Revathi Nagarajan,^a Antoni P. A. Hendrickx^{b‡} and Karthe Ponnuraj^{a*}

^aCentre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai 600 025, India, and

^bDepartment of Microbiology, University of Chicago, Chicago, IL 60637, USA

‡ Present address: Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands.

Correspondence e-mail: pkarthe@hotmail.com

Received 2 February 2013

Accepted 27 February 2013

Cloning, expression, purification, crystallization and preliminary crystallographic analysis of the N-terminal domain of serine glutamate repeat A (SgrA) protein from *Enterococcus faecium*

Serine glutamate repeat A (SgrA) protein is an LPxTG surface adhesin of *Enterococcus faecium* and is the first bacterial nidogen-binding protein identified to date. It has been suggested that it binds to human nidogen, the extracellular matrix molecule of basal lamina, and plays a key role in the invasion and colonization of eukaryotic host cells. SgrA_{28–288}, having both a putative ligand-binding A domain and repetitive B domain, was expressed in *Escherichia coli* and purified using Ni-affinity and hydrophobic interaction chromatography. Further, the putative ligand-binding region, rSgrA_{28–153}, was subcloned, overexpressed and purified in both native and selenomethionine-derivative forms. The native rSgrA_{28–153} protein crystallized in the monoclinic space group *P*₂₁ and diffracted to 3.3 Å resolution using an in-house X-ray source, with unit-cell parameters *a* = 35.84, *b* = 56.35, *c* = 60.20 Å, β = 106.5°.

1. Introduction

Enterococci are Gram-positive facultatively anaerobic oval cocci and are commensals of the gastrointestinal tract of humans and other animals as well as those of insects and nematodes (Murray, 1990). During the last three decades, *Enterococcus faecium* has emerged as an opportunistic pathogen and is a common cause of nosocomial infections such as septicaemia, endocarditis and urinary-tract infections in immunocompromised patients (Richards *et al.*, 2000). The hospital-associated *E. faecium* strains are highly resistant to ampicillin, ciprofloxacin and vancomycin (Arias & Murray, 2012) and are enriched in LPxTG-type surface proteins (Hendrickx *et al.*, 2007). The treatment of enterococcal infections has become challenging as resistance to newly developed antibiotics has been observed (Hendrickx, Luit-Asbroek *et al.*, 2009).

It has been proposed that several enterococcal genes, including those encoding cell-wall-anchored LPxTG-type surface proteins, are involved in the pathogenesis of enterococcal infections (Hendrickx, Luit-Asbroek *et al.*, 2009; Hendrickx, Willems *et al.*, 2009). Adherence is the critical first step in bacterial pathogenesis or infection. Most pathogenic bacteria have been shown to recognize and adhere to various components of extracellular matrix (ECM), leading to colonization of the host tissue (Patti & Höök, 1994; Foster & Höök, 1998; Murray, 2000). ECM includes glycoproteins such as collagen, fibrinogen, fibronectin, nidogen and laminin, the latter two forming the underlying basement membrane of epithelial and endothelial cells (Hay, 1991).

Serine glutamate repeat A (SgrA) is an LPxTG surface adhesin of *E. faecium*. It consists of 324 amino acids and is a marker of multi-drug-resistant and hospital-associated *E. faecium*. It has an N-terminal signal peptide, a nonrepetitive A domain, a repetitive B domain containing eight repeats of Ser–Ser–Glu–Ser–Ser–Thr followed by a cell-wall sorting signal (CWS), which is composed of highly conserved LPxTG-like (Leu–Pro–*x*–Thr–Gly, where *x* denotes any amino acid) sortase substrate motif, a hydrophobic domain and a short C-terminal charged tail (Hendrickx, Willems *et al.*, 2009; Fig. 1). After translocation of the precursor protein, the LPxTG motif is cleaved by a housekeeping sortase, which subsequently anchors the surface protein to the cell-wall peptidoglycan. SgrA is thought to bind to protein ligands of ECM especially fibrinogen and nidogen (Hendrickx, Luit-Asbroek *et al.*, 2009). Fibrinogen is a large 340 kDa

