Crystallization and preliminary crystallographic analysis of a surface antigen glycoprotein, SAG19, from *Eimeria tenella*

Nur Zazarina Ramly,^{a,b}
Sergey N. Rouzhenikov,^a
Svetlana E. Sedelnikova,^a
Patrick J. Baker,^a Yock-Ping
Chow,^c Kiew-Lian Wan,^{c,d} Sheila
Nathan^{c,d} and David W. Rice^{a*}

^aKrebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, England, ^bFaculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Bandar Baru Nilai, Negeri Sembilan DK, Malaysia, ^cSchool of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor DE, Malaysia, and ^dMalaysia Genome Institute, 43000 Kajang, Selangor DE, Malaysia

Correspondence e-mail: d.rice@sheffield.ac.uk

Received 11 September 2013

Accepted 29 October 2013

Coccidiosis in chickens is caused by the apicomplexan parasite *Eimeria tenella* and is thought to involve a role for a superfamily of more than 20 cysteine-rich surface antigen glycoproteins (SAGs) in host–parasite interactions. A representative member of the family, SAG19, has been overexpressed in *Escherichia coli*, purified and crystallized by the hanging-drop method of vapour diffusion using ammonium sulfate as the precipitant. Crystals of SAG19 diffracted to beyond 1.50 Å resolution and belonged to space group *I*4, with unit-cell parameters $a = b = 108.2$, $c = 37.5$ Å. Calculation of possible values of V_M suggests that there is a single molecule in the asymmetric unit.

1. Introduction

The phylum Apicomplexa represents a group of obligate intracellular parasites that include *Plasmodium*, *Toxoplasma* and *Eimeria*, the causative agents of malaria, toxoplasmosis and coccidiosis, respectively (Augustine, 2001; Madigan *et al.*, 2006; Ménard, 2007). Coccidiosis is an economically important intestinal disease that infects poultry with three species of the genus *Eimeria* (*E. tenella*, *E. maxima* and *E. acervulina*) contributing significantly to observed infections (Shirley *et al.*, 2004). Whilst the duration of an *Eimeria* infection is usually fairly short (4–6 d depending on the species), coccidiosis can be serious owing to the burden on the host of massive parasite production during the intracellular phase of the pathogen's life cycle (Morris & Gasser, 2006). Thus, the worldwide economic burden of *E. tenella* is estimated to be in excess of £2 billion per annum (Shirley *et al.*, 2004), with the financial impact of coccidiosis in the United Kingdom being approximately equivalent to 4.5% of the profits from sales of live birds (Williams, 1999; Shirley *et al.*, 2004). Various anti-parasitic drugs have been used in attempts to control coccidiosis, but the emergence of resistance has lessened the effectiveness of a number of the antiparasitic agents (Shirley *et al.*, 2007).

The life cycle of *Eimeria* is relatively complex and involves both endogenous and exogenous stages of development inside and outside the host, respectively (Jeurissen *et al.*, 1996). Whilst some aspects of the biology of these developmental stages are understood, the molecular basis of host–parasite interactions is largely undefined. In the related apicomplexan parasite *Toxoplasma gondii*, a superfamily of more than 160 cysteine-rich surface antigen glycoproteins (SAGs) with a molecular weight generally around 40 kDa and sharing no less than 24% sequence identity have been identified (Jung *et al.*, 2004). These proteins have been implicated in host–parasite interactions, where they are thought to be involved in the initial attachment of the parasite to the host (Lekutis *et al.*, 2001; Carruthers & Boothroyd, 2007). Common features of the sequence of these SAGs include an N-terminal signal sequence, 12 conserved cysteine residues which form six disulfide bridges and a glycosylphosphatidylinositol (GPI) anchor for membrane attachment (Jung *et al.*, 2004; Crawford *et al.*, 2009). Studies have established that the pattern of SAG expression varies dependent upon the stage of development with, for example, SAG1, BSR4 and sporoSAG being expressed in the tachyzoite, bradyzoite and sporozoite stages of growth, respectively. The three-dimensional

