Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Ostertagia ostertagi* ASP-1: insights into disulfide-mediated cyclization and dimerization

Jimmy Borloo,[a]* Peter Geldhof,[a] Iris Peelaers,[a] Frederik Van Meulder,[a] Paul Ameloot,[b,c] Nico Callewaert,[b,c,d] Jozef Vercruysse,[a] Edwin Claerebout,[a] Sergei V. Strelkov[e] and Stephen D. Weeks[e]*

[a]Laboratory of Parasitology, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, B-9820 Merelbeke, Belgium, [b]Unit for Medical Biotechnology, Department for Molecular Biomedical Research, VIB, Technologiepark 927, B-9052 Ghent-Zwijnaarde, Belgium, [c]Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium, [d]Laboratory for Protein Biochemistry and Biomolecular Engineering, Department of Biochemistry and Microbiology, Ghent University, Belgium, and [e]Laboratory for Biocrystallography, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, O&N II, Herestraat 49 Box 822, B-3000 Leuven, Belgium

Correspondence e-mail: jimmy.borloo@ugent.be, stephen.weeks@pharm.kuleuven.be

Received 29 August 2012
Accepted 6 December 2012

PDB Reference: ASP-1, 4g2u

The cysteine-rich secretory/antigen 5/pathogenesis-related 1 (CAP) protein superfamily is composed of a functionally diverse group of members that are found in both eukaryotes and prokaryotes. The excretome/secretome of numerous helminths (parasitic nematodes) contains abundant amounts of CAP members termed activation-associated secreted proteins (ASPs). Although ASPs are necessary for the parasitic life cycle in the host, the current lack of structural and functional information limits both understanding of their actual role in host–parasite interactions and the development of new routes in controlling parasitic infections and diseases. Alleviating this knowledge gap, a 1.85 Å resolution structure of recombinantly produced *Oo*-ASP-1 from *Ostertagia ostertagi*, which is one of the most prevalent gastrointestinal parasites in cattle worldwide, was solved. Overall, *Oo*-ASP-1 displays the common hallmark architecture shared by all CAP-superfamily members, including the N-terminal CAP and C-terminal cysteine-rich domains, but it also reveals a number of highly peculiar features. In agreement with studies of the natively produced protein, the crystal structure shows that *Oo*-ASP-1 forms a stable dimer that has been found to be primarily maintained *via* an intermolecular disulfide bridge, hence the small interaction surface of only 306.8 Å^2. Moreover, unlike any other ASP described to date, an additional intramolecular disulfide bridge links the N- and C-termini of each monomer, thereby yielding a quasi-cyclic molecule. Taken together, the insights presented here form an initial step towards a better understanding of the actual biological role(s) that this ASP plays in host–parasite interactions. The structure is also essential to help to define the key regions of the protein suitable for development of ASP-based vaccines, which would enable the current issues surrounding anthelmintic resistance in the treatment of parasitic infections and diseases to be circumvented.

1. Introduction

The cysteine-rich secretory/antigen 5/pathogenesis-related 1 proteins, collectively termed CAP-superfamily proteins, are found in both eukaryotes and prokaryotes and are involved in processes as diverse as reproduction, cancer and immune regulation (reviewed in Gibbs *et al.*, 2008). Besides the CAP nomenclature, members are also often referred to as sperm-cell glycoprotein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS; Pfam PF00188) proteins, further reflecting the vast functional diversity in proteins that this superfamily harbours (Cantacessi *et al.*, 2009). Although the sequence identity between different CAP proteins can be low, the crystal structures of several members have recently been elucidated (Asojo, 2011; Asojo *et al.*, 2005; Fernández *et al.*, 1997; Guo *et al.*, 2005;

Henriksen *et al.*, 2001; Osman *et al.*, 2011; Serrano *et al.*, 2004; Shikamoto *et al.*, 2005; Wang *et al.*, 2005) and show a highly conserved architecture. These structures reveal that CAP proteins essentially have a two-domain organization consisting of (i) an approximately 20 kDa structurally conserved N-terminal CAP domain (also termed the SCP or PR-1 domain) that has an α-β-α sandwich fold (Fernández *et al.*, 1997) in which three antiparallel β-sheets are enclosed by four α-helices and that is commonly held together by several disulfide bridges and (ii) an approximately 6 kDa C-terminal cysteine-rich domain comprised of a conserved hinge region and a variable C-terminal extension domain. Although a limited number of CAP proteins have been found to lack a signal peptide, preserving them in specific intracellular compartments such as, for instance, the Golgi complex (Eberle *et al.*, 2002), CAP-superfamily members typically carry a signal sequence directing them to the extracellular space, where the presence of multiple dithioether bonds in both domains is likely to have an important role in stabilizing the overall fold of the protein.

Both of the constituent domains of CAP proteins contain conserved sequence motifs that are found in all members of this protein family. The N-terminal CAP domain has several signature motifs that can be used to identify superfamily members. The PROSITE database (Sigrist *et al.*, 2010) lists two such motifs, CAP1 [(GDER)(HR)(FYWH)(TVS)(QA)(LIVM)(LIVMA)W*xx*(STN)] and CAP2 [(LIVMFYH)(LIVMFY)*x*C(NQRHS)Y*x*(PARH)*x*(GL)N(LIVMFYWDN)], in which one of the given residues in parentheses is present and *x* can be any amino acid. In addition, although not adopted as part of the consensus motifs of the superfamily within PROSITE, two further motifs have been documented: CAP3 [(H)(N)*xx*(R)] and CAP4 [(G)(EQ)(N)(ILV)] (Gibbs *et al.*, 2008). The cysteine-rich C-terminal domain shows conserved cysteine spacing in the hinge region, most frequently as a $Cx_2Cx_{3-7}Cx_5C$ motif, although additional cysteines may be present in the sequence. Whereas the presence of the CAP and C-terminal domains allows the identification of CAP-superfamily members, the underlying role of the consensus motifs is poorly understood. The sequence properties of these motifs are often used for classification purposes (Osman *et al.*, 2011), particularly based on the occurrence of a conserved histidine residue in the CAP1 motif as well as the number of disulfide bridges that are distributed along the protein sequence (see §3.5 for further details). However, whether and how such divisions translate into differences at the functional level is still not understood.

