rm work}/R_{\rm free}$ (%)	21.0/25.1	18.6/22.3	19.8/23.1	20.6/25.0
Average B value (Å^2)	24.2	27.4	24.9	56.6
R.m.s.d. for bonds (Å)	0.014	0.014	0.014	0.016
R.m.s.d. for angles (°)	1.468	1.430	1.465	1.486
No. of protein atoms	2896	2922	11380	4610
No. of metal/water molecules	1/220	1/237	4/759	9/79
Ramachandran plot, residues in (%)				
Favored regions	88.4	89.3	87.9	85.1
Allowed regions	11.6	10.7	12.1	14.5
Disallowed regions	0.0	0.0	0.0	0.4
PDB code	3tvy	—	3twd	3txa

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ are the intensities of symmetry-related reflections and $\langle I(hkl)\rangle$ is the average intensity over all observations.

and purification of wild-type and Cys-mutant rGBS104$_{212–581}$, rGBS104$_{48–863}$ and rGBS104$_{48–729}$ were performed using a similar protocol to that used for rGBS104$_{212–863}$. However, hydrophobic interaction chromatography was required to remove contaminants in the final stage of the purification of rGBS104$_{48–729}$. After the gel-filtration step, the protein in a solution of high ionic strength (50 m*M* potassium phosphate buffer with 1.5 *M* ammonium sulfate pH 7.0; buffer *A*) was applied onto a Phenyl Superose HR 5/5 column (Pharmacia) equilibrated at the same ionic strength. The protein was eluted using a linear gradient to a solution of low ionic strength (0–100% buffer *B*; 50 m*M* potassium phosphate buffer pH 7.0).

2.2. Crystallization and data collection

Preliminary crystallization trials for all of the purified recombinant proteins were carried out by the sitting-drop vapor-diffusion method using a Phoenix robot (Art Robbins Instruments) at 295 and 277 K. In addition, manual screening by the hanging-drop vapor-diffusion method using various commercial crystallization kits (Wizard I and II from Emerald BioSystems, Crystal Screen and Crystal Screen 2 from Hampton Research and The PEGs and PEGs II Suites from Qiagen) was performed. All recombinant proteins were stored in 20 m*M* Tris buffer pH 8.0, 100 m*M* NaCl for crystallization trials. During the final optimization, 1 µl protein solution and 1 µl reservoir solution were mixed and equilibrated against 1 ml reservoir solution.

Crystals of native GBS104N3 (rGBS104$_{212–581}$) and of its SeMet derivative appeared in various conditions after 5–15 d, but only the rod-shaped crystals obtained using 0.1 *M* HEPES pH 7.5, 20% PEG 8000 with 10 mg ml^{-1} protein solution were suitable for data collection. Crystals of GBS104N3mut (the Cys mutant of rGBS104$_{212–581}$) were obtained using 0.1 *M* HEPES pH 7.5, 20% PEG 3000, 0.2 *M* NaCl, 10 m*M* spermine tetrahydrochloride with 24 mg ml^{-1} protein solution. Crystals of GBS104N2N3 (rGBS104$_{140–729}$, a breakdown product of rGBS104$_{48–729}$) were obtained using 0.1 *M* Tris pH 8.0, 15% PEG 4000, 0.2 *M* lithium sulfate, 10 m*M* $CdCl_2$ with 10 mg ml^{-1} protein solution.

Diffraction data sets for the GBS104N3 and GBS104N3mut crystals were collected at the home-source facility using an R-AXIS IV image-plate detector mounted on a Rigaku rotating-anode X-ray generator operating at 100 mA and 50 kV with 25% ethylene glycol (EG) as a cryoprotectant. The GBS104N3 crystals belonged to space group $P2_12_12$ with one molecule in the asymmetric unit; the GBS104N3mut crystals belonged to space group $P1$ with four molecules in the asymmetric unit. Single-wavelength anomalous dispersion (SAD) data for SeMet GBS104N3 crystals and native data for GBS104N2N3 crystals in space group $P3_121$ were collected on the NE-CAT 24ID synchrotron beamline at the Advanced Photon Source (APS), Chicago, USA using 20% ethylene glycol and PEG 400 as cryoprotectants, respectively. The data collected using the in-house X-ray source were processed with *d*TREK* (Pflugrath, 1999) and the data collected at the synchrotron were processed using *HKL*-2000 (Otwinowski & Minor, 1997). Relevant diffraction data-collection and processing details are presented in Table 1.

2.3. Structure determination, model building and refinement

The GBS104N3 structure was determined by SAD phasing using SeMet GBS104N3 crystals. Phase calculation and initial model building were carried out using *PHENIX* (Adams *et al.*, 2010). Five selenium sites were used in the *PHENIX* phase

calculation (FOM of 0.33) and the initial model consisted of 400 residues with R_{work} and R_{free} values of 39.0 and 42.0%, respectively. Final model building and refinement were carried out using *Coot* (Emsley *et al.*, 2010; Emsley & Cowtan, 2004) and *REFMAC* (Murshudov *et al.*, 2011) from the *CCP*4 suite (Winn *et al.*, 2011). The SeMet GBS104 structure was used in the refinement of the native GBS104N3 structure. The quality of the final models was examined in *Coot* and checked with *PROCHECK* (Laskowski *et al.*, 1993). The GBS104N3mut structure was solved by molecular replacement using the GBS104N3 structure as a search model.

The GBS104N2N3 structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007). Firstly, the GBS104N3 domain was placed, and the N2 domain was then modeled based on the difference electron-density map and the homologous N2 domain of the RrgA crystal structure (PDB entry 2ww8; Izore *et al.*, 2010). Metal atoms and water molecules were added during the final stages of refinement using *REFMAC* from the *CCP*4 suite (Winn *et al.*, 2011). We used the default values for *B*-factor restraints in refinement with *REFMAC*: 1.5, 3.0 and 4.5 for main-chain bonds, side-chain bonds and side-chain angles, respectively. The program *PROCHECK* (Laskowski *et al.*, 1993) was used to analyze the stereochemical quality of the crystal structure, and the Ramachandran plots for GBS104N3 and GBS104N2N3 were of good quality (Table 1).

Full-length *S. pneumoniae* RrgA exhibits 51% primary sequence identity to *S. agalactiae* GB104 overall, and 47 and 50% sequence identity between the respective N1 (108 residues) and N4 (270 residues) domains (Fig. 2). We built models of the GBS104 N1 and N4 domains using the corresponding domains of RrgA (PDB entry 2ww8) as backbone templates with the help of the *FFAS* server (Zhang, Thiele *et al.*, 2009) and *MODELLER* (Martí-Renom *et al.*, 2000) (r.m.s.d.s of 1.0 and 1.1 Å for N4 and N1, respectively, between the two models). The full-length GBS104 model was built with the help of the *FFAS* server and *Coot* (Emsley *et al.*, 2010) by combining the N2 + N3 crystal structure and the models of the N1 and N4 domains and was idealized using *REFMAC* (Murshudov *et al.*, 2011) while monitoring the quality with the help of a Ramachandran plot.

2.4. Immunofluorescence microscopy

GBS must cross the human blood–brain barrier (BBB) to gain access to the central nervous system (CNS). The BBB is primarily composed of a single layer of brain microvascular endothelial cells (hBMECs). Isolated and cultured hBMECs have been used to identify virulence genes in various bacterial species that promote cellular invasion. We treated hBMECs with $rGBS104_{29–854}$ and its fragments fixed with 4% paraformaldehyde and stained with anti-PilA antiserum (Banerjee *et al.*, 2011) and Alexa Fluor 488 conjugated goat anti-rabbit mAb (Invitrogen, Carlsbad, USA). DAPI (4,6-diamidino-2-phenylindole) was used to stain DNA. Cover slips were mounted on glass slides using Vectashield (Vector Labs, Burlingame, USA) and visualized using a confocal laser scanning microscope (Model No. DMIRE2; Leica Microsystems).

2.5. Binding of recombinant GBS104 proteins to host ECM proteins

To assess collagen binding, ELISA plates coated with collagen I (BD Biosciences) or BSA (2 µg per well) were blocked with PBS/1% BSA and treated with various concentrations of purified $rGBS104_{48–863}$ and its fragments diluted in PBS. After subsequent washing and incubation with anti-His Ab, the ELISA plate was treated with a secondary Ab conjugated to horseradish peroxidase (HRP). The chromophore was developed using 3,3′,5,5′-tetramethylbenzidine (Sigma, St Louis, USA) and the absorbance was measured at 450 nm.

The solid-phase binding assay used by Mishra *et al.* (2011) was used to determine the binding properties of recombinant $rGBS104_{48–863}$ and its fragments to the extracellular matrix proteins (Sigma–Aldrich). Briefly, 96-well high-binding plates were coated with 100 µl fibronectin (50 µg ml^{-1}), fibrinogen (10 µg ml^{-1}), laminin (50 µg ml^{-1}) and collagen type I (50 µg ml^{-1}), centrifuged at 2000 rev min^{-1} for 5 min and incubated for 2 h at 300 K. Samples were blocked with 5% milk for 1 h at room temperature and were washed twice with phosphate-buffered saline (PBS) containing 0.2% Tween-20 (PBST). Recombinant $rGBS104_{48–863}$ and its fragments were added at a concentration of 5 µ*M*. The plates were centrifuged and incubated at 310 K for 1 h followed by washing with PBST and incubation with anti-penta-His-HRP labeled antibodies. After 1 h incubation at room temperature, the wells were washed three times with PBST and incubated with tetramethylbenzidine for 30 min. The reactions were stopped using sulfuric acid and read at OD_{450}. All assays were performed in triplicate in two independent experiments.

2.6. Examination of IL-8 induction

Concentrations of IL-8 in hBMEC supernatants collected 4 h post-incubation with PilA, $rGBS104_{48–863}$ or its fragment and mutant constructs were measured using enzyme-linked immunosorbent assays (ELISA) according to the manufacturer's instructions (R&D Systems, Minneapolis, USA).