plasma protein and plays an important role in homeostasis and coagulation. The two homologous proteins nidogen-1 (entactin) and nidogen-2 are sulfated monomeric glycoproteins of 150 and 200 kDa, respectively. Nidogens are found ubiquitously in basement membranes. Previous study reveals that rSgrA_{28–288} binds to α and β chains of fibrinogen and to nidogen (Hendrickx, Luit-Asbroek *et al.*, 2009). The N-terminal region of SgrA (A domain, residues 28–153) is considered to be the putative ligand-binding domain (Hendrickx, Luit-Asbroek *et al.*, 2009).

Enterococci have been associated with the formation of biofilms on various kinds of in-dwelling medical devices, such as artificial hip prostheses, intra-uterine devices, prosthetic heart valves, central venous catheters and urinary catheters, leading to many infections (Donlan, 2002; Paganelli *et al.*, 2012). Previously, it was shown that SgrA mediates biofilm formation onto a polystyrene abiotic surface but not onto biotic surfaces such as human intestinal epithelial cells, human bladder cells and kidney cells (Hendrickx, Luit-Asbroek *et al.*, 2009). This suggests that SgrA may also play a role in adhesion to medical-device-related infections by forming a biofilm.

Here, we report the cloning, expression, purification, crystallization and preliminary X-ray analysis of the ligand-binding region of recombinant SgrA (rSgrA_{28–153}). Understanding the structure of rSgrA_{28–153} and subsequent biochemical studies with human nidogen and fibrinogen will be useful in the development of novel therapeutics against enterococcal infections.

2. Materials and methods

2.1. Cloning and expression of rSgrA_{28–288} and rSgrA_{28–153}

The DNA sequence encoding both the A and B regions of SgrA (residues 28–288; molecular mass 31 kDa) was amplified by polymerase chain reaction (PCR) from genomic DNA of *E. faecium* TX0016. To generate a C-terminally His₆-tagged fusion protein, the PCR product was cloned into a *Bam*HI/*Xho*I-digested expression vector pET-24b(+). The resulting plasmid was transformed into the expression host *Escherichia coli* BL21(DE3) and protein expression was carried out in Luria–Bertani (LB) medium. A 10 ml overnight culture was prepared and transferred to 1 l LB medium supplemented with 50 μ g ml^{−1} kanamycin and allowed to grow at 310 K with shaking (180 rev min^{−1}) until an optimum optical density (OD₆₀₀ of 0.5–0.6) was reached. At this point, expression of the recombinant protein was induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 310 K. After 5 h of incubation with shaking, bacterial cells were harvested by centrifugation at 4000 rev min^{−1} for 20 min at 277 K.

The primers 5′-CGC **GGA TCC** AAT GAA CGG GCA AAT GAG AAT-3′ and 5′-CGG **CTC GAG** GCT ATC CGT ACT GCT TGG-3′ were designed for PCR amplification of *sgrA* segment 28–153. The *Bam*HI and *Xho*I restriction sites in the forward and reverse primers, respectively, are shown in bold. The plasmid DNA encoding the full-length recombinant *sgrA* gene (both A and B regions) in the

pET-24b(+) vector was used as a template. After digestion with *Bam*HI and *Xho*I restriction enzymes, the PCR product was cloned into the expression vector pET-28a(+). The construct was confirmed by DNA sequencing and allows expression of the protein with both N and C-terminal 6 $\times$ His tags.

E. coli BL21 (DE3) cells were transformed with the plasmid pET-28a(+)-*sgrA*_{28–153} and grown on LB agar plates containing 50 μ g ml^{−1} kanamycin. Following overnight incubation at 310 K, single colonies were transferred into 10 ml LB broth containing 50 μ g ml^{−1} kanamycin and grown overnight at 310 K with shaking (180 rev min^{−1}). The cells were diluted 1/100 in flasks containing 1 l LB broth supplemented with 50 μ g ml^{−1} kanamycin and cultivated at 310 K with continuous shaking (180 rev min^{−1}) until an OD_{600nm} of 0.6 was reached. The cells were subsequently treated with 1 mM IPTG to induce the expression of recombinant SgrA A domain (rSgrA_{28–153}). After 5 h of incubation with continuous shaking (180 rev min^{−1}) at 310 K, the cells were harvested by centrifugation at 4000 rev min^{−1} for 20 min at 277 K.