The excretory/secretory (ES) products of helminths (a term used for parasitic worms) of the phylum Nematoda contain abundant amounts of CAP-superfamily proteins that are more commonly referred to as activation-associated secreted proteins or *Ancylostoma*-secreted proteins (ASPs), the latter reflecting the organism, the dog hookworm *Ancylostoma caninum*, in which these proteins were initially discovered (Hawdon *et al.*, 1996). Their localization to the ES protein fraction is most likely linked to their *in vivo* function, as ASPs are thought to play an important role in the transition from the free-living stage to the parasitic stage during invasion of the host (Datu *et al.*, 2008; Hawdon *et al.*, 1996; Moser *et al.*, 2005). Nevertheless, although their importance has been generally acknowledged, little is known of their actual roles in parasitic invasion and/or host avoidance. In contrast to the significant number of CAP protein structures that have been elucidated, to date only a limited number of three-dimensional protein structures of helminth ASPs have been determined (Asojo, 2011; Asojo *et al.*, 2005; Osman *et al.*, 2011). These structures hint at a role for ASPs and ASP-like proteins in a diverse range of activities, for example as antagonistic ligands of complement receptor 3 (CR3), preventing chemotaxin binding and thereby altering the immune cascade (Asojo *et al.*, 2005), and/or as an immunomodulator by blocking the release of hydrogen peroxide from activated neutrophils (Moyle *et al.*, 1994) and inhibiting the aggregation and adhesion of platelets (Del Valle *et al.*, 2003). However, conclusive evidence and insights related to their exact function(s) have yet to be obtained.

To date, ASPs have been found in three configurations: either as double-domain ASPs composed of two distinct but related CAP domains or as C-type or N-type single-domain ASPs, where the identity of the latter two depends on their homology to either the C- or the N-terminus of the double-domain ASPs, respectively (Geldhof *et al.*, 2003). In this report, we focus on ASP-1 from *Ostertagia ostertagi* (*Oo*-ASP-1), an N-type single-domain ASP and one of the most abundant proteins in the excretome/secretome of this widespread gastrointestinal helminth that is responsible for major losses in cattle productivity worldwide. Similar to other single-domain CAP proteins (Asojo, 2011), *Oo*-ASP-1 was also found to dimerize in solution (Meyvis *et al.*, 2007). Although the importance of this property remains unclear, we sought to explore the mechanism(s) behind the dimerization process and use this information to shed light on the rationale behind it. Here, we describe the high-resolution three-dimensional structure of a *Pichia pastoris*-produced version of *Oo*-ASP-1 that is biologically equivalent to the natively produced protein both in its N-glycosylation profile and its quaternary structure; that is, as a dimer (Meyvis *et al.*, 2007, 2008).

2. Materials and methods

2.1. Expression and purification of recombinant *Oo*-ASP-1

Oo-ASP-1 was expressed in *P. pastoris* as follows: its coding sequence was PCR-amplified from a previously described construct (Geldhof *et al.*, 2003) and subsequently inserted as a *Xho*I–*Not*I fragment into the *Pichia* expression vector pPIC9 (Invitrogen Ltd, Paisley, Scotland). The resulting expression plasmid, pPIC9-*Oo*ASP1, was used to transform *P. pastoris* GnM5 strain (Jacobs *et al.*, 2009) by electroporation. Individual clones growing on minimal medium plates were isolated and tested for secretion of *Oo*-ASP-1 by SDS–PAGE followed by Coomassie Brilliant Blue staining or rabbit anti-*Oo*-ASP-1 immunoblotting. The glycoform of the secreted proteins was assessed using DSA–FACE glycan profiling (Jacobs *et al.*,

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	0.98
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 75.18, b = 76.29, c = 81.53
Resolution range (Å)	37.59–1.85 (1.95–1.85)
Unique reflections	37078 (4489)
Reflections in the R_{free} set	1851
Multiplicity	3.1 (2.2)
Completeness† (%)	91.7 (77.6)
Mean $I/\sigma(I)$	8.9 (2.5)
R_{merge}‡ (%)	6.3 (41.5)
R_{meas}§ (%)	7.6 (53.2)
Wilson B factor (Å^2)	18.7
Refinement	
No. of non-H atoms	
Protein	3301
Carbohydrate	14
Solvent	458
Sulfate	5
Average B factor/r.m.s.d. (Å^2)	
Protein (aniso)¶	26.39/3.52
Carbohydrate (iso)	75.55/5.86
Sulfate (iso)	57.9/4.02
Solvent (iso)	34.81
Root-mean-square deviations from ideal geometry	
Bond lengths (Å)	0.08
Bond angles (°)	1.194
R_{work}/R_{free} (%)	19.17/24.20

† The completeness of the data set was compromised by the presence of ice rings. ‡ $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of reflection hkl. § $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $N(hkl)$ is the number of observations of reflection hkl. ¶ The protein-atom positions were refined using TLS-restrained anistropic refinement. TLS groups were defined using the *PHENIX* package (Adams *et al.*, 2010).