3. Results

3.1. Structure of GBS104 fragments determined by X-ray crystallography

The tip pilin GBS104 consists of 890 residues (Fig. 1*c*). Recombinant protein (residues 29–854; $rGBS104_{29–854}$) composed of N1, N2, N3 and N4 domains and containing an N-terminal His tag and lacking the N-terminal signal peptide and the C-terminal sorting motif was expressed in *E. coli* and purified using chromatographic techniques. However, as the yield of $rGBS104_{29–854}$ was not sufficient for crystallization trials, a construct containing residues 212–854 ($rGBS104_{212–854}$) was expressed and used in crystallization trials. Thin plate-like crystals belonging to space group $P2_12_12$

which diffracted to 2.0 Å resolution (Table 1) were obtained in four weeks. The unit-cell volume and solvent-content calculations suggested the presence of an approximately 40 kDa peptide instead of the expected 72 kDa in the asymmetric unit. SDS–PAGE analysis of the protein used for crystallization and the dissolved crystals revealed that the rGBS104$_{212–854}$ protein had degraded into a stable fragment of molecular weight ∼40 kDa during crystallization. The crystal structure of this ∼40 kDa fragment was solved by the single-wavelength anomalous dispersion (SAD) method using diffraction data collected from an SeMet crystal on the NE-CAT 24ID synchrotron beamline (Table 1). The final model consisted of residues 220–584, corresponding to the N3 domain of GBS104 (Figs. 1*c* and 2). The model was refined using 2.0 Å resolution diffraction data collected using an in-house Rigaku rotating-anode generator and an R-AXIS IV imaging-plate detector (Table 1).

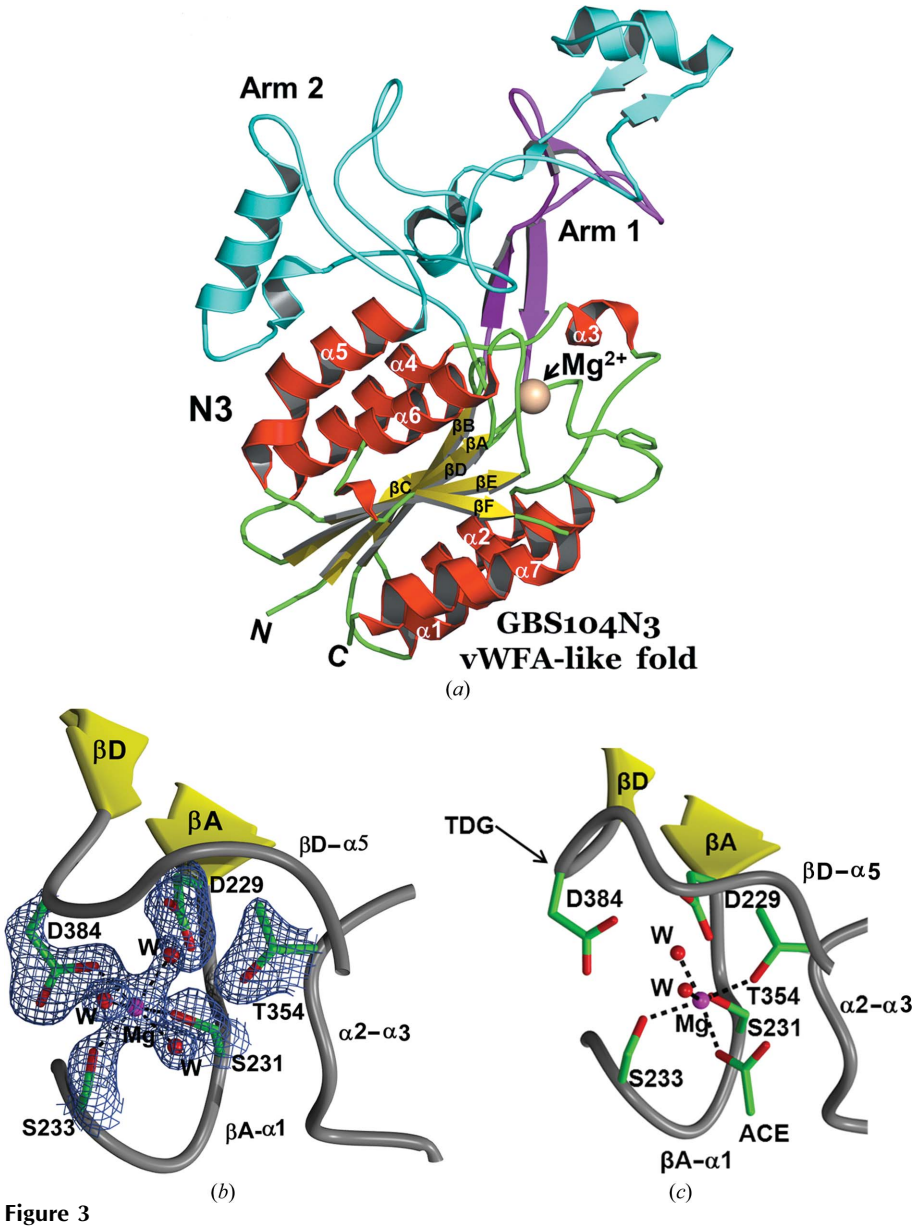

Figure 3
(*a*) Ribbon representation of the GBS104 N3 domain, which exhibits a vWFA-like domain fold. The six β-strands βA–βF, represented in yellow, of the central β-sheet are surrounded by seven (α1–α7) amphipathic α-helices, which are depicted in red. Two extended arms near the metal-ion-dependent adhesion site (MIDAS) of the N3 domain are labeled Arm 1 (magenta) and Arm 2 (cyan). The magnesium ion (Mg^{2+}) present in the MIDAS is shown in ivory. The MIDAS of GBS104 N3 is shown in two conformations: (*b*) a closed/low-affinity state in the wild-type structure and (*c*) an open/high-affinity state in the mutant structure. The magnesium ions are shown as pink spheres. The residues involved in metal coordination are represented by sticks and the ionic bonds to the metal are illustrated as broken lines. In the closed conformation (*b*) the magnesium ion is coordinated by six O atoms from the side chains of Ser231, Ser233 and Asp384 and three water molecules. The electron density displayed around the MIDAS is from an ($F_o - F_c$) simulated-annealing OMIT electron-density map computed from a model (GBS104 N3) without the side chains of the MIDAS residues and is shown contoured at 3.0σ. In the open conformation (*c*) generated by locking down the α7 helix using an engineered disulfide bridge the magnesium ion is coordinated by Ser231, Ser233 and Thr354, with two water molecules and an acetate ion from the solvent completing the coordination of the Mg^{2+} ion. The Thr354 moves in to directly coordinate to Mg^{2+} and Asp384 moves out, losing its direct coordination and interacting with the metal through one water molecule. The displacement of Asp384 in the TDG motif is marked by an arrow.

3.2. The structure of the vWFA-like (N3) domain

The overall structure of GBS104N3 adopts a Rossmann or dinucleotide-binding fold similar to that observed in classical vWFA and integrin I domains, with a six-stranded β-sheet hydrophobic core surrounded by seven amphipathic α-helices (Fig. 3*a*). The helices and strands are labeled α1–7 and βA–F, similar to the structurally homologous $\alpha M\beta_2$ I domain (∼22% primary sequence identity; Lee, Rieu *et al.*, 1995; Plow *et al.*, 2000) and human complement C2 and factor B central domains (22 and 18% primary-sequence identity, respectively; Krishnan, Xu *et al.*, 2007; Ponnuraj *et al.*, 2004). The central β-sheet contains five parallel β-strands and one short antiparallel strand (βC) at one edge in the following order: βC–βB–βA–βD–βE–βF. The short antiparallel strand βC is distorted where an extended loop (the 'first arm') is inserted (described below). The N- and C-termini are located close to each other, as observed in a typical vWFA domain (Fig. 3*a*).

Interestingly, the N3 domain of GBS104 displays two 'inserted arms' (Fig. 3*a*) that are not observed in the abovementioned eukaryotic vWFA domains (the $\alpha M\beta_2$ I domain and the central vWFA-like domains of complement factor B and C2), and both 'arms'

exhibit well defined electron density except for a small loop (residues 453–457). These two segments, composed of nearly 175 residues, increase the overall dimensions of the N3 domain to 70 × 55 × 40 Å compared with the 40 × 40 × 30 Å typical of vWFA domains. The first arm (277–315), which is an elongated region with two β-hairpins, consists of 39 residues and is inserted as a linking region between βB and βC. The longer second arm (389–513) inserted between βD and the α5 helix is composed of one short hairpin, two three-turn helices and several flexible loops. These two inserted arms are stretched away from the MIDAS, which is present on a face with a crevice that is suitable for metal binding at the head of the central β-sheet. However, there are large buried interfaces between the first arm and the core (buried surface area of 1522 Å^2), between the second arm and the core (2112 Å^2), and between the first and second arms (1720 Å^2).

3.3. The MIDAS motif

The MIDAS motif in GBS104N3 is composed of residues Asp229-*X*-Ser231-*X*-Ser233 in the βA–α1 loop, Thr354 in the α2–α3 loop and Asp384 in the βD–α5 loop (Fig. 3*b*). This MIDAS arrangement is similar to that observed in the vWFA and I domains, in which it plays a key role in the recognition of ECM proteins such as collagen, fibrinogen and fibronectin (Plow *et al.*, 2000; Loftus *et al.*, 1994). The electron-density maps calculated using SeMet phases showed the presence of density higher than 7σ for the MIDAS, even though no metal ions were included in the purification and crystallization of the recombinant protein. However, the density for two of the three water molecules coordinating to this position is weak. Diffraction data collected from crystals soaked in 5 m*M* $MgCl_2$ revealed the presence of Mg^{2+} (Table 1) with complete octahedral coordination (average bond length of 2.1 Å), and the three water molecules and MIDAS residues had better density with no significant positional or conformational changes.

In the GBS104N3 crystal structure the residues Ser231, Ser233 and Asp384 make direct ionic bonds to the Mg^{2+} ion through their hydroxyl O atoms, while the side chains of Thr354 and Asp229 make water-mediated contacts. The arrangement of the MIDAS residues and the backbone conformation of Gly385 (confirmed by TDG dihedral angle analysis; Table 2) clearly indicate that the vWFA-like domain of GBS104N3 is in the ligand-free closed conformation (Fig. 3*b*).

Table 2
Comparison of the backbone torsion angles of the TDG motif.