structures of each of these three SAGs have been determined at high resolution to show that the fold comprises two related domains each forming a β -sandwich and corresponding to the N-terminal (D1) and C-terminal (D2) segments of the sequence (He *et al.*, 2002; Crawford *et al.*, 2009, 2010). However, the *T. gondii* SAGs are thought to differ in quaternary structure, with both SAG1 and BSR4 being dimeric (He *et al.*, 2002; Crawford *et al.*, 2009), whereas sporoSAG is monomeric (Crawford *et al.*, 2010). Despite the similarities in overall structure, comparisons show that the molecular surfaces of these SAGs are quite distinct. Thus, whilst the surface of sporoSAG is predominantly negatively charged, that of SAG1 and BSR4 is more positive (He *et al.*, 2002; Crawford *et al.*, 2009, 2010). This has led to the suggestion that the different SAGs play distinct roles in the biology of *T. gondii* (Crawford *et al.*, 2009, 2010).

Transcriptome sequencing of *E. tenella* has led to the identification of a superfamily of cysteine-rich surface antigen glycoproteins, with more than 20 members having been identified (Tabarés *et al.*, 2004). Like the SAGs from *T. gondii*, the *E. tenella* SAGs possess an N-terminal signal sequence and a GPI anchor (Tabarés *et al.*, 2004). Sequence comparisons show that these cysteine-rich SAGs can be divided into two subfamilies (family A and family B) within which considerable sequence similarity can be seen (Tabarés *et al.*, 2004). However, sequence comparisons between the families show that they share little or no similarity and furthermore the sequences of *Eimeria* and *Toxoplasma* SAGs cannot be aligned with each other with any degree of confidence, so that the relationship of their structures is unclear. *E. tenella* SAGs are among the major surface molecules of the parasite and many of these are expressed during the development of second-generation merozoites, making them good targets for host innate and adaptive immunity immune responses. However, their precise role in infection remains ill defined (Tabarés *et al.*, 2004). A modelling study on *E. tenella* SAG1 has suggested that parasite invasion may be initiated by sporozoite attachment to the negatively charged proteoglycans on the surface of the host cell (Jahn *et al.*, 2009).

In order to contribute a better understanding of the role of these SAGs in *E. tenella* and to permit their structures to be compared with those from other apicomplexa, structural studies on representative *E. tenella* SAGs have been initiated. This report describes the first step in the structural investigation of one of the surface antigen glycoproteins, namely the crystallization and initial crystallographic analysis of the 26 kDa *E. tenella* protein SAG19.

2. Materials and methods

2.1. Cloning and overexpression

A construct of *E. tenella* SAG19 was made by amplifying the gene from a merozoite cDNA library using the oligonucleotides 5'-**CCATGG**CGGCCGACCAGACTTCTC-3' (forward) and 5'-**CTCGAG**TGCTTCCAATCCCCACAGAGCATT-3' (reverse) in which the oligonucleotides contained *Nco*I (**CCATGG**) and *Xho*I (**CTCGAG**) restriction-enzyme sites, respectively (Chow *et al.*, 2011). The purified DNA fragment (710 bp) was cloned into the pTZ57R/T cloning vector (Fermentas) and subsequently subcloned into the pET-32b(+) expression vector (Novagen) to facilitate expression by fusion, at the N-terminus of the coding sequence, of a cleavable Trx-His-S tag (Fig. 1). The construct was designed to express the tag followed by residues 159–391 of SAG19 together with an additional three-residue linker and a hexahistidine tag at the C-terminus of the SAG19 sequence. The plasmid was isolated and purified and its DNA sequence was confirmed. pET-32b(+)-SAG19 was transformed into *Escherichia coli* strain Rosetta gami 2 (DE3) for overexpression. This strain was chosen as it provides a suitable vehicle to facilitate the folding of proteins containing disulfide bridges, which it was anticipated would be present in SAG19, and to translate rare codons. A single colony of the transformant was inoculated into Luria–Bertani (LB) medium containing 100 $\mu\text{g ml}^{-1}$ carbenicillin and 34 $\mu\text{g ml}^{-1}$ chloramphenicol and the culture was grown overnight at 310 K in an incubator shaker at 250 rev min $^{-1}$. 10 ml of this culture was inoculated into 500 ml LB medium supplemented with carbenicillin and chloramphenicol as above. Cultures were grown at 310 K (250 rev min $^{-1}$) until an OD $_{600}$ of 0.6 was attained, at which point SAG19 expression was induced by the addition of 0.1 mM isopropyl β -D-1-thiogalactopyranoside and the culture was grown for an additional 20 h at 293 K (200 rev min $^{-1}$). The soluble protein was then analysed by SDS–PAGE and the overexpressed protein was identified as a large band at the expected molecular weight of the construct (43 kDa; Fig. 2).