2009). To obtain secreted recombinant *Oo*-ASP-1 with a GalGlcNAcMan5 glycoform (Meyvis *et al.*, 2008), the GnM5 strain secreting *Oo*-ASP-1 was transformed with pGlyco-SwitchGalT-1 (Jacobs *et al.*, 2009). Individual clones were selected on YPD plates containing nourseothricin and tested for recombinant *Oo*-ASP-1 secretion and glycoform. A GalGnM5 glyco-engineered clone expressing *Oo*-ASP-1 was freshly grown on a plate and used to inoculate a shake-flask culture with BMGY medium. After 48 h growth at 301 K, the cells were pelleted by centrifugation for 5 min at 1519*g*, after which the cells were resuspended in BMMY and further grown at 301 K. Every 12 h extra methanol [0.5%(*v*/*v*)] was added to the culture and the cells were finally pelleted after 48 h of induction. The cell medium was harvested and filtered over a 0.2 µm membrane. Upon the addition of ammonium sulfate to 50% saturation at 277 K, *Oo*-ASP-1 precipitated and was concentrated in a pellet fraction by centrifugation at 18 000*g* for 15 min. This pellet was dissolved in 40 m*M* sodium acetate buffer pH 4.4 and the remaining ammonium sulfate was removed by gel filtration using a Sephadex G25 column (GE Healthcare Biosciences AB, Uppsala, Sweden). This fraction was then applied onto a SP-Sepharose column (GE Healthcare Biosciences AB) equilibrated in 40 m*M* sodium acetate buffer pH 4.4 and the bound *Oo*-ASP-1 was eluted employing a gradient to 1 *M* sodium chloride (NaCl) in the same buffer. Fractions containing recombinant *Oo*-ASP-1 were pooled, dialysed against 25 m*M* Tris–HCl pH 7.5 and loaded onto a Mono Q column (GE Healthcare Biosciences AB). *Oo*-ASP-1 was eluted from this column using a gradient to 1 *M* NaCl in the same buffer. Fractions containing recombinant *Oo*-ASP-1 were pooled and buffer-exchanged to PBS by gel filtration on a Sephadex G25-column (GE Healthcare Biosciences AB).

2.2. Gel electrophoresis and immunoblotting

Reducing and nonreducing one-dimensional SDS–PAGE was carried out according to Laemmli (1970): in the latter case reducing agent was excluded from the Laemmli sample-loading buffer and the samples were not heat denatured. When immunoblotting specifically for *Oo*-ASP-1, following SDS–PAGE adult *O. ostertagi* ES proteins (Geldhof *et al.*, 2003) were blotted from the polyacrylamide gel to a polyvinylidene fluoride (PVDF) membrane for 1 h at 1 mA per cm^2 of gel surface followed by a 1 h blocking step in PBS + 0.2%(*v*/*v*) Tween 80 at room temperature (RT) and incubation of the membrane with the primary antibody, *i.e.* specific bovine antibodies against *Oo*-ASP-1 [Meyvis *et al.*, 2007; 1:500 dilution in PBS + 0.2%(*v*/*v*) Tween 80], for 1 h at RT. The membrane was then extensively washed in PBS + 0.2%(*v*/*v*) Tween 80 followed by incubation in horseradish peroxidase (HRP) conjugated rabbit anti-bovine antibody solution [1:5000 dilution in PBS + 0.2%(*v*/*v*) Tween 80] for 1 h at RT, allowing subsequent chemiluminescent detection of *Oo*-ASP-1.

2.3. Analysis of N-glycosylation by PNGaseF treatment

N-Glycans were removed from purified recombinant *Oo*-ASP-1 by treatment with recombinant PNGaseF as described previously (Vanderschaeghe *et al.*, 2010). Briefly, 6 µg *Oo*-ASP-1 was incubated for 30 min in denaturation buffer (50 m*M* sodium phosphate pH 7, 0.5% SDS, 40 m*M* DTT) at 310 K. Subsequently, NP-40 (to a final concentration of 1%) and 200 units of PNGaseF were added. After incubation for 3 h, the samples were heat-denatured at 368 K and analyzed by SDS–PAGE. Untreated samples were treated identically except for the exclusion of PNGaseF.

2.4. Crystallization and data collection

Recombinant *Oo*-ASP-1 was concentrated to 13 mg ml^{-1} using Amicon regenerated cellulose-based centrifugal filter devices (Millipore Corporation, Billerica, Massachusetts, USA) with a 10 kDa molecular-mass cutoff. Following initial screening, optimized crystals were produced by mixing an equal volume of the protein solution with precipitant (100 m*M* bis-Tris pH 6.5, 100 m*M* ammonium sulfate, 25% PEG 3350) in a hanging-drop vapour-diffusion setup at 293 K; crystals typically appeared after 2 d. For data collection, the crystals were briefly transferred into a cryoprotectant solution composed of the precipitant solution with 20%(*v*/*v*) ethylene glycol added to it, followed by flash-cooling in liquid nitrogen. X-ray diffraction data were collected at 100 K using the PROXIMA 1 facilities (SOLEIL Synchrotron, Saint-Aubin, France) and data-collection statistics are given in Table 1.

2.5. Structure determination and refinement

The obtained data were processed using the *XDS* (Kabsch, 2010) and *SCALA* (Evans, 2006) software packages. Phasing of the data was performed by molecular replacement using a model based on the N-terminal CAP domain of *Na*-ASP-1 (32% sequence identity; PDB entry 3nt8; Asojo, 2011) and the program *BALBES* (Long *et al.*, 2008) employing the York Structural Biology Laboratory software server. The initial model was then applied as an input for complete model building using the *phenix.mr_rosetta* algorithm in *PHENIX* (DiMaio *et al.*, 2011). This model was further improved by iterative rounds of manual model building using *Coot* (Emsley *et al.*, 2010) and additional refinement using the *phenix.refine* package in *PHENIX* (Adams *et al.*, 2010). Factors reflecting the quality of the final structure are provided in Table 1. Atomic coordinates and structure factors were deposited in the Protein Data Bank (PDB) as entry 4g2u. Structural figures were prepared using the *PyMOL* molecular-graphics system (v.1.3; Schrödinger LLC). Structure-based sequence alignments of single-domain SCP/TAPS were generated manually based on the available structures in the Protein Data Bank. Secondary-structure prediction-based sequence alignments were prepared using the *DSSP* algorithm (available at the PDB website) and the *PSIPRED* software (Jones, 1999). Interaction-surface predictions were performed using *PISA* (Krissinel & Henrick, 2007).