	T		D		G	
	φ (°)	ψ (°)	φ (°)	ψ (°)	φ (°)	ψ (°)
GBS104	−149.3	157.0	−121.0	19.1	−167.6	153.4
GBS104 (Mg^{2+})	−143.6	155.0	−125.2	20.5	−162.3	146.3
RrgA (2ww8)	−143.7	158.1	−120.4	21.4	−171.0	153.5
CR3 open (1ido)	−153.4	169.8	−94.5	3.6	73.8	27.7
CR3 closed (1jlm)	−144.3	150.6	−115.4	11.7	−162.1	158.2
α2 open (1dzi)	−160.6	167.5	−93.6	4.4	74.5	12.3
α2 closed (1aox)	−140.3	154.8	−122.6	22.2	−159.2	152.3
GBS104N3mut	−167.5	162.8	−95.9	2.8	79.1	24.8

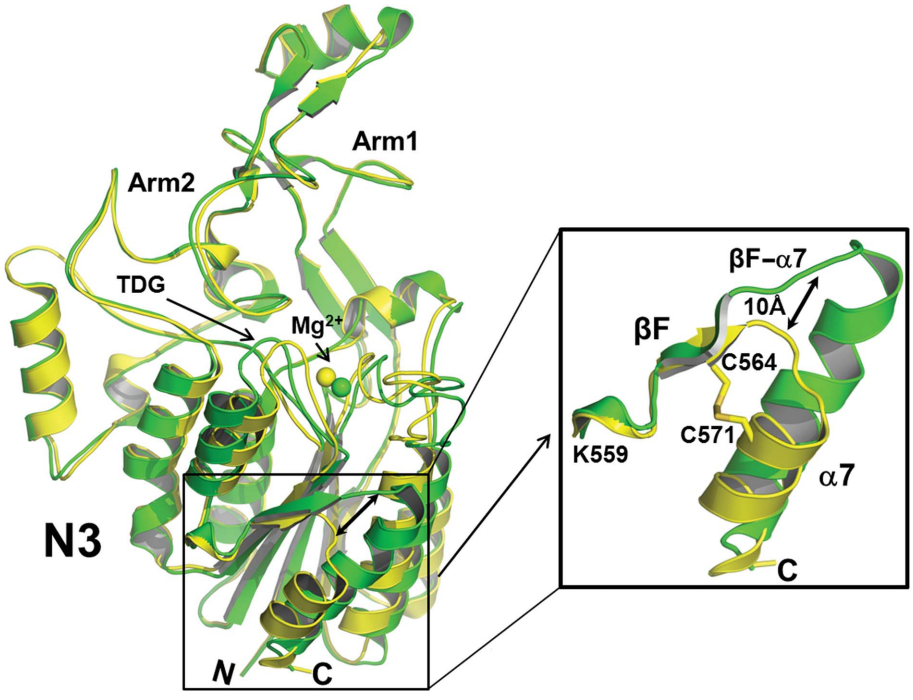

Figure 4
Comparison of the wild-type GBS104N3 structure with that of its engineered Cys mutant. (*a*) Superposition of the GBS104N3mut (yellow) and the GBS104N3wt (green) structures. The major structural differences (α7 helix shift and TDG displacement) are marked by arrows. (*b*) Enlarged view of the shift of the βF–α7 loop and α7 helix away from the MIDAS by ∼10 Å between the open (yellow) and closed (green) conformations.

3.4. Mutations to generate an open conformation of the vWFA-like (N3) domain

Large conformational changes were observed in the integrin I domains upon ligand binding (Lee, Bankston *et al.*, 1995), and numerous crystal structures with open (ligand-bound) and closed (ligand-free) conformations are available in the PDB. The transition from closed/low affinity to open/high affinity is thought to increase the electrophilicity of the MIDAS and to increase its affinity towards the ligand (Plow *et al.*, 2000).

The MIDAS-mediated protein–protein interactions are best understood for the αMβ2 integrin I domain (Shimaoka, Lu *et al.*, 2002). The main structural difference between the two conformations (ligand-free closed and ligand-bound open) is found at the C-terminal α-helix, which undergoes a 10 Å shift away from the ligand-bound MIDAS. The α7-helix shift, coupled with changes in the Mg^{2+}-binding site

and a 180° backbone flip of a conserved Gly (in βD) are the hallmarks of the transition between the closed and open states of vWFA/I domains. Similar changes have been proposed for MIDAS-mediated C2 and factor B binding to C4b and C3b, respectively, during C3 convertase formation in the human complement cascade (Janssen *et al.*, 2009; Krishnan, Xu *et al.*, 2007; Ponnuraj *et al.*, 2004; Rooijakkers *et al.*, 2009). The ligand-bound conformational changes in the I domain could be reproduced with the use of engineered disulfide bonds (Lu, Shimaoka, Ferzly *et al.*, 2001). The open, intermediate and closed conformations of many I domains have been generated by linking the C-terminal α7 helix and the preceding βF strand with the help of an introduced disulfide bond in order to understand the structural rearrangement between the low-affinity and the high-affinity states (Lu, Shimaoka, Zang *et al.*, 2001; McCleverty & Liddington, 2003; Ponnuraj *et al.*, 2004; Shimaoka *et al.*, 2001, 2003; Shimaoka, Lu *et al.*, 2002; Shimaoka, Tagaki *et al.*, 2002). To investigate whether a similar mechanism-based adhesive affinity is possible for the GBS104 vWFA-like N3 domain towards ECM proteins (as discussed below), we mutated Thr564 and Lys571 to Cys residues in the N3 domain (GBS104N3mut). The crystals of this mutant, which were produced with $CaCl_2$ in the crystallization conditions, diffracted to 2.0 Å resolution on the NE-CAT 24-ID beamline at APS and belonged to space group *C*2 with two molecules in the asymmetric unit. Another crystal form generated using spermine tetrahydrochloride as an additive diffracted to 2.0 Å resolution at the home-source facility and belonged to space group *P*1 with four molecules in the asymmetric unit (Table 1). We used the latter crystal structure in the following discussion because its model quality was better than that of the space-group *C*2 crystals.

In the GBS104N3mut structure, the four molecules in the asymmetric unit are tightly packed with their extended 'arms' directed toward the center of the four molecules. Molecules *A* and *B* are related by noncrystallographic twofold symmetry, as are molecules *C* and *D*. We refer to the well refined molecule *A* in the following discussion of the GBS104N3mut structure. Its overall structure is similar to that of the wild type, with the following differences: rearrangements in the MIDAS, of the

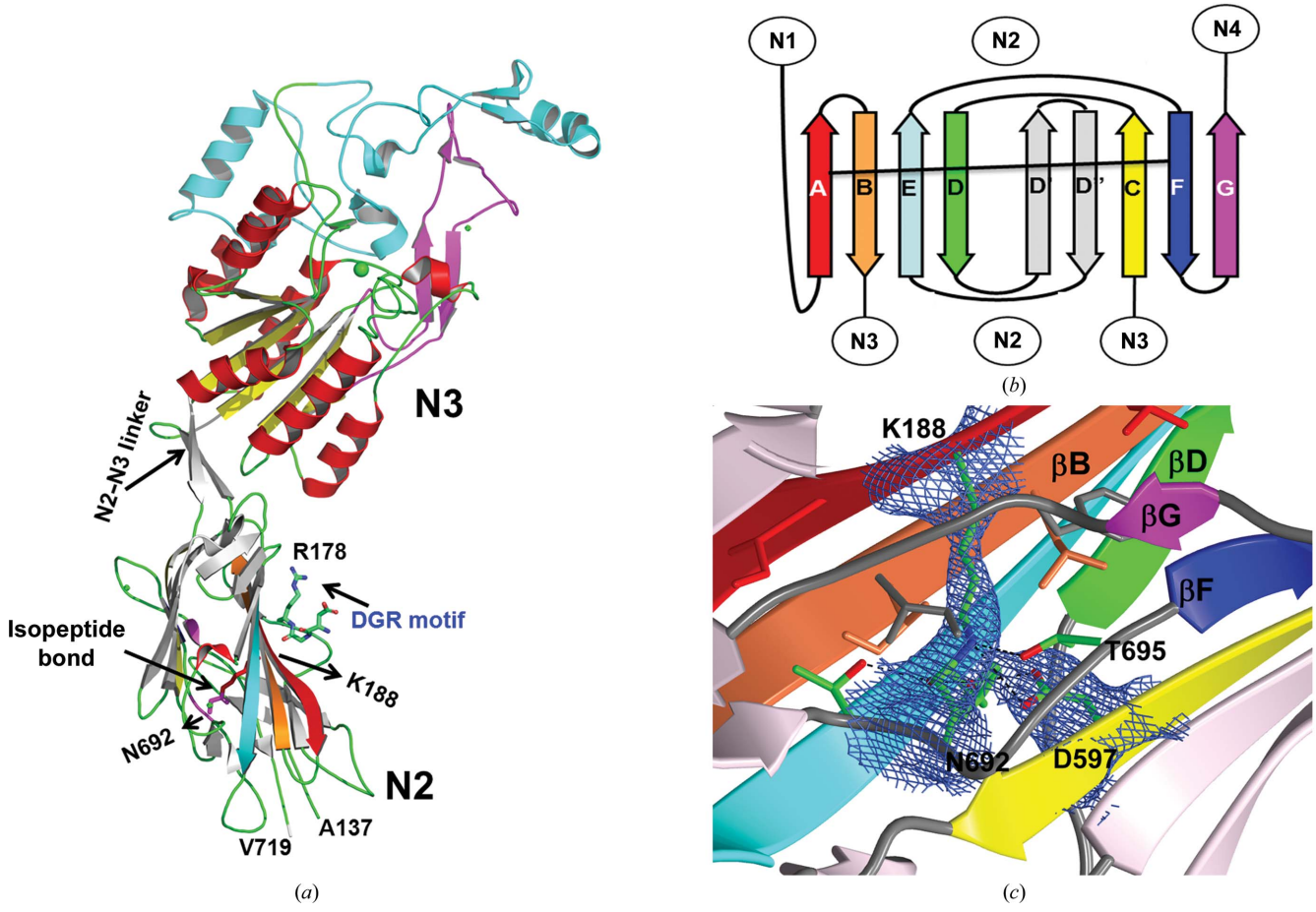

Figure 5
(*a*) Ribbon representation of the GBS104N2N3 structure. The secondary-structural elements in the N3 domain are colored as in Fig. 3(*a*). The β-strands of the N2-domain CnaA-type fold are colored in rainbow style from red to violet. Residues of the isopeptide bond and DGR motif are shown as sticks and marked by arrows. (*b*) Topology diagram of core β-strands in the CnaA-type fold of the N2 domain colored as in (*a*). The horizontal solid line shows the approximate position of the isopeptide bond. (*c*) The isopeptide bond in the N2 domain, the conserved acidic residue and the surrounding hydrophobic residues are shown using sticks. The electron density displayed around the isopeptide bond is from an ($F_o - F_c$) OMIT electron-density map computed from a model (GBS104N2N3) without the side chains of the residues involved in the isopeptide bond and is shown contoured at 3.0σ.

position of the C-terminal α7-helix position and in the TDG motif (Table 2; Fig. 4*a* and 4*b*). The differences between the wild-type and mutant structures around the MIDAS suggest that the mutant is in the open conformation, as anticipated. The Mg^{2+} ion in GBS104N3mut is directly coordinated by the O atoms of Ser231, Ser233 and Thr354 and is indirectly coordinated by Asp384 and Asp229 through water molecules. The sixth MIDAS coordination site is occupied by a pseudo-ligand (Fig. 3*c*), the electron-density shape of which resembles an acetate ion. The Mg^{2+} ion is shifted towards Thr354 by approximately 2.5 Å.