2.2. Purification of full-length SgrA (rSgrA_{28–288})

The cell pellet was resuspended in 25 ml cold lysis buffer (20 mM Tris–HCl pH 6.5, 500 mM NaCl, 0.01% Triton X-100, 5% glycerol) with 1 mM phenylmethanesulfonyl fluoride (PMSF) and ultrasonicated for 2 min with 25% amplitude (20 s on, 10 s off) on ice. This lysate solution was spun down at 10 000 rev min^{−1} for 45 min at 277 K. The supernatant and pellet were analysed on a 15% SDS–PAGE gel and the protein was detected in the soluble fraction. The supernatant was passed through a 0.22 μ m membrane filter to remove any cell debris in preparation for protein purification. The first step involved in protein purification was immobilized metal-affinity chromatography (IMAC) using an Ni–NTA column (5 ml HisTrap column, GE Healthcare) which was equilibrated in lysis buffer to capture the target 6 $\times$ His-tagged fusion protein from the supernatant. The bound protein was eluted using a linear gradient of 0–1 M imidazole in the lysis buffer. The fractions containing target protein were verified on 15% SDS–PAGE, pooled, concentrated to approximately 2 ml and then subjected to ammonium sulfate precipitation. In an initial analysis it was found that the protein remains soluble up to 1.5 M ammonium sulfate and precipitates at 2 M. Hence, the protein was precipitated with 2 M ammonium sulfate and the pellet was resuspended in 2 ml lysis buffer containing 1.5 M ammonium sulfate. This solution was centrifuged at 10 000 rev min^{−1} for 10 min at 277 K. The supernatant containing protein was filtered and loaded onto a pre-equilibrated Phenyl Sepharose HP column (GE Healthcare). The protein was eluted using a gradient of 1.5–0.0 M ammonium sulfate in lysis buffer. The peak fractions were verified on 15% SDS–PAGE and pure fractions were pooled and dialysed overnight against buffer consisting of 20 mM Tris–HCl pH 6.5, 200 mM NaCl, 0.01% Triton X-100, 5% glycerol to remove excess sodium chloride and ammonium sulfate. The purified protein was concentrated to 500 μ l and the concentration of the protein was measured using a UV spectrophotometer (A₂₈₀) and assuming a calculated absorption coefficient of 0.772 (*ProtParam*; Gasteiger *et al.*, 2005) with a final yield of 40 mg ml^{−1} per litre of culture. Crystallization trials were carried out with an initial concentration of 25 mg ml^{−1}.

2.3. Purification of the A domain of SgrA (rSgrA_{28–153})

For purifying rSgrA_{28–153}, a similar purification protocol to that used for rSgrA_{28–288} was carried out with slight changes in the buffer condition. In brief, buffer A consisting of 20 mM Tris–HCl pH 7.5,

Figure 1

Structural organization of the SgrA surface protein of *E. faecium*. The signal peptide, cell-wall anchoring domain, membrane-spanning domain and positively charged C-terminal domain are indicated by S, W, M and C, respectively. The nonrepetitive ligand-binding region followed by the repeat region is shown.

Table 1

Data-collection statistics.

Values in parentheses are for the highest-resolution shell.