3. Results and discussion

3.1. Recombinant production of *Oo*-ASP-1

As a secreted protein, it was hypothesized that *Oo*-ASP-1 would contain both disulfide bonds and N-linked glycans. Examination of the primary sequence shows that the protein contains 13 cysteine residues, allowing the possibility of up to six dithioether bonds. Comparison of the sequence with those of ASPs with known structures shows that ten of the cysteines are highly conserved and form disulfide bonds in the homologues. In addition, the native form of *Oo*-ASP-1 has previously been shown to be glycosylated (Meyvis *et al.*, 2008), specifically at Asn9 and Asn37. Consequently, in order to produce protein suitable for crystallographic studies with correct co-translational and post-translational modifications, we employed a modified strain of the methylotrophic yeast *P. pastoris* as a host for recombinant production. Following expression and purification of *Oo*-ASP-1, gel electrophoretic analysis verified that the protein was of adequate purity to permit subsequent crystallization trials (Supplementary Fig. S1[1]). Glycan profiling and PNGaseF treatment indicated that the majority of the recombinant protein had two identical hybrid-type N-glycan moieties per monomer (Asn9 and Asn37; Supplementary Fig. S1*a*), as found in the native protein (Meyvis *et al.*, 2008). Notably, however, a lower band was observed for PNGaseF-untreated recombinant *Oo*-ASP-1, as shown in Supplementary Fig. S1(*a*), which may represent a population of protein lacking one of the two N-glycans. Interestingly, Meyvis *et al.* (2007) previously reported that *Oo*-ASP-1 exists as a dimeric species, a finding that was corroborated here upon comparison of reducing and non-reducing electrophoretic migration patterns of recombinant *Oo*-ASP-1, additionally hinting towards the disulfide bridge-based character of the dimer (Supplementary Fig. S1*b*).

3.2. Structure of *Oo*-ASP-1

In our hands, the recombinant *Oo*-ASP-1 readily crystallized, yielding suitably diffracting crystals (Table 1). Although it bears low sequence identity to other parasitic ASPs, the *Oo*-ASP-1 structure was solved by molecular replacement using a model based on the N-terminal CAP domain of *Na*-ASP-1 (PDB entry 3nt8; Asojo, 2011). The structure contains two equivalent molecules of *Oo*-ASP-1 (r.m.s.d. of 0.314 Å) in the asymmetric unit. The first monosaccharide of a single carbohydrate chain is observed for only one of the monomers in the crystal structure (Fig. 1*a*). The quality of the electron-density map for this GlcNAc residue is considerably poorer than the surrounding protein density, a result that is likely to reflect the high mobility commonly associated with carbohydrate chains. Indeed, there is no detectable density for the equivalent residue in the other protein monomer or for the second putative N-glycosylation site of both protein chains (Asn37). For the latter, the sites are situated in a predicted loop region (residues 31–60) of the protein that is partially disordered in both chains in the crystal.

The structure of each individual monomer shows the classical architecture of CAP-superfamily members encompassing a CAP domain and a hinge region (Fig. 1*b*). In addition to these structurally conserved regions, a single α-helix is observed at the C-terminal end of the protein (α-helix 5; Figs. 1*b* and 2*a*), a feature that to our knowledge is unique to *Oo*-ASP-1 and may confer a specific functional role to this protein.

The CAP domain displays the highly conserved α-β-α sandwich fold (Fernández *et al.*, 1997) representative of all CAP-superfamily members (Fig. 1*b*). Overall, the secondary-structural elements are arranged in the order α_1–α_2–β_1–α_3–α_4–β_2–β_3–β_4–β_5–α_5, where the core CAP domain (α_1–α_2–β_1–α_3–α_4) is stabilized by three dithioether bonds, as shown in Figs. 1(*c*) and 2(*a*). Disulfide bridge 3, which cross-links β-strands 2 and 3, is conserved in almost all CAP-superfamily members. Similarly, dithioether bond 2, which links β-sheet 2 to the loop C-terminally adjacent to α-helix 2, is conserved in a large number of CAP proteins, although it is absent in the venom proteins and a specific subgroup of ASPs (*i.e.* group 2 ASPs; see §3.5). Both of these disulfide bridges are located in the α-β-α core and are most likely to be involved in stabilization of the structure. Dithioether bond 1, which links the N-terminus of α-helix 2 to the N-terminal loop of the protein, acts in tethering the N-terminus in close proximity to α-helix 1, thereby stabilizing the loop/3_{10}-helix region between the latter and Cys4 (Figs. 1*a* and 1*c*). Although displaying some diversity

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MH5075). Services for accessing this material are described at the back of the journal.

across ASPs, this loop has a P$x_{4–5}$S/T consensus sequence; *Oo*-ASP-1 and *Teladorsagia circumcincta* ASP are the only known ASP members that incorporate an N-glycosylation acceptor sequence in this region (Fig. 2*a* and Supplementary Fig. S2).

Comparative analysis of ASPs with known structures, including that of *Oo*-ASP-1, allows the mapping of four previously identified CAP-specific conserved sequence motifs (Gibbs *et al.*, 2008) onto the reported sequences and structures (Figs. 2*a* and 2*b*). Although the established motifs are located within the core of the CAP domain, there is considerable sequence diversity amongst species. More specifically, the established motifs employed by the PROSITE database (CAP1 and CAP2) bear significant shortcomings, particularly