One hydrophobic Ile316 residue present at the α7-helix C-terminal end of the αM I domain was observed to be buried in a highly conserved hydrophobic pocket and is displaced by Leu312 during the transition from the closed/low-affinity to the open/high-affinity state (Xiong *et al.*, 2000). The similarly buried Ile583 in GBS104N3 is observed to be displaced by Tyr579 in the GBS104N3mut structure. Moreover, a conserved Phe302 residue at the N-terminal end of the α7 helix in the αM integrin I domain (closed), near its MIDAS, is solvent-exposed upon conversion to the open conformation (Lee, Bankston *et al.*, 1995). A similar positional shift is also observed for Phe275 of the βE–α6 loop in the transition to the open conformation of the αM integrin I domain. Residues Thr569 and Val538 of GBS104 N3 occupy the exact positions of residues Phe302 and Phe275 in the αM integrin I-domain structure and undergo similar conformational changes in the transition from a closed to an open state.

3.5. The N2-domain structure

While we were completing the GBS104N3mut crystal structure, the structure of the full-length minor pilin RrgA from *S. pneumoniae* (PDB entry 2ww8), which exhibits 51% sequence identity to GBS104 (Fig. 2), was reported (Izoré *et al.*, 2010). Homology modeling of GBS104 suggested that, similar to RrgA, its N1 and N2 domains are made of two distant segments of protein primary structure and are analogous to the intertwined β-sheet domains that have been observed in complement proteins C3, C4 and C5 and cobra venom factor (Janssen *et al.*, 2006; Krishnan *et al.*, 2009). We designed a new construct of GBS104, rGBS104$_{48–863}$, based on the primary-sequence alignment of GBS104 and RrgA, which yielded needle-shaped crystals that diffracted to 5 Å resolution at the home source. SDS–PAGE analysis revealed that the crystals are of a 75 kDa fragment and not the full-length construct that was used for crystallization. We suspected the loss of the N4 domain (about 15 kDa) during crystallization or purification. We obtained better crystals with the redesigned construct rGBS104$_{48–729}$, which lacked the N4 domain, in the presence of $CdCl_2$ as an additive, and the crystals diffracted to 2.6 Å resolution on the NE-CAT 24-ID beamline at the APS. These crystals belonged to space group $P3_121$, with one molecule per asymmetric unit (Table 1). The structure was solved by molecular replacement using the GBS104 N3 domain as a search model. The difference electron density confirmed the presence of the N2 domain; there was no density for the N1 domain. N-terminal sequencing of the recombinant protein obtained after re-dissolving the crystals confirmed its start as Glu133. The loss of the N1 domain could arise from cleavage (Lys132-

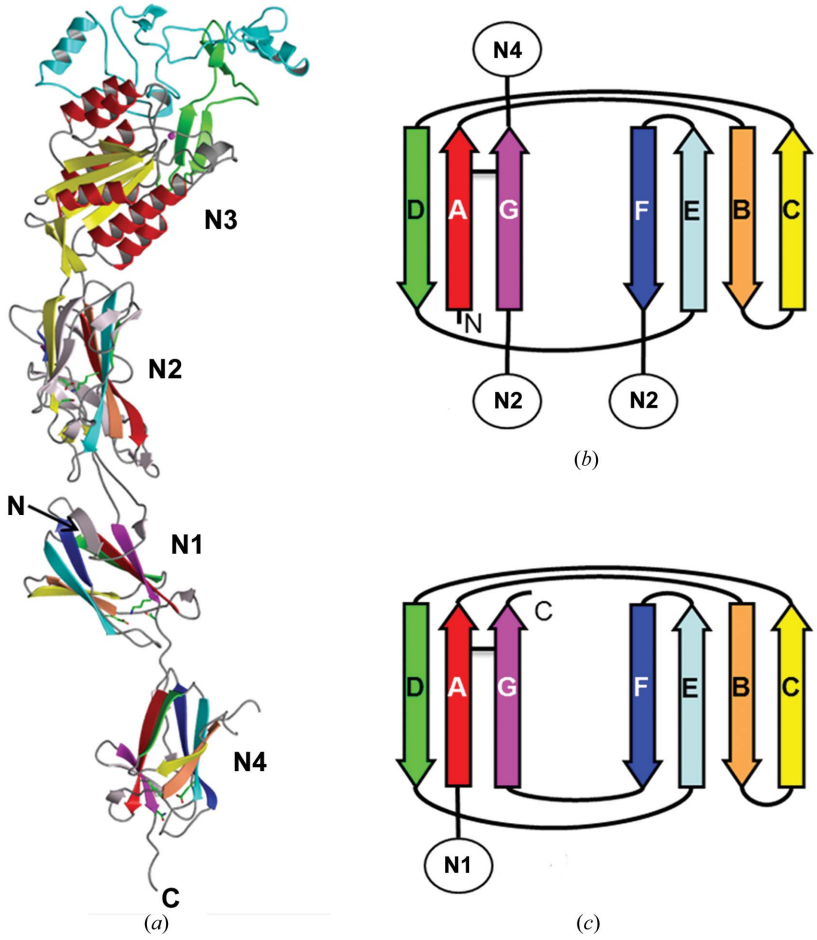

Figure 6
(*a*) Ribbon representation of the full-length GBS104 structure generated by combining the crystal structure of the N2 + N3 domains with the models of the N1 and N4 domains, which were modeled using the *S. pneumoniae* RrgA crystal structure with the help of *MODELLER* and the *FFAS* server (Zhang, Thiele *et al.*, 2009). The secondary-structural elements are colored in rainbow style from red to violet. The observed isopeptide bond in the N2 domain and the putative isopeptide bonds in the N1 and N4 domains are shown as stick models. Topology diagrams are shown for the CnaB-type fold for (*b*) the N1 domain and (*c*) the N4 domain. These are colored in rainbow style as in Fig. 5(*a*). The termini of the N1 and N4 domains are labeled with the names of the domain to which they are linked. The horizontal solid lines indicate the positions of putative intramolecular isopeptide bonds.

Glu133) by a protease contaminant during crystallization. We built a model of the N2 domain with the help of difference electron-density maps, using the homologous RrgA model as a guide, and traced the polypeptide from Ala137 to Val719.

The GBS104 N2 domain displays an IgG-like fold which is made up of two distinct parts of the same chain. The N-terminal half of the GBS104 N2-domain crystal structure consists of 73 residues (137–210) and the C-terminal half consists of 130 residues (590–719). The IgG-like fold of the GBS104 N2 domain is similar to the CnaA-type fold (Vengadesan & Narayana, 2011) and has two β-sheets: strands ABED form β-sheet I and strands CFG form β-sheet II. As observed in other typical CnaA-type fold domain structures, extra strands (D′ and D″) are observed between strands E and D which align with strands CFG at the interface of sheets I and II (Figs. 5*a* and 5*b*). Interestingly, additional β-strands and loops which are distant from the core are formed by the N-terminal residues 137–182. A D-type isopeptide bond, which has often been observed in the CnaA-type folds of several Gram-positive pilins and MSCRAMMs (Hendrickx *et al.*, 2011; Kang & Baker, 2011; Vengadesan & Narayana, 2011), is observed to connect two β-sheets of the N2 domain involving Lys188 and Asn692 from the two antiparallel strands A (sheet I) and F (sheet II), respectively (Fig. 5*a*). The catalytic Asp597 present on strand C (sheet II) is surrounded by many hydrophobic residues (Fig. 5*c*).

The N3 domain is inserted between strands B and C of the N2 domain through two short β-hairpins (201–214 and 585–589). Interestingly, we observed an inverse RGD motif formed by residues 176–178 at the N-terminal end of strand A in the N2 domain (Fig. 5*a*). The DGR motif is often observed and has been implicated in integrin binding and cellular adhesion (Gao & Brigstock, 2006), although its neighbors and the kinetics of its transformation into *iso*DGR, which can mimic the RGD motif for binding to integrins (Curnis *et al.*, 2006), affects its role in adhesion (Corti & Curnis, 2011). This DGR motif is observed distant from the β-core and on an Ω-loop that is easily accessible to the solvent and targeting receptors. The difference electron-density maps revealed the presence of several cadmium ions, which is not surprising given that $CdCl_2$ was an additive in crystallization. These ions all seem to stabilize the structure and crystal packing, mostly interacting with His residues and waters. Nine cadmium sites were associated with the GBS104N2N3 structure in one asymmetric unit with a >7σ level in the difference electron density. Crystals generated in the absence of $CdCl_2$ were of poor diffraction quality, suggesting that $CdCl_2$ is required for the stabilization of flexible regions and for stable crystal packing.