2.2. Purification

For purification, the cell paste was suspended in 50 mM Tris–HCl pH 8.0 (5–10 ml per gram of cell paste) and the cells were disrupted using a sonicator (Soniprep 150, three cycles of 15 μm amplitude for 15 s). The soluble protein was separated from the cell debris by

```

1   MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGLTVAKLN
61  IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHSHH
      ↓
121 HHSSGLVPRSGMKETAAAKFERQHMDSPDLGTDDDDKAMAAAPDFSSALSLRSSTATSQ
181 QNSLSTNIFASGDVSPQTPTPPQADEKTEDCLAIINKLRSENLKDLLGTLAKAEDTEVTE
241 SLKAIEEPASPTAPKIAVTLAGSNVDTCSGEGANAKKYPGLVIPFPHDTEFNCLALI
301 QATYTAGLDHLKQSNFEPSTGTYDVENAPFNNVNASNVAFLLSKSKKVSCAATKDCCKAG
361 HDVLFCFYFIDPLRKEDKPFTAELYNALWGLEALEHHHHHHH

```

Figure 1

Sequence of the fusion construct for SAG19 including all of the associated tags. The arrow indicates the expected enterokinase cleavage site of the N-terminal Trx-His-S tag.

Table 1
Data-collection and processing statistics for *E. tenella* SAG19 crystals.

Values in parentheses are for the outermost resolution shell.

Space group	I4
Molecules per asymmetric unit	1
Unit-cell parameters (Å)	$a = b = 108.2$, $c = 37.5$
X-ray source	I04, Diamond Light Source
Detector	ADSC Q315r
Temperature (K)	100
Wavelength (Å)	0.92
Resolution (Å)	15.00–1.50 (1.53–1.50)
No. of observations	210483 (15178)
No. of unique reflections	34939 (2303)
Completeness (%)	99.9 (99.9)
Mean $I/\sigma(I)$	25.3 (2.8)
$R_{\text{merge}}^{\dagger}$	0.033 (0.508)
Multiplicity	6.0 (6.0)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of a unique reflection hkl and $\langle I(hkl) \rangle$ is the average over all symmetry-related observations of that reflection.

centrifugation at 24 500 rev min⁻¹ and 277 K for 10 min (JA-25.50, J-25 Avanti centrifuge). Subsequently, the protein was loaded onto a nickel-based resin (5 ml His-Trap HP cartridge, GE Healthcare), washed with 0.5 M NaCl, 50 mM Tris–HCl pH 8.0 and eluted with the same buffer containing 500 mM imidazole. The fractions containing SAG19 were concentrated using a Vivaspın concentrator fitted with a 30 kDa molecular-weight cutoff filter (Sartorius, Germany) and the buffer was exchanged to cleavage buffer (50 mM Tris–HCl pH 8.0, 1 mM CaCl₂). Subsequently, the protein was cleaved by recombinant enterokinase (Invitrogen, USA) to remove the N-terminal Trx–His–S tag according to the manufacturer's specifications but modifying the protocol to replace 1 U of enzyme per 20–50 µg of protein with 1 U of enzyme per 1.5 mg of protein. SAG19 was further purified using an anion-exchange resin (6 ml Resource Q column, GE Healthcare) and eluted using a gradient from 0 to 0.35 M NaCl in 50 mM Tris–HCl pH 8.0. SAG19 eluted from the column at approximately 0.1 M NaCl and fractions containing the protein were pooled. The protein concentrations were estimated by the method of Bradford (1976). Prior to crystallization, the sample was concentrated to 11 mg ml⁻¹ using a

Vivaspin concentrator with a 10 kDa molecular-weight cutoff filter (Sartorius, Germany).