Wavelength (Å)	1.5418
Space group	$P2_1$
Unit-cell parameters (Å, °)	$a = 35.84$, $b = 56.35$, $c = 60.20$, $\beta = 106.5$
Resolution range (Å)	30.00–3.3 (3.42–3.30)
$R_{\text{merge}}^{\dagger}$ (%)	12.4 (22.1)
Total no. of reflections	12056
No. of unique reflections	3530
Mean $I/\sigma(I)$	4.2 (1.9)
Completeness (%)	98.9 (95.0)
Multiplicity	3.3

$$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}.$$

500 mM NaCl, 0.01% Triton X-100, 5% glycerol was used for IMAC and hydrophobic interaction chromatography (HIC). It was found that rSgrA_{28–153} precipitates at 1 M ammonium sulfate and is soluble at 0.5 M ammonium sulfate. The fractions containing target protein obtained from the IMAC column were subjected to a Phenyl Sepharose HP column pre-equilibrated with buffer A containing 0.5 M ammonium sulfate. The bound protein was eluted with a linear gradient of 0.5–0.0 M ammonium sulfate in buffer A. The peak fractions were analysed on 15% SDS–PAGE and pure fractions were pooled and dialysed overnight against buffer A to remove ammonium sulfate. The concentration of rSgrA_{28–153} was measured using a UV spectrophotometer (A_{280}) assuming a calculated absorption coefficient of 1.49 (ProtParam; Gasteiger *et al.*, 2005). A final yield of 6 mg protein was obtained from a 500 ml culture and was used for crystallization trials.

2.4. Crystallization of rSgrA_{28–288} and rSgrA_{28–153}

Crystallization experiments were carried out using the hanging-drop vapour-diffusion method with Hampton Research screens (Crystal Screen, Crystal Screen 2, PEG/Ion, PEG/Ion 2 and SaltRx) and home-made screens such as ammonium sulfate *versus* pH and polyethylene glycol *versus* pH for both the rSgrA_{28–288} and the rSgrA_{28–153} proteins. Extensive crystallization trials were carried out for both rSgrA_{28–288} and rSgrA_{28–153}. However, rSgrA_{28–288} did not yield any crystals. In the case of rSgrA_{28–153} thin needle-like crystals

were obtained with PEG 2000 monomethyl ether (PEG 2K MME) as a precipitant.

2.5. Data collection and processing

For X-ray data collection, native rSgrA_{28–153} crystals were soaked for 30 s in reservoir solution supplemented with 20% glycerol and flash-cooled in a nitrogen-gas stream at 100 K. Diffraction data were collected at our in-house data-collection facility using a MAR345 image-plate detector and a Bruker Microstar copper rotating-anode generator operating at 60 mA and 45 kV. A total of 180 frames were collected with an oscillation step of 1° and an exposure of 180 s per frame. The crystal-to-detector distance was maintained at 150 mm. The diffraction data were processed and scaled using *imosflm* (Battye *et al.*, 2011). The crystal parameters and data-collection statistics are summarized in Table 1.

2.6. Preparation of selenomethionine derivative of rSgrA_{28–153}

rSgrA_{28–153} contains four methionine residues. A selenomethionine derivative was prepared by growing cells with pET-28a(+)-sgrA_{28–153} plasmid in 1× M9 minimal medium (Sambrook *et al.*, 1989) supplemented with 50 µg ml^{−1} kanamycin. The incorporation of selenomethionine was carried out using the methionine-auxotroph strain of *E. coli* B834. For this, the pET-28a(+)-sgrA_{28–153} gene construct was transformed into competent B834 strain and expression was performed in 1× M9 minimal medium. SeMet-rSgrA_{28–153} was prepared in the following way: 10 ml LB medium supplemented with 50 µg ml^{−1} kanamycin was inoculated with B834 harbouring pET-28a(+)-sgrA. The overnight-grown culture was spun down at 4000 rev min^{−1} for 10 min, resuspended in 10 ml 1× M9 minimal medium and transferred to 1 l of 1× M9 minimal medium supplemented with kanamycin. After 1 h, 40 mg of selenomethionine was added to the 1 l culture. The culture was grown for an additional 9 h followed by overnight induction with 1 mM IPTG. Following this, harvesting the cells and purification of the protein were carried out in the same way as for the native protein. However, 5 mM β-mercaptoethanol was added to the buffer to avoid the oxidation of selenomethionine.