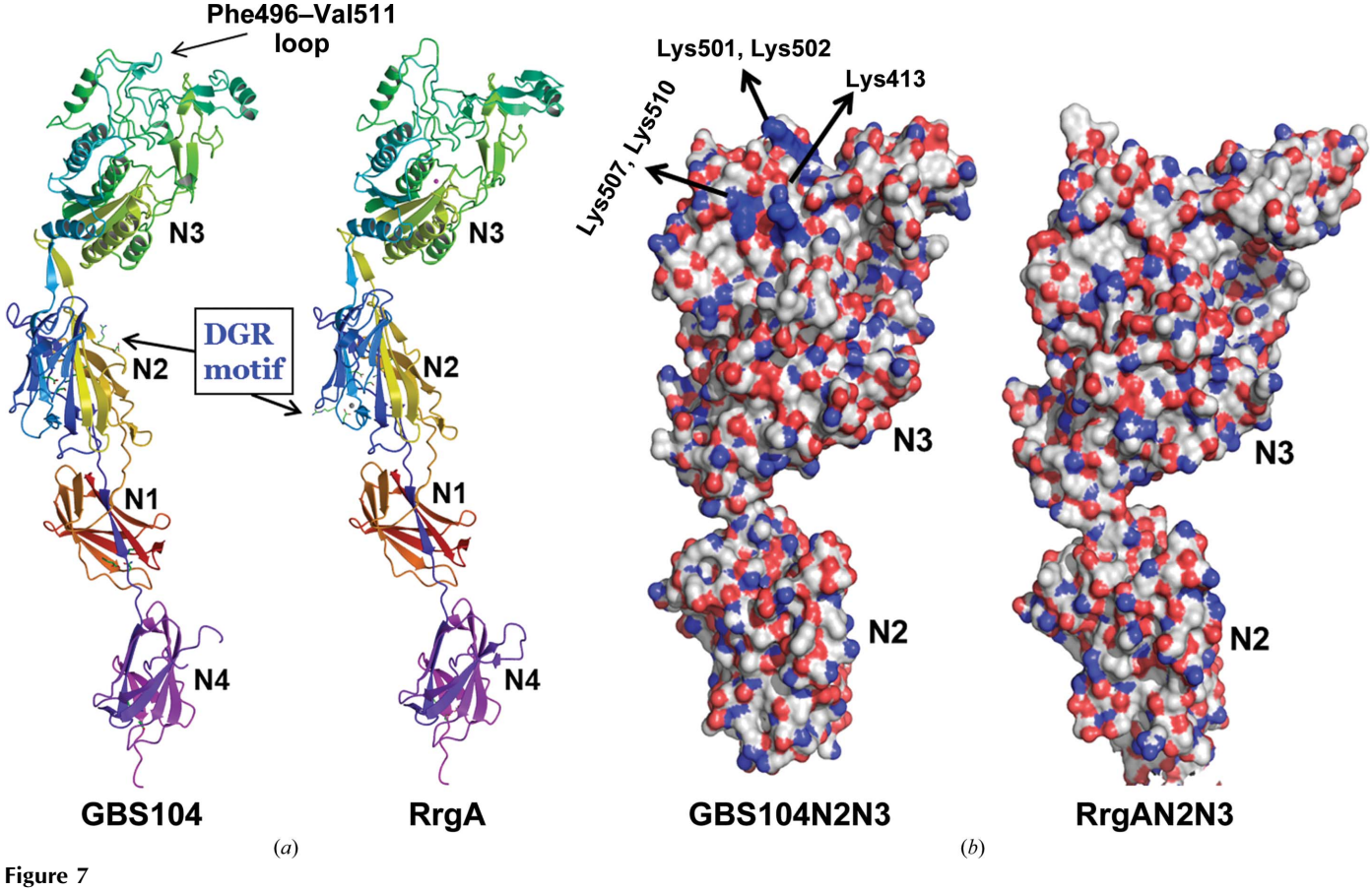

Figure 7
Comparison of the GBS104 and RrgA structures. (*a*) The RrgA structure (PDB entry 2ww8) is superimposed on the GBS104 model and translated laterally to show their domains in similar orientations. Individual domains are labeled and the respective DGR motifs are marked by arrows. (*b*) The atomic structure of GBS104N2N3 and RrgAN2N3 in surface representation (C atoms are in gray, O atoms are in red and N atoms are in blue). The Lys residues in the Phe496–Val511 loop of the GBS104 N3 domain are marked.

3.6. Models of the N1 and N4 domains

The GBS104 N1 and N4 domains were modeled with the help of *MODELLER* (Martí-Renom *et al.*, 2000) and the *FFAS* server (Zhang, Thiele *et al.*, 2009) using the homologous RrgA N1 and N4 domains (47 and 50% primary sequence identity, respectively). A model of full-length GBS104, shown in Fig. 6(*a*), was built by combining the GBS104N2N3 crystal structure with the models of the N1 and N4 domains. The GBS104 N1-domain model is composed of two distinct parts of the primary chain, similar to the N2 domain. However, approximately 90% of the N1 domain is composed of the N-terminal segment 37–136, and the remainder is composed of the C-terminal segment 720–730 (Fig. 1). The overall fold of the N1 domain resembles a CnaB-type fold (Fig. 6*b*) and is composed of two β-sheets, with strands DAG as β-sheet I and CBEF as β-sheet II. An E-type isopeptide bond is predicted to link Lys45 of strand A and Asn729 of strand G, which are two adjacent parallel strands of β-sheet I, while the catalytic Glu94 is present on strand E of β-sheet II. In addition, the predicted isopeptide bond in the N1-domain model connects the N- and C-terminal parts of GBS104, which furnish the two adjacent parallel strands of N1 β-sheet I. In contrast, the N4-domain model consists of only C-terminal residues (731–856). It displays a CnaB-type fold (Fig. 6*c*) with a putative E-type isopeptide bond between Lys740 and Asn850 present on the parallel strands A and G, respectively, of β-sheet I. Glu798, which is the catalytic residue for this putative isopeptide bond, is located on strand E of β-sheet II.

3.7. Comparison of the full-length GBS104 and RrgA models

The full-length GBS104 model and the crystal structure of the *S. pneumoniae* tip pilin RrgA are shown in Fig. 7(*a*). Superposition of the GBS104N2N3 and RrgAN2N3 coordinates gives an r.m.s.d. of 1.83 Å for 554 aligned residues. The two inserted arms in the respective N3 domains are almost identical in length and orientation. However, there are five exposed Lys residues (499, 501, 502, 507 and 510) in the extended 496–511 loop of GBS104 that are absent in the corresponding loop of RrgA (Figs. 2, 7*a* and 7*b*). The MIDAS arrangement, C-terminal α7 helix and the TDG-motif configuration (Table 2) are similar in the two pilins.

The isopeptide bonds observed in the GBS104 and RrgA N2-domain crystal structures are similarly positioned, as are the predicted isopeptide bonds in the modeled GBS104 N4-domain structure and the observed RrgA N4-domain crystal structure. We predicted one isopeptide bond in the GBS104 N1-domain model between Lys45 and Asn729. However, a similar isopeptide bond is not observed in the RrgA N1-domain crystal structure because the corresponding residues are replaced by Glu54 and Lys731, respectively.

The DGR motif (176–178) present at the N-terminal end of strand A is observed to project away from the core of the GBS104 N2 domain. The corresponding region in the RrgA N2 domain is three residues shorter than the region in the GBS104 N2 domain and does not have a DGR motif. Nevertheless, a DGR motif (613–615) is present on the opposite face of the RrgA N2-domain core on a CD loop (Fig. 7*a*). However, the tip pilin PilA (the PI-1 ortholog of GBS104) from GBS strain NCTC10/84 (Doran *et al.*, 2003) does not have a DGR motif in its N2 domain. Two calcium ion-

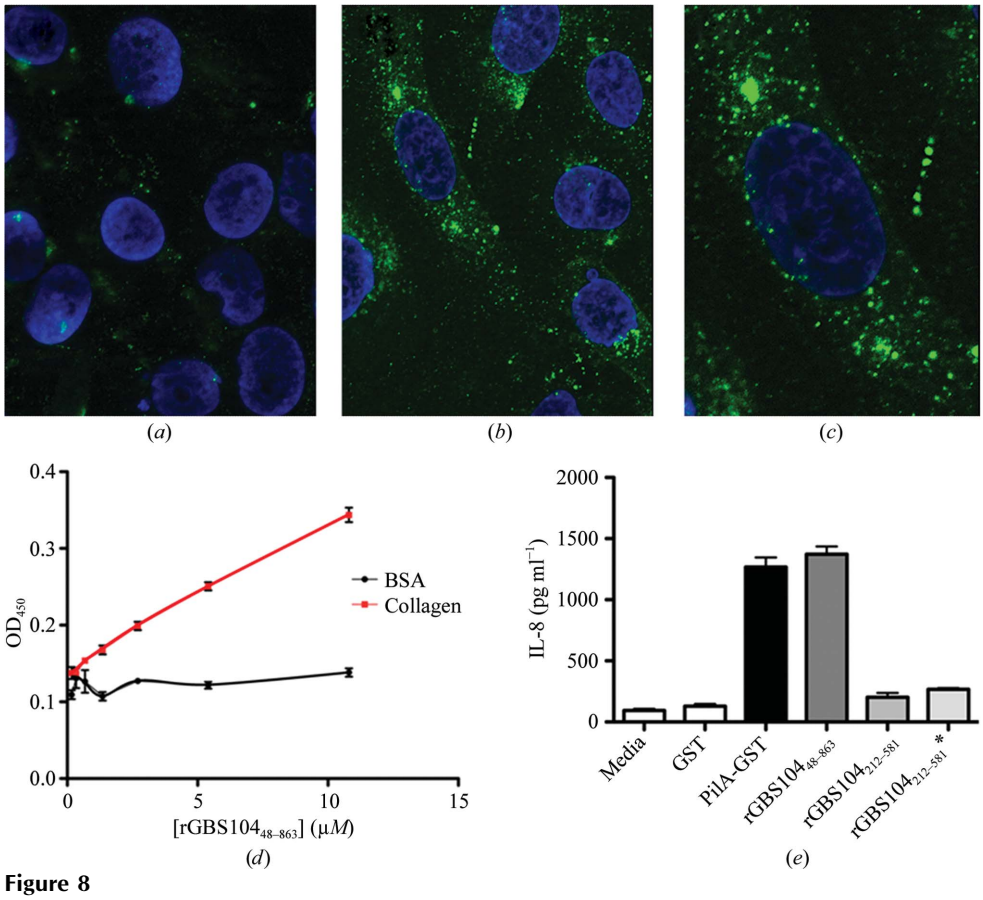

Figure 8
GBS104 binds to endothelial cells and collagen and promotes chemokine secretion. Binding of purified $rGBS104_{48–863}$ protein to hBMEC cells is shown in this figure. Cells were treated with $rGBS104_{48–863}$ and were then probed with (*a*) anti-IgG only or (*b*, *c*) control or primary antiserum. The nuclei were stained with DAPI. A magnification of the cells in (*b*) is shown in (*c*). (*d*) Dose-dependent binding of $rGBS104_{48–863}$ protein to collagen I compared with BSA was measured by ELISA. (*e*) IL-8 secretion by hBMEC cells upon incubation with GBS pilin proteins, recombinant PilA, $rGBS104_{48–863}$, $GBS104_{212–581}$ and $GBS104_{212–581}$* (GBS104N3mut) is shown. Concentrations of IL-8 were measured by ELISA using hBMEC supernatants collected after 4 h incubation.

binding sites noted in the N2 and N3 domains of the RrgA crystal structure are not observed in GBS104N2N3 because the corresponding metal-coordinating residues are not fully conserved.