2.3. Crystallization and preliminary X-ray analysis of SAG19

Initial crystallization trials were carried out using crystallization kits from The JCSG+ and PACT Suites (Qiagen, Germany) on a Matrix Hydra II (Thermo Fisher Scientific, USA) using the standard sitting-drop vapour-diffusion technique by adding 0.2 µl of the concentrated SAG19 in 50 mM Tris–HCl pH 8.0, 0.1 M NaCl to an equal volume of the precipitant and then equilibrating the drops against the same precipitant (100 µl) at 290 K. Whilst no crystals were initially observed, one of the conditions from The JCSG+ Suite (3.15 M ammonium sulfate, 0.1 M citrate buffer pH 5) looked promising and was chosen for further optimization by variation of the pH and precipitant concentrations using the hanging-drop vapour-diffusion technique.

For data collection, crystals were loop-mounted and placed in a cryosolution consisting of 2.6 M ammonium sulfate, 0.1 M citrate buffer pH 4.5, 20% glycerol for approximately 1 min at room temperature prior to cooling in a nitrogen-gas stream at 100 K (Oxford Cryosystems Cryostream). Data were collected from crystals of SAG19 using an ADSC Q315r detector on beamline I04 at the Diamond Light Source for a total of 150 images each at 1° oscillation per image.

3. Results and discussion

Recombinant SAG19 protein was expressed, purified to homogeneity and the N-terminal Trx–His–S tag was removed. Following cleavage, a protein band of ~26 kDa was observed by SDS–PAGE and this reduction in molecular weight is consistent with the complete removal of the N-terminal Trx–His–S tag. Approximately 8 mg cleaved SAG19 could be purified from 3 g cell paste. The purified cleaved SAG19 protein at a concentration of 11 mg ml⁻¹ and an estimated purity of at least 90% (Fig. 2) was used for crystallization. Initial screening was performed at 290 K and the SAG19 protein precipitated out. One of the conditions, 3.15 M ammonium sulfate, 0.1 M citrate buffer pH 5, looked promising and was further optimized. This resulted in the growth of small thin needles in the presence of 2.4–2.8 M ammonium sulfate at pH 4.5. These were further improved by the inclusion of an additional buffer exchange

Figure 2
SDS–PAGE (12%) analysis showing stages in purification of SAG19. Lane 1, protein markers (Mark1); lane 2, crude extract; lane 3, eluate from the Ni column; lane 4, the products of the cleavage of the SAG19 fusion protein by enterokinase; lanes 5–13, fractions 11, 13, 15, 18, 20, 22, 24, 26 and 30 following anion-exchange chromatography on a Resource Q column (GE Healthcare). Fractions 18–22 (lanes 8–10) were pooled to provide a sample of SAG19 for crystallization. Arrows on the left indicate the position and molecular weight (kDa) of selected markers. Arrows on the right indicate the position of the fusion construct and its cleaved products.

Figure 3
Crystals of SAG19 grown from 2.6 M ammonium sulfate, 0.1 M citrate buffer pH 4.5.

Figure 4

A representative 1° oscillation image of data collected from a crystal of SAG19 at the Diamond Light Source. The diffraction extends to a resolution of ~ 1.5 Å, which is indicated by the blue circle.

prior to crystallization to replace the buffer used during the anion-exchange chromatography (50 mM Tris-HCl pH 8.0) with 20 mM phosphate buffer pH 6.5. Following this modification to the purification protocol, crystals with maximum dimensions of $\sim 0.25 \times 0.03 \times 0.01$ mm were grown from 2.6–2.7 M ammonium sulfate in 0.1 M citrate buffer pH 4.5 (Fig. 3). Data collection was carried out to a resolution of 1.5 Å at the Diamond Light Source using an ADSC Q315r detector (Fig. 4).