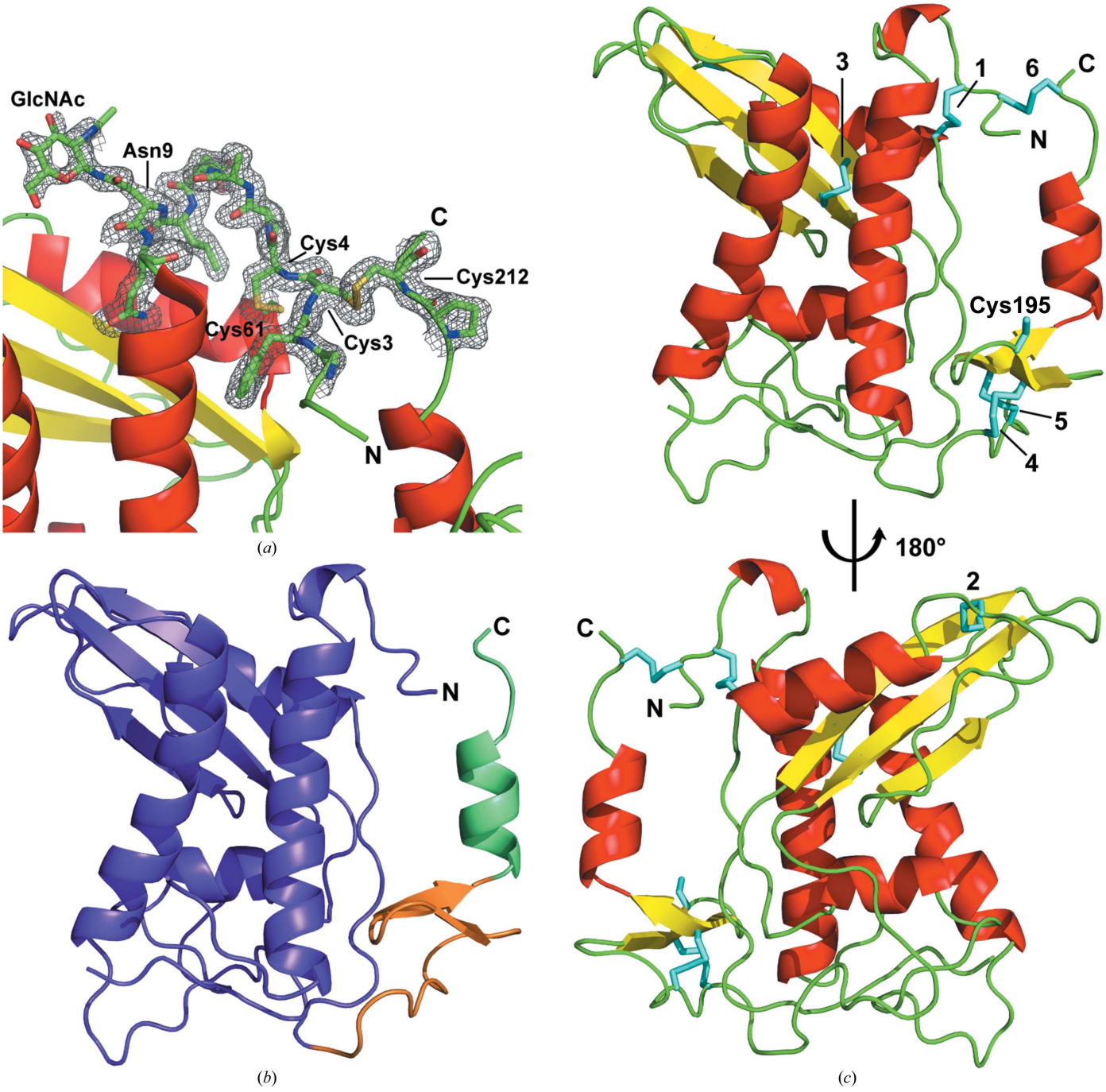

Figure 1
Structure of the *Oo*-ASP-1 monomer. (*a*) Map quality of the *Oo*-ASP-1 termini. A $2F_o - F_c$ map of the only observed *N*-acetylglucosamine (GlcNAc) moiety and residues 1–10 and 211–212 of chain *B* is shown contoured at 1σ. The corresponding residues are shown in stick representation, while the remaining structure is shown as ribbons. (*b*) Ribbon diagram of *Oo*-ASP-1 in which the protein has been divided into three regions based on structural homology to other members of the CAP superfamily: the CAP domain (purple), the hinge region (orange) and a C-terminal nonconserved region (green). (*c*) The disulfide-bridge pattern of *Oo*-ASP-1. The side chains of all cysteine residues are shown as stick models (cyan). The disulfide bonds are numbered according to the order shown in Fig. 2(*a*).