3. Results and discussion

The rSgrA_{28–288} protein consists of two regions: the N-terminal putative ligand-binding A domain and the C-terminal repetitive B domain. To facilitate purification, the rSgrA_{28–288} protein was expressed with an LEHHHHHH motif at its C-terminus. The purity of rSgrA_{28–288} used in the crystallization experiments was checked by 15% SDS–PAGE analysis and the 31 kDa protein migrated at around 43 kDa (Fig. 2). Also, multiple bands were observed below the 43 kDa band, indicating that the protein is likely to be degraded or unstable. A similar degradation pattern was observed in a family of Group B Streptococcal surface proteins of repetitive nature, namely Alpha-C and Rib (Wästfelt *et al.*, 1996), and recently in FbsA, the fibrinogen-binding protein (Ponnuraj & Ragunathan, 2011). Thus, in rSgrA_{28–288} the presence of the C-terminal repetitive B domain which consists of eight repeats of a Ser–Ser–Glu–Ser–Ser–Thr motif could be the reason for instability of the protein and its failure to crystallize.

In an attempt to crystallize the ligand-binding A domain (devoid of the B repeat region), rSgrA_{28–153} was cloned and expressed in *E. coli* BL21 (DE3) cells. The PCR-amplified sgrA segment 28–153 was inserted into the pET28a vector using its *Bam*HI and *Xho*I sites. The sequence upstream of the *Bam*HI site encodes a 6×His tag followed by additional residues which include a thrombin cleavage site. The

Figure 2

SDS–PAGE of purified rSgrA_{28–288} and rSgrA_{28–153}. Lane 1, molecular-weight marker (labelled in kDa); lane 2, final purified rSgrA_{28–288} (10 mg ml^{−1}); lane 3, rSgrA_{28–153} (8 mg ml^{−1}). 10 µl of protein mixed with 5 µl of loading dye was loaded in lanes 2 and 3.

Figure 3
Thin needle-like crystals of rSgrA₂₈₋₁₅₃ of *E. faecium*.

sequence downstream of the *XhoI* site encodes another 6×His tag followed by a stop codon. Thus, the rSgrA₂₈₋₁₅₃ construct contains 126 residues of SgrA plus 42 residues [34 residues at the N-terminus (MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS) and eight residues (LEHHHHHH) at the C-terminus] from the cloning vector. These tags were not removed prior to crystallization. Mass spectrometry confirmed the molecular mass of 20 kDa and the protein was found to migrate at around 25 kDa in 15% SDS-PAGE (Fig. 2). The protein was purified to homogeneity and screened for crystallization.

Initial crystals of rSgrA₂₈₋₁₅₃ appeared after 3 months using PEG 2K MME as a precipitant. Various optimization procedures such as the addition of reducing agents and salts to the reservoir solution and microseeding were tried in order to obtain diffraction-quality crystals. Finally, a drop consisting of 1 µl protein solution and 1 µl reservoir solution (40% PEG 2K MME, 100 mM HEPES pH 7.5) seeded with microseeds gave crystals that were suitable for data collection (Fig. 3). The drop was equilibrated against 1 ml reservoir solution and crystals were obtained in 2 d. A native data set was collected from a single crystal to a resolution of 3.3 Å. The crystal lattice was primitive monoclinic, with unit-cell parameters $a = 35.84$, $b = 56.35$, $c = 60.20$ Å, $\beta = 106.5^\circ$, and the space group was deduced to be $P2_1$. The values for the crystal volume per unit protein mass (V_M , Matthews coefficient) and solvent content were computed to be $2.94 \text{ Å}^3 \text{ Da}^{-1}$ and 58.14%, respectively, which correspond to one molecule per asymmetric unit. Alternatively, if there are two molecules in the asymmetric unit, the values are $1.47 \text{ Å}^3 \text{ Da}^{-1}$ and 16.35%, respectively (Matthews, 1968).