3.8. Identification of potential GBS104 host receptors

Tamura *et al.* (1994) identified that GBS strain COH1, isolated from an infected newborn infant, binds lung epithelial A549 cells using its surface protein components and organelles. In addition, Tamura & Rubens (1995) demonstrated that all GBS strains from human sources adhered specifically to immobilized fibronectin at levels varying from 4 to 60%. Recently, PI-1 orthologs from GBS strain NCTC10/84 have been investigated for their role in central nervous system infections using an *in vitro* model of the blood–brain barrier (BBB; Doran *et al.*, 2003) consisting of hBMEC cells that retain the morphological and functional characteristics of primary brain endothelial cells. Banerjee *et al.* (2011) also investigated the BBB endothelium response to GBS NCTC10/84 PilA (the PI-1 ortholog of GBS104) and identified its role in promoting the neutrophilic inflammatory response of acute bacterial meningitis. They also identified PilA interactions with collagen, which engaged integrins and the integrin-signaling machinery that contributes to the pathogenesis of meningitis *in vivo*.

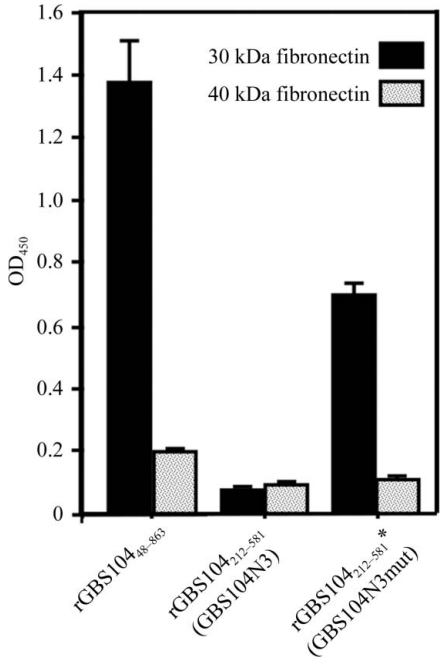

Figure 9
Binding of GBS104 to fibronectin determined by ELISA. 30 and 40 kDa fragments of fibronectin were used in this assay. The values shown are the averages of two independent experiments, each performed in triplicate. The full-length $rGBS104_{48–863}$ exhibits high affinity compared with the recombinant GBS104N3 fragment ($rGBS104_{212–581}$) towards the 30 kDa fragment of fibronectin. However, the same domain locked in the open/high-affinity state, GBS104N3mut (labeled $rGBS104_{212–581}$* in the figure), exhibits higher affinity towards the same 30 kDa fibronectin fragment compared with the wild type.

With the help of immunofluorescence microscopy, we determined that the GBS104 protein binds directly to host cells, specifically to hBMEC cells (Figs. 8*a*, 8*b* and 8*c*). Direct binding of purified GBS104 protein to collagen I was also assessed in an ELISA-based assay, in which we observed a dose-dependent increase in the binding of the GBS104 protein to collagen compared with BSA control (Fig. 8*d*). Furthermore, treatment of hBMEC with PilA and GBS104 proteins for 4 h resulted in a significant induction of IL-8 secretion compared with GST-protein and medium-only controls (Fig. 8*e*).

We confirmed the association between rGBS104 and immobilized fibronectin (FN) as having the highest affinity compared with collagen and other ECM proteins such as fibrinogen (data not shown). To localize the FN binding on GBS104, we tested the recombinants $rGBS104_{48–863}$ and rGBS104N3 ($GBS104_{212–581}$), together with rGBS104N3mut ($rGBS104_{212–518}$*), using the ELISA method (Fig. 9). Fig. 9 suggests that $rGBS104_{48–863}$ shows the highest affinity for 30 kDa FN and that the N3 domain ($rGBS104_{212–58}$) exhibits low affinity. However, the GBS104 N3-domain mutant GBS104N3mut locked in the high-affinity open conformation exhibits a considerably higher affinity compared with the wild type.

4. Discussion

The pili are critical determinants for host–pathogen interactions and investigation of their assembly and biogenesis is an emerging and exciting field. In this context, the structural and biological understanding of Gram-positive pili is an intriguing area with many unknowns compared with the assembly of Gram-negative pili, which is well understood (Kline *et al.*, 2010; Krishnan & Narayana, 2011; Telford *et al.*, 2006). There are no similarities between the Gram-negative and Gram-positive type I pili polymerization modes (chaperone-ushered *versus* sortase-linked) and individual component associations (hydrophobic interactions *versus* covalent linkages), but the pili of both pathogen types host specific and characteristic adhesins at the pilus tip that are implicated in pathogen–host interactions. Glycoproteins and glycolipids present on epithelial cells have been shown to act as host receptors for Gram-negative pathogenic type I pili components (Zhou *et al.*, 2001; Xie *et al.*, 2006; Hung *et al.*, 2002), while the host partners for Gram-positive pili are not known. In this report, we present crystal structures and a model of the Gram-positive *S. agalactiae* tip pilin GBS104 which, together with the minor pilin GBS52, has been identified as being essential for adhesion to lung epithelial cells (Krishnan, Gaspar *et al.*, 2007). Interestingly, GBS104 and its structural homologue *S. pneumoniae* tip pilin RrgA (Izoré *et al.*, 2010) are assembled of eukaryotic IgG-like and vWFA-like domains, which are known for their adhesive functions. Surface proteins and pili of *S. pneumoniae*, *S. pyogenes* and *S. agalactiae* have been implicated in adherence to ECM proteins (Tamura & Rubens, 1995; Kreikemeyer *et al.*, 2005; Banerjee *et al.*, 2011) and their tip pilins host adhesive

vWFA-like domains, which could be a result of convergent evolution or could perhaps have arisen from lateral or horizontal gene transfer from higher organisms.

Interestingly, GBS104 exhibits a new class of IgG-C variants compared with two other GBS pilins (GBS52 and GBS80). The GBS104 N2 domain with a CnaA-type fold is constituted of two distinct segments separated by almost 400 residues. One segment (one third of the total residues) comes from the N-termini and the other two thirds come from the C-terminus; a D-type isopeptide bond facilitates stabilization of the association of these two unequal parts. The N1-domain model that we have built exhibits a similar CnaB-type fold with an isopeptide bond linking neighboring β-strands of a sheet that is formed by two separate and remote segments (Fig. 1*c*).

However, the N1 domain of the homologous RgrA contains no isopeptide bonds and we suggest that the absence and the presence of the isopeptide bonds in the N1 domains of RrgA and GBS104, respectively, may impact on their stability. The *S. pneumoniae* pilin SpaA, which has no isopeptide bonds in the N1 domain, was stable and resistant to proteolysis (Kang, Paterson *et al.*, 2009). While crystallization of full-length SpaA was possible, the full-length pilins *S. agalactiae* GBS80 and *A. naeslundii* FimA with putative isopeptide bonds in their N1 domains are unstable and were found to be degraded during crystallization. Alternatively, their crystallization was possible after limited proteolytic cleavage of the N1 domains (Mishra *et al.*, 2011; Vengadesan *et al.*, 2011). Therefore, similar instability problems may be the reason for the breakdown of the recombinant rGBS104$_{29–729}$ and rGBS104$_{48–863}$ proteins, while recombinant RrgA is stable during crystallization.

Prokaryotic vWFA domains are rare, while they are extensively observed as adhesion domains in eukaryotic