The data were processed and scaled using *DENZO* and *SCALE-PAK*, respectively (Otwinowski & Minor, 1997). Analysis of the merging statistics indicated that the crystals belong to the body-centred tetragonal crystal system with unit-cell parameters $a = b = 108.2$, $c = 37.5$ Å. Analysis of the pattern of systematic absences clearly showed that the $00l$ reflections had an $l = 2n$ condition, confirming the space group as *I*4. Calculation of the possible values of V_M suggests that the crystal contains one molecule in the asymmetric unit, with a corresponding V_M of $2.11 \text{ Å}^3 \text{ Da}^{-1}$, equal to a solvent content of 41.8% (Matthews, 1976). The crystallographic parameters

and data-collection statistics are given in Table 1. Attempts to solve the structure by sulfur-SAD techniques are currently being pursued in the hope that a structure of SAG19 will shed light on the biological role of this protein superfamily in host invasion by *E. tenella*.

NZR would like to thank the Ministry of Higher Education of Malaysia and Islamic Science University of Malaysia for financial support. DWR, SN and KLW would like to thank the PMI-2 initiative of the British Council for funding. YPC, SN and KLW also acknowledge support from the Ministry of Science, Technology and Innovation, Malaysia. We acknowledge the Diamond Light Source for synchrotron beam time and thank Mark Williams for assistance in operating station I04.

References

- Augustine, P. C. (2001). *Int. J. Parasitol.* **31**, 1–8.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Carruthers, V. & Boothroyd, J. C. (2007). *Curr. Opin. Microbiol.* **10**, 83–89.
- Chow, Y.-P., Wan, K.-L., Blake, D. P., Tomley, F. & Nathan, S. (2011). *PLoS One*, **6**, e25233.
- Crawford, J., Grujic, O., Bruic, E., Czjzek, M., Grigg, M. E. & Boulanger, M. J. (2009). *J. Biol. Chem.* **284**, 9192–9198.
- Crawford, J., Lamb, E., Wasmuth, J., Grujic, O., Grigg, M. E. & Boulanger, M. J. (2010). *J. Biol. Chem.* **285**, 12063–12070.
- He, X.-L., Grigg, M. E., Boothroyd, J. C. & Garcia, K. C. (2002). *Nature Struct. Biol.* **9**, 606–611.
- Jahn, D., Matros, A., Bakulina, A. Y., Tiedemann, J., Schubert, U., Giersberg, M., Haehnel, S., Zoufal, K., Mock, H.-P. & Kipriyanov, S. M. (2009). *Parasitol. Res.* **105**, 655–668.
- Jeurissen, S. H. M., Janse, E. M., Vermeulen, A. N. & Vervelde, L. (1996). *Vet. Immunol. Immunopathol.* **54**, 231–238.
- Jung, C., Lee, C. Y.-F. & Grigg, M. E. (2004). *Int. J. Parasitol.* **34**, 285–296.
- Lekutis, C., Ferguson, D. J. P., Grigg, M. E., Camps, M. & Boothroyd, J. C. (2001). *Int. J. Parasitol.* **31**, 1285–1292.
- Madigan, M. T., Martinko, J. M. & Brock, T. D. (2006). *Brock Biology of Microorganisms*, 11th ed. Upper Saddle River: Pearson Prentice Hall.
- Matthews, B. W. (1976). *Annu. Rev. Phys. Chem.* **27**, 493.
- Ménard, R. (2007). *Curr. Opin. Microbiol.* **10**, 346–348.
- Morris, G. M. & Gasser, R. B. (2006). *Biotechnol. Adv.* **24**, 590–603.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Shirley, M. W., Ivens, A., Gruber, A., Madeira, A. M. B. N., Wan, K.-L., Dear, P. H. & Tomley, F. M. (2004). *Trends Parasitol.* **20**, 199–201.
- Shirley, M. W., Smith, A. L. & Blake, D. P. (2007). *Vaccine*, **25**, 5540–5547.
- Tabarés, E., Ferguson, D., Clark, J., Soon, P.-E., Wan, K.-L. & Tomley, F. (2004). *Mol. Biochem. Parasitol.* **135**, 123–132.
- Williams, R. B. (1999). *Int. J. Parasitol.* **29**, 1209–1229.