Based on the crystallization experiments of the rSgrA₂₈₋₂₈₈ and rSgrA₂₈₋₁₅₃ proteins it is evident that although both proteins were expressed with identical tags on their C-termini, the crystals were only obtained with rSgrA₂₈₋₁₅₃, which also has an N-terminal His tag. In the longer construct, rSgrA₂₈₋₂₈₈, the His tag is probably not intact because it was fused with the repetitive B domain which is likely to be unstable. Thus, the rSgrA₂₈₋₂₈₈ sample used for crystallization might contain different molecular species of rSgrA₂₈₋₂₈₈, each with a different number of B repeats. The protein heterogeneity, as observed by a laddering pattern in SDS-PAGE (Fig. 2), could be the reason for the failure of rSgrA₂₈₋₂₈₈ to crystallize even after several

rounds of purification and crystallization trials with various protein concentrations.

The crystallized construct rSgrA₂₈₋₁₅₃ does not have a homologous structure in the PDB and therefore determination of its structure by molecular replacement is not possible. It contains four methionine residues which can be modified to selenomethionine (SeMet) in order to resolve the phase problem using the multi-wavelength anomalous dispersion (MAD) or single-wavelength anomalous dispersion (SAD) techniques. The SeMet-substituted protein was prepared using a methionine-auxotroph strain in minimal medium (M9 medium) as described in §2.6. Crystals of SeMet-substituted rSgrA₂₈₋₁₅₃ were obtained under similar conditions to those used for the native protein, but were not of diffraction quality. However, the incorporation of selenomethionine into the crystal was detected using X-ray fluorescence spectroscopy at the synchrotron beamline. Further crystallization trials are under way in order to obtain diffraction-quality crystals of SeMet-incorporated rSgrA₂₈₋₁₅₃. A three-wavelength MAD data set or a single-wavelength SAD data set will be collected to solve the structure of the protein. If diffraction-quality SeMet-rSgrA₂₈₋₁₅₃ crystals are not obtained, halide-ion soaking and xenon derivatization will be attempted with native crystals to solve the phase problem. If these methods also fail, nuclear magnetic resonance (NMR) will be attempted to obtain a solution structure of rSgrA₂₈₋₁₅₃.

KP thanks the Department of Biotechnology (DBT), Government of India, for financial support in the form of a grant. RN thanks the University Grants Commission (UGC), India, for providing a fellowship.

References

- Arias, C. A. & Murray, B. E. (2012). *Nature Rev. Microbiol.* **16**, 266–278.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Donlan, R. M. (2002). *Emerg. Infect. Dis.* **8**, 881–890.
- Foster, T. J. & Höök, M. (1998). *Trends Microbiol.* **6**, 484–488.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
- Hay, E. D. (1991). *Cell Biology of Extracellular Matrix*. New York: Plenum Press.
- Hendrickx, A. P. A., van Luit-Asbroek, M. V., Schapendonk, C. M. E., van Wamel, W. J. B., Braat, J. C., Wijlans, L. C., Bonten, M. J. M. & Willems, R. J. L. (2009). *Infect. Immun.* **77**, 5097–5106.
- Hendrickx, A. P. A., van Wamel, W. J. B., Posthuma, G., Bonten, M. J. M. & Willems, R. J. L. (2007). *J. Bacteriol.* **189**, 8321–8332.
- Hendrickx, A. P. A., Willems, R. J. L., Bonten, M. J. M. & van Schaik, W. (2009). *Trends Microbiol.* **17**, 423–430.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Murray, B. E. (1990). *Clin. Microbiol. Rev.* **3**, 46–65.
- Murray, B. E. (2000). *N. Engl. J. Med.* **342**, 710–721.
- Paganelli, F. L., Willems, R. J. & Leavis, H. L. (2012). *Trends Microbiol.* **20**, 40–49.
- Patti, J. M. & Höök, M. (1994). *Curr. Opin. Cell Biol.* **6**, 752–758.
- Ponnuraj, K. & Ragunathan, P. (2011). *Protein J.* **30**, 159–166.
- Richards, M. J., Edwards, J. R., Culver, D. H. & Gaynes, R. P. (2000). *Infect. Control Hosp. Epidemiol.* **21**, 510–515.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.
- Wästfelt, M., Stålhammar-Carlén, M., Delisse, A. M., Cabezon, T. & Lindahl, G. (1996). *J. Biol. Chem.* **271**, 18892–18897.