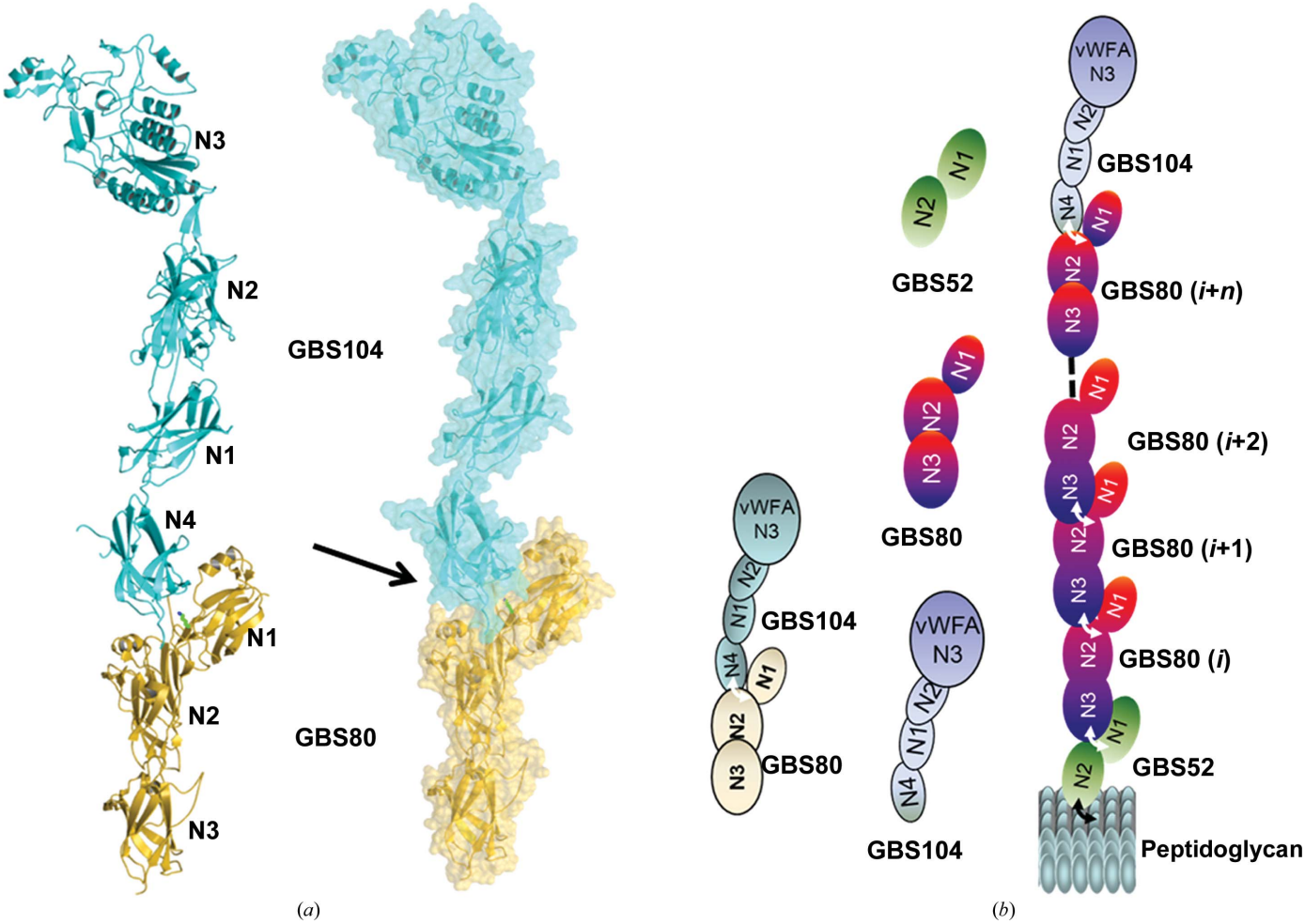

Figure 10
A model of the association of the major pilin GBS80 with the minor pilins GBS52 and GBS104. (*a*) A schematic representation of the possible association of tip pilin GBS104 with the major pilin GBS80 *via* an intermolecular isopeptide bond (black arrow) between the GBS80 pilin motif (IYP**K**N) Lys199 and the sorting-motif (FPK**T**G) Thr of GBS104. (*b*) A structural model of GBS104–GBS80–GBS52 association and pili anchoring to the peptidoglycan. The three structures are linked through yet to be observed intermolecular isopeptide bonds. The linking of the GBS52 pilin-like motif to the GBS80 sorting motif terminates pilus assembly when the former is presented by the housekeeping sortase SrtA. The polymerized pili with GBS52 at the base are transferred and linked to the cell-surface peptidoglycan when the acylated SrtA (attached to polymerized pili) is resolved by the cell-wall nucleophilic amino-acid cross-bridge. The hydrophobic GBS52 C-terminal extension, with the sorting motif, could facilitate stabilization of cell-wall and pili-base interactions. However, if GBS52 is presented by a pilus-specific sortase instead of a housekeeping sortase, and the acylated enzyme is resolved by the pilin motif of GBS80, the polymerization of the pilus shaft can continue without stopping and thus a model of intermittent distribution of GBS52 along the pilus shaft is also possible.

proteins. The GBS104 N3 domain exhibiting a vWFA-like fold is present at the head of the spear-shaped GBS104 (Fig. 10*a*). It is composed of three IgG-C variants arranged in tandem (N4 + N1 + N2) and the vWFA-like N3. The role of the N1, N2 and N4 domains could be similar to that of the *S. aureus* collagen-binding CNA B-region, which is an extension that pushes the key ligand-binding area away from the bacterial cell surface (Deivanayagam *et al.*, 2000). Incidentally, the vWFA-like fold of the N3 domain in GBS104 and RrgA is unique compared with the eukaryotic I domains, which have two extended arms that are shaped and charged, perhaps to capture and hold ECM proteins or other ligands. The tip adhesins *S. agalactiae* GBS104, *S. pneumoniae* RrgA and Spy0125 of *S. pyogenes* strain SF370 (which is absent in many other strains; Smith *et al.*, 2010) exhibit significant sequence identity (Fig. 2) among their vWFA-like domains (>50%; Izoré *et al.*, 2010). However, some patches of positively charged and polar residues present in the C-terminal extended arm 2 sets GBS104 apart from RrgA and could have an impact on its ligand specificity and pathogenicity. GBS pili have also been implicated in biofilm formation. However, the vWFA-like domain of tip pilin (strain NEM316) implicated in binding to epithelial cells is not necessary for biofilm formation (Konto-Ghiorghi *et al.*, 2009).

The PI-1 pathogenicity island of GBS strain 2603V/R, also known as the *srtC1-srtC2* locus, has genes for GBS52, GBS80 and GBS104 and sortases that are necessary for pilus assembly (GBSSrtC1 and GBSSrtC2). The locus for the *srtA* gene, which encodes the housekeeping sortase GBSSrtA, is far from the *srtC1-srtC2* locus (Dramsi *et al.*, 2006). With the crystal structures of three (full-length or fragment) pilins including GBS52 (Krishnan, Gaspar *et al.*, 2007), GBS80 (Vengadesan *et al.*, 2011) and GBS104 and of the sortases GBSSrtA and GBSSrtC1 (Khare *et al.*, 2011) of GBS strain 2603V/A in hand, we can present a model of full-length GBS PI-1 pilus assembly (Figs. 10*a* and 10*b*). The major pilin GBS80 forms the pilus shaft and two minor pilins that are predicted to be adhesins are present at either end of the GBS80 shaft (Mandlik *et al.*, 2008).

The GBS cell-surface extension pili are implicated in life-threatening infections in newborns and their components are considered to be potential vaccine candidates. Maione *et al.* (2005) first explored the possibility of pilus components eliciting antibodies that provide protection against GBS infections in mouse offspring. Interestingly, multiple GBS strains conserve one of the three known pilus islands with minimal variations in the pilus components, and vaccines consisting of a combination of conserved pilus components provide protection against all GBS strains (Margarit *et al.*, 2009). Although the crystal structures of GBS104 from *S. agalactiae* and RrgA from *S. pneumoniae* appear to be quite similar, there are structural differences, both subtle and large, between them, especially in the loops around the MIDAS. Therefore, vaccines directed against the domains that are involved in host-receptor binding can be pathogen-specific compared with those directed against the structurally homologous N1, N2 and N4 domains. We suggest that the information regarding the localization of ECM protein binding on GBS104, which we will determine more specifically in the future, together with the identification of host proteins that are receptors for the N2-domain DGR motif, could be helpful and may lead to successful structural vaccinology against GBS infections (Nuccitelli *et al.*, 2011).

We thank members of our laboratory for critical reviews of the manuscript and discussion, and Monique Stins and Kwang Sik Kim for providing hBMEC. We collected two diffraction data sets (GBS104N3-SeMet and GBS104N2N3) on the NE-CAT 24ID, and we are thankful to Dr Rajashankar for helping us with the data collection. This work was supported by US Public Health Service grants R01NS051247 to KSD from the National Institute of Neurological Disorders and Stroke, AI061381 to HT-T and AI037251 to SVLN from the National Institute of Allergy and Infectious Diseases.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Banerjee, A., Kim, B. J., Carmona, E. M., Cutting, A. S., Gurney, M. A., Carlos, C., Feuer, R., Prasadarao, N. V. & Doran, K. S. (2011). *Nature Commun.* **2**, 462.
Bolaños, M., Cañas, A., Santana, O. E., Pérez-Arellano, J. L., de Miguel, I. & Martín-Sánchez, A. M. (2001). *Eur. J. Clin. Microbiol. Infect. Dis.* **20**, 837–839.
Budzik, J. M., Poor, C. B., Faull, K. F., Whitelegge, J. P., He, C. & Schneewind, O. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 19992–19997.
Corti, A. & Curnis, F. (2011). *J. Cell Sci.* **124**, 515–522.
Craig, L., Pique, M. E. & Tainer, J. A. (2004). *Nature Rev. Microbiol.* **2**, 363–378.
Curnis, F., Longhi, R., Crippa, L., Cattaneo, A., Dondossola, E., Bachi, A. & Corti, A. (2006). *J. Biol. Chem.* **281**, 36466–36476.
Deivanayagam, C. C., Rich, R. L., Carson, M., Owens, R. T., Danthuluri, S., Bice, T., Höök, M. & Narayana, S. V. (2000). *Structure*, **8**, 67–78.
Dermer, P., Lee, C., Eggert, J. & Few, B. (2004). *J. Pediatr. Nurs.* **19**, 357–363.
Doran, K. S., Liu, G. Y. & Nizet, V. (2003). *J. Clin. Invest.* **112**, 736–744.
Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prévost, M.-C., Kojadinovic, M., Lalioui, L., Poyart, C. & Trieu-Cuot, P. (2006). *Mol. Microbiol.* **60**, 1401–1413.
Edwards, M. S., Rench, M. A., Palazzi, D. L. & Baker, C. J. (2005). *Clin. Infect. Dis.* **40**, 352–357.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fischetti, V. A., Pancholi, V. & Schneewind, O. (1990). *Mol. Microbiol.* **4**, 1603–1605.
Gao, R. & Brigstock, D. R. (2006). *Gut*, **55**, 856–862.
Heelan, J. S., Hasenbein, M. E. & McAdam, A. J. (2004). *J. Clin. Microbiol.* **42**, 1263–1264.
Hendrickx, A. P. A., Budzik, J. M., Oh, S.-Y. & Schneewind, O. (2011). *Nature Rev. Microbiol.* **9**, 166–176.
Hung, C.-S., Bouckaert, J., Hung, D., Pinkner, J., Widberg, C., DeFusco, A., Auguste, C. G., Strouse, R., Langermann, S., Waksman, G. & Hultgren, S. J. (2002). *Mol. Microbiol.* **44**, 903–915.
Izoré, T., Contreras-Martel, C., El Mortaji, L., Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A. M. & Dessen, A. (2010). *Structure*, **18**, 106–115.
Janssen, B. J., Christodoulidou, A., McCarthy, A., Lambris, J. D. & Gros, P. (2006). *Nature (London)*, **444**, 213–216.