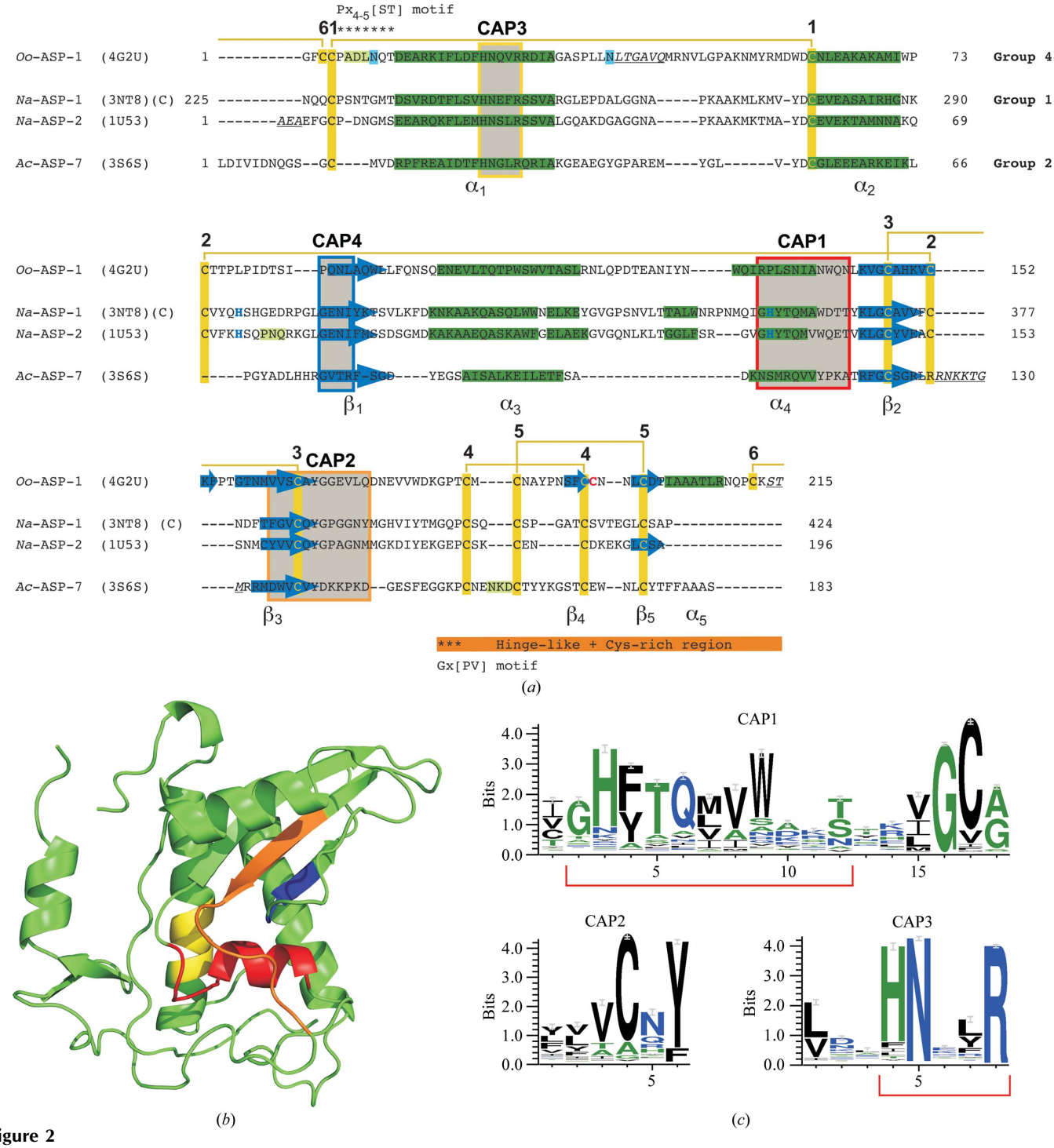

Figure 2
Sequence features of *Oo*-ASP-1. (*a*) Comparison of sequence and structural features of ASP proteins with known structures obtained by structural alignment of *Oo*-ASP-1, *N. americanus* ASP-1 (*Na*-ASP-1; PDB entry 3nt8), *N. americanus* ASP-2 (*Na*-ASP-2; PDB entry 1u53) and *A. caninum* ASP-7 (*Ac*-ASP-7; PDB entry 3s6s). *DSSP*-assigned secondary structures are shaded in dark green (α-helix), dark blue (β-strands) and light green (3_{10}-helices). Unresolved regions are shown in italics and underlined. Whereas *Na*-ASP-2 has a single monomer in the asymmetric unit, for *Oo*-ASP-1, *Na*-ASP-1 and *Ac*-ASP-7 assignments for chains *B*, *A* (C-terminal CAP domain) and *B* of the two available chains in the asymmetric units are shown, respectively. CAP motifs 1–4 are shown in grey boxes with outline colours corresponding to those in (*b*). Cysteine residues are in yellow and are numbered above according to the disulfide bridge to which they belong. Conserved histidine residues (in group 1 ASPs) are highlighted in dark blue. For *Oo*-ASP-1 the cysteine residue involved in dimerization (Cys195) is shown in red. The two N-glycosylated residues (Asn9 and Asn37) are shaded in light blue. Regions covered by the P*x*$_{4–5}$(ST) and G*x*(PV) motifs are indicated by asterisks. (*b*) Ribbon diagram of *Oo*-ASP-1 with CAP motifs 1, 2, 3 and 4 highlighted in red, orange, yellow and blue, respectively. (*c*) Sequence-logo representation of the consensus residues for metazoan CAP1, CAP2 and CAP3 motifs. The stack height, measured in bits, corresponds to the conservation at that position (where the maximum value for proteins is approximately 4.32). Metazoan-specific CAP sequences were extracted from the set composing Pfam family PF00188 (Punta *et al.*, 2012) and initially aligned using *MUSCLE* (Edgar, 2004). The alignment was manually curated using *Jalview* (Waterhouse *et al.*, 2009) and then employed as a basis for creating consensus logos through *WebLogo* 3.3 (Crooks *et al.*, 2004). Underlined residues for CAP1 and CAP3 correspond to the motifs currently recognized by PROSITE.

as they fail to recognize a significant number of ASPs, including *Oo*-ASP-1, as members of the CAP superfamily. Through analysis of known metazoan CAP proteins, we were able to distil more refined and expanded motifs for both the CAP1 and the non-adopted CAP3 motifs. Particularly, the currently used CAP1 motif, which covers the well conserved α-helix 4, identifies only 80% of the nonredundant CAP protein structures in the PDB. Taking the available structures, including that of *Oo*-ASP-1, into consideration, the CAP1 motif can be expanded to (GDERNK)(HRSPE)(FYWHLM)*xx*(TVSRA)(QANL)(LIVMA)(LIVMA)(WNYS)*xx*(STNAV)(VILMFA)G(CV)(AGYS). This pattern provides extra diversity to the N-terminus to accommodate the sequence variability observed in this region across species, but additionally includes extremely conserved glycine and cysteine residues at the C-terminal end of the motif that offset the nonspecificity introduced by the first change (Fig. 2*c*). These latter residues are found in the middle of β-strand 2, where the cysteine is part of disulfide bridge 3. Analysis of the structure suggests that the small (or absent) side chains of the residues on either side of this cysteine permit α-helix 2 to pack closer to the β-sheet. The CAP3 motif, which has previously been documented (Gibbs *et al.*, 2008) but has not been adopted by PROSITE, serves well in recognizing CAP proteins since it is located in the structurally conserved N-terminal α-helix 1.

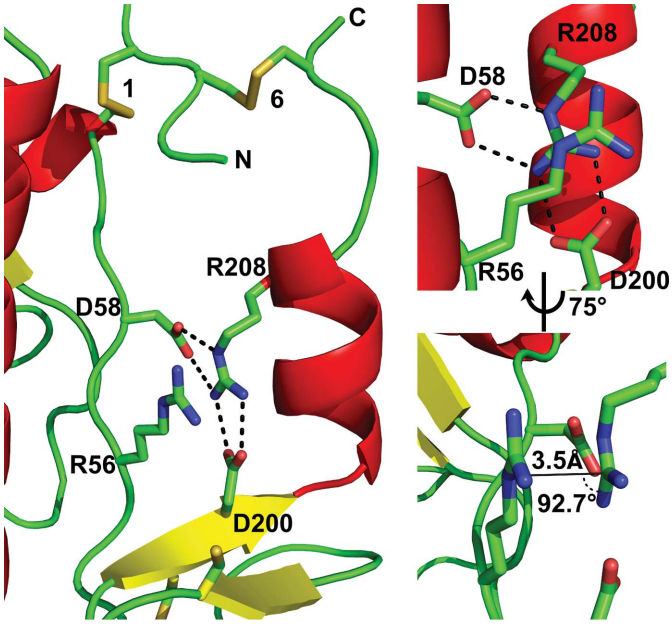

Figure 3
Interactions observed at the C-terminus of *Oo*-ASP-1. (*a*) The C-terminus of *Oo*-ASP-1 is tethered to the N-terminus by a disulfide bond. Ribbon diagram of *Oo*-ASP-1 in which the two disulfide bonds are shown as stick models (yellow) and numbered according to Fig. 1. Residues mediating additional contact between the terminal helix and the CAP and hinge regions of *Oo*-ASP-1, as outlined in the text, are shown as stick models. The noncovalent interactions between the residues are shown as black dashed lines. (*b*) Two views of the planar stacking observed for Arg56 and Arg208. Top panel, view along the axis orthogonal to that of the two guanidinium groups. Lower panel, side view of the parallel stacking interaction.