Janssen, B. J., Gomes, L., Koning, R. I., Svergun, D. I., Koster, A. J., Fritzinger, D. C., Vogel, C. W. & Gros, P. (2009). *EMBO J.* **28**, 2469–2478.

Johri, A. K., Paoletti, L. C., Glaser, P., Dua, M., Sharma, P. K., Grandi, G. & Rappuoli, R. (2006). *Nature Rev. Microbiol.* **4**, 932–942.

Kang, H. J. & Baker, E. N. (2011). *Trends Biochem. Sci.* **36**, 229–237.

Kang, H. J. & Baker, E. N. (2012). *Curr. Opin. Struct. Biol.* **22**, 200–207.

Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. (2007). *Science*, **318**, 1625–1628.

Kang, H. J., Middleditch, M., Proft, T. & Baker, E. N. (2009). *Biopolymers*, **91**, 1126–1134.

Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H. & Baker, E. N. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 16967–16971.

Khare, B., Krishnan, V., Rajashankar, K. R., I-Hsiu, H., Xin, M., Ton-That, H. & Narayana, S. V. (2011). *PLoS One*, **6**, e22995.

Kline, K. A., Dodson, K. W., Caparon, M. G. & Hultgren, S. J. (2010). *Trends Microbiol.* **18**, 224–232.

Konto-Ghiorghi, Y., Mairey, E., Mallet, A., Duménil, G., Caliot, E., Trieu-Cuot, P. & Dramsi, S. (2009). *PLoS Pathog.* **5**, e1000422.

Kreikemeyer, B., Nakata, M., Oehmcke, S., Gschwendtner, C., Normann, J. & Podbielski, A. (2005). *J. Biol. Chem.* **280**, 33228–33239.

Krishnan, V., Gaspar, A. H., Ye, N., Mandlik, A., Ton-That, H. & Narayana, S. V. L. (2007). *Structure*, **15**, 893–903.

Krishnan, V. & Narayana, S. V. L. (2011). *Adv. Exp. Med. Biol.* **715**, 175–195.

Krishnan, V., Ponnuraj, K., Xu, Y., Macon, K., Volanakis, J. E. & Narayana, S. V. L. (2009). *Structure*, **17**, 611–619.

Krishnan, V., Xu, Y., Macon, K., Volanakis, J. E. & Narayana, S. V. L. (2007). *J. Mol. Biol.* **367**, 224–233.

Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.

Lauer, P., Rinaudo, C. D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A. R., Mora, M., Rappuoli, R., Grandi, G. & Telford, J. L. (2005). *Science*, **309**, 105.

Lee, J.-O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. (1995). *Structure*, **3**, 1333–1340.

Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995). *Cell*, **80**, 631–638.

Lin, F.-Y. C., Azimi, P. H., Weisman, L. E., Philips, J. B. III, Regan, J., Clark, P., Rhoads, G. G., Clemens, J., Troendle, J., Pratt, E., Brenner, R. A. & Gill, V. (2000). *Clin. Infect. Dis.* **31**, 76–79.

Linke, C., Young, P. G., Kang, H. J., Bunker, R. D., Middleditch, M. J., Caradoc-Davies, T. T., Proft, T. & Baker, E. N. (2010). *J. Biol. Chem.* **285**, 20381–20389.

Loftus, J. C., Smith, J. W. & Ginsberg, M. H. (1994). *J. Biol. Chem.* **269**, 25235–25238.

Lu, C., Shimaoka, M., Ferzly, M., Oxvig, C., Takagi, J. & Springer, T. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 2387–2392.

Lu, C., Shimaoka, M., Zang, Q., Takagi, J. & Springer, T. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 2393–2398.

Luo, B.-H., Carman, C. V. & Springer, T. A. (2007). *Annu. Rev. Immunol.* **25**, 619–647.

Luo, B.-H., Takagi, J. & Springer, T. A. (2004). *J. Biol. Chem.* **279**, 10215–10221.

Maione, D. *et al.* (2005). *Science*, **309**, 148–150.

Mandlik, A., Das, A. & Ton-That, H. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 14147–14152.

Margarit, I. *et al.* (2009). *J. Infect. Dis.* **199**, 108–115.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F. & Sali, A. (2000). *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.

McCleverty, C. J. & Liddington, R. C. (2003). *Biochem. J.* **372**, 121–127.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Mishra, A., Devarajan, B., Reardon, M. E., Dwivedi, P., Krishnan, V., Cisar, J. O., Das, A., Narayana, S. V. & Ton-That, H. (2011). *Mol. Microbiol.* **81**, 1205–1220.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nuccitelli, A., Cozzi, R., Gourlay, L. J., Donnarumma, D., Necchi, F., Norais, N., Telford, J. L., Rappuoli, R., Bolognesi, M., Maione, D., Grandi, G. & Rinaudo, C. D. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 10278–10283.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.

Plow, E. F., Haas, T. A., Zhang, L., Loftus, J. & Smith, J. W. (2000). *J. Biol. Chem.* **275**, 21785–21788.

Ponnuraj, K., Xu, Y., Macon, K., Moore, D., Volanakis, J. E. & Narayana, S. V. L. (2004). *Mol. Cell*, **14**, 17–28.

Proft, T. & Baker, E. N. (2009). *Cell. Mol. Life Sci.* **66**, 613–635.

Rooijakkers, S. H., Wu, J., Ruyken, M., van Domselaar, R., Planken, K. L., Tzekou, A., Ricklin, D., Lambris, J. D., Janssen, B. J., van Strijp, J. A. & Gros, P. (2009). *Nature Immunol.* **10**, 721–727.

Rosini, R., Rinaudo, C. D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzo, C., Brettoni, C., Buccato, S., Margarit, I., Grandi, G. & Telford, J. L. (2006). *Mol. Microbiol.* **61**, 126–141.

Sauer, F. G., Mulvey, M. A., Schilling, J. D., Martinez, J. J. & Hultgren, S. J. (2000). *Curr. Opin. Microbiol.* **3**, 65–72.

Schuchat, A. (1998). *Clin. Microbiol. Rev.* **11**, 497–513.

Sendi, P., Johansson, L. & Norrby-Teglund, A. (2008). *Infection*, **36**, 100–111.

Shimaoka, M., Lu, C., Palframan, R. T., von Andrian, U. H., McCormack, A., Takagi, J. & Springer, T. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6009–6014.

Shimaoka, M., Lu, C., Salas, A., Xiao, T., Takagi, J. & Springer, T. A. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 16737–16741.

Shimaoka, M., Takagi, J. & Springer, T. A. (2002). *Annu. Rev. Biophys. Biomol. Struct.* **31**, 485–516.

Shimaoka, M., Xiao, T., Liu, J.-H., Yang, Y., Dong, Y., Jun, C.-D., McCormack, A., Zhang, R., Joachimiak, A., Takagi, J., Wang, J.-H. & Springer, T. A. (2003). *Cell*, **112**, 99–111.

Smith, W. D., Pointon, J. A., Abbot, E., Kang, H. J., Baker, E. N., Hirst, B. H., Wilson, J. A., Banfield, M. J. & Kehoe, M. A. (2010). *J. Bacteriol.* **192**, 4651–4659.

Spraggon, G., Koesema, E., Scarselli, M., Malito, E., Biagini, M., Norais, N., Emolo, C., Barocchi, M. A., Giusti, F., Hilleringmann, M., Rappuoli, R., Lesley, S., Covacci, A., Masignani, V. & Ferlenghi, I. (2010). *PLoS One*, **5**, e10919.

Tamura, G. S., Kuypers, J. M., Smith, S., Raff, H. & Rubens, C. E. (1994). *Infect. Immun.* **62**, 2450–2458.

Tamura, G. S. & Rubens, C. E. (1995). *Mol. Microbiol.* **15**, 581–589.

Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R. & Grandi, G. (2006). *Nature Rev. Microbiol.* **4**, 509–519.

Ton-That, H., Marraffini, L. A. & Schneewind, O. (2004). *Mol. Microbiol.* **53**, 251–261.

Ton-That, H. & Schneewind, O. (2003). *Mol. Microbiol.* **50**, 1429–1438.

Ton-That, H. & Schneewind, O. (2004). *Trends Microbiol.* **12**, 228–234.

Vengadesan, K., Ma, X., Dwivedi, P., Ton-That, H. & Narayana, S. V. L. (2011). *J. Mol. Biol.* **407**, 731–743.

Vengadesan, K. & Narayana, S. V. L. (2011). *Protein Sci.* **20**, 759–772.

Whittaker, C. A. & Hynes, R. O. (2002). *Mol. Biol. Cell*, **13**, 3369–3387.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wu, H. & Fives-Taylor, P. M. (2001). *Crit. Rev. Oral Biol. Med.* **12**, 101–115.

Xie, B., Zhou, G., Chan, S.-Y., Shapiro, E., Kong, X.-P., Wu, X.-R., Sun, T.-T. & Costello, C. E. (2006). *J. Biol. Chem.* **281**, 14644–14653.

Xiong, J.-P., Li, R., Essafi, M., Stehle, T. & Arnaout, M. A. (2000). *J. Biol. Chem.* **275**, 38762–38767.

Zhang, H., Liu, J., Yang, W., Springer, T., Shimaoka, M. & Wang, J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 18345–18350.
Zhang, Y., Thiele, I., Weekes, D., Li, Z., Jaroszewski, L., Ginalski, K., Deacon, A. M., Wooley, J., Lesley, S. A., Wilson, I. A., Palsson, B., Osterman, A. & Godzik, A. (2009). *Science*, **325**, 1544–1549.
Zhou, G., Mo, W.-J., Sebbel, P., Min, G., Neubert, T. A., Glockshuber, R., Wu, X.-R., Sun, T.-T. & Kong, X.-P. (2001). *J. Cell Sci.* **114**, 4095–4103.