Despite its short length, this motif contains three well conserved residues (Fig. 2*c*) which, combined with the additionally conserved L/V that we observe three residues upstream in the metazoan-wide alignment (Fig. 2*c*), are all found on the same face of α-helix 1 pointing towards the core of the CAP domain. Like CAP1, we therefore propose that the CAP3 motif be extended to (LV)*xx*HN*xx*R. Contrasting with the analyses of the CAP1 and CAP3 motifs, sequence alignment suggests that only a smaller region of the currently employed CAP2 motif is actually conserved amongst metazoans (Fig. 2*c*), where the central cysteine is part of the β-sheet stabilizing disulfide bridge 3 (Figs. 1*c* and 2*a*). Finally, our analyses also confirmed the presence of the CAP4 motif (data not shown) that is part of β-strand 2 (Figs. 2*a* and 2*b*); however, it could not be extended beyond this pattern of four residues.

In comparison to the relatively conserved nature of the CAP domain, the C-terminal hinge region initiated by the G*x*(PV) motif preceding dithioether bond 4 (Fig. 2*a*) displays significantly higher levels of sequence flexibility, mainly owing to the frequent loops and the general lack of substantial secondary-structural elements. In terms of stabilization of the cysteine-rich region, *Oo*-ASP-1 contains two intra-domain dithioether bonds arranged in a C*x*C*x*$_7$C*x*$_4$C pattern (disulfide bridges 4 and 5; Figs. 1*c* and 2*a*), thereby largely conforming to the generally observed hinge-like sequence of CAP-superfamily members (Gibbs *et al.*, 2008).

3.3. An additional intramolecular dithioether bond in *Oo*-ASP-1 yields a cyclic molecule

Apart from the five hallmark intramolecular disulfide bonds present in *Oo*-ASP-1 and common to CAP-superfamily members, an additional dithioether bond was revealed in the protein crystal structure in which cysteine residues 3 and 212 are oxidized (Fig. 1*a*). The formation of a disulfide bond between these two residues, which are in close proximity to the N- and C-termini, results in a quasi-cyclic molecule (dithioether bond 6; Figs. 1*c* and 2*a*). Given the fact that *Oo*-ASP-1 is secreted *in vivo* by *O. ostertagi* during the infection process in the host, thereby hinting at an intimate involvement in host–parasite interplay, a means of protection for the molecule in this harsh environment would be desirable. Although a relatively rare event, protein cyclization is known to be an efficient route towards stabilization and a lower susceptibility towards proteolysis (Conlan *et al.*, 2010), and mechanisms as diverse as peptide bonding, amino-acid ligation, enzymatic conversion-driven cyclization and disulfide bonding have been revealed to be adopted by proteins seeking increased stability (Conlan *et al.*, 2010). For these reasons, it seems plausible that the introduction of an additional intramolecular dithioether bond, as shown here for *Oo*-ASP-1, contributes substantially to permitting this protein to maintain activity in a hostile environment such as the host–parasite interface.

In addition to a putative role in improving the overall stability of the protein, disulfide bond 6 also stabilizes the

Table 2
PISA analysis of currently available ASP structures.

Protein	PDB code	Molecules in asymmetric unit	Interface area† (Å^2)	Interface area as percentage of total
Oo-ASP-1	4g2u	2	306.8	2.8
Na-ASP-1‡	3nt8	2	486.0	5.4
Na-ASP-2	1u53	1	485.7	4.9
Ac-ASP-7	3s6s	2	418.2	4.8
	3s6u	2	470.1	5.2
	3s6v	2	423.8	4.8

† The interface surface areas for *Oo*-ASP-1 and *Na*-ASP-1 were calculated for the disulfide-linked and peptide-linked domains, respectively. For the remaining structures, the largest surface area is reported. ‡ Structure of the two-CAP-domain *Na*-ASP-1. To determine the interaction surface between the two domains, the linker was removed from the PDB file and each domain was treated as an individual chain.

position of the C-terminal α-helix 5 (Figs. 1 and 3). This structural element is to the best of our knowledge unique to *Oo*-ASP-1 and may have some functional importance. Utilizing a single arginine residue, the C-terminal α-helix adopts only two additional interactions: one with the CAP domain (Arg208–Asp58 salt bridge) and another with the hinge region (Arg208–Asp200 salt bridge). This interaction appears to be strained, as Arg208 is found in close proximity to Arg56 of the CAP domain (Fig. 3*b*). However, arginine pairs have been observed for other proteins in which, as observed in the *Oo*-ASP-1 structure, the two planar guanidinium groups are parallel (Neves *et al.*, 2012). In *Oo*-ASP-1 the N$^{\varepsilon}$ group of Arg56 sits along an orthogonal axis to the C$^{\zeta}$ of Arg208 (Fig. 3*b*), with a distance separating the two atoms that is similar to that observed for arginine and aromatic side-chain stacking (Flocco & Mowbray, 1994). The coordination of the Arg208 with the aforementioned aspartate residues and the solvent exposure of both arginines are in good agreement with reported strategies used to neutralize the electrostatic charge of this like pair of amino acids in other proteins (Neves *et al.*, 2012).

Whereas the possible benefits in terms of protein stability provided by the extra cyclizing dithioether bond in *Oo*-ASP-1 can be taken at face value, there may be more than meets the eye, as at the level of protein function it has been proposed that the usually more flexible C-terminal cysteine-rich region of ASPs mainly governs enzymatic activity and/or binding-partner selectivity (Osman *et al.*, 2011). Even though these postulations await experimental proof, it is likely that the covalent restraint of the C-terminal region of *Oo*-ASP1 and the presence of α-helix 5 would suggest an activity or binding partner that is considerably different from other ASPs.

3.4. The quaternary structure of *Oo*-ASP-1 reveals the mechanism of dimerization

The fact that the *Oo*-ASP-1 sequence bears an uneven number of cysteine residues (a total of 13), together with the observation that electrophoretic analysis under nonreducing conditions reveals *Oo*-ASP-1 migrating as a dimer (Supplementary Fig. S1*b*), led us to further investigate the possible occurrence of an inter-monomer disulfide-linked dimeric

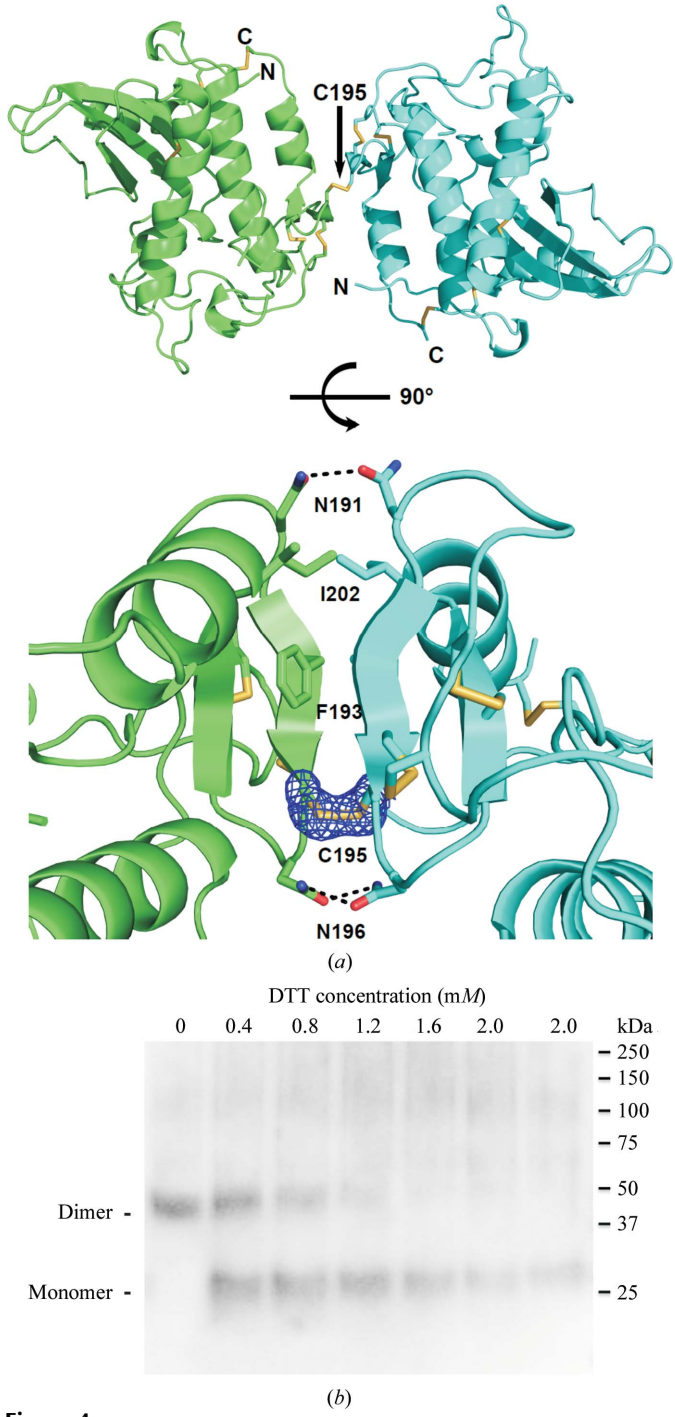

Figure 4
Disulfide-linked dimeric conformation of *Oo*-ASP-1. (*a*) Top panel, a ribbon diagram of the *Oo*-ASP-1 dimer in which the intermolecular dithioether bond between monomer *A* (green) and monomer *B* (cyan) is shown as a stick model. Lower panel, the dimer interface is stabilized by a limited number of interactions. The principal residues identified by *PISA* (Krissinel & Henrick, 2007) as involved in the dimer interface are shown as a stick model. A $2F_o - F_c$ map is shown for the intermolecular disulfide bond contoured at 1σ. (*b*) Native *Oo*-ASP-1 forms a disulfide-bridged dimer. The *Oo*-ASP-1 dimer-to-monomer transition was monitored by the addition of dithiotreitol (DTT) to separate aliquots of adult *O. ostertagi* ES protein samples in the concentration range 0–2 m*M* in increments of 400 µ*M* without concomitant heating. One final 2 m*M* DTT-treated aliquot was subjected to a heating step at 368 K for 5 min (rightmost lane). Subsequently, the samples were analysed by SDS–PAGE followed by *Oo*-ASP-1-specific immunoblotting.

species. Conclusive evidence regarding the quaternary-structural organization was obtained after inspection of the *Oo*-ASP-1 crystal structure (Fig. 4*a*). Interestingly, a dimer interface was identified that occurs between the two chains in the asymmetric unit and is primarily found on a sole intermolecular disulfide bridge based on covalent linkage of Cys195, located in the hinge region of the protein (Fig. 2*a*), from both monomers. The resultant dimer has a remarkably small interface area of 306.8 Å^2 (Table 2) with limited additional interactions (Fig. 4*b*), an arrangement that most likely gives rise to substantial rotational freedom between the two monomers in solution. In order to verify that the observed disulfide-linked dimer conformation is physiologically relevant, we sought to demonstrate its presence in the *O. ostertagi* excretion/secretion fraction. As Fig. 4(*c*) demonstrates, not only is *Oo*-ASP-1 present in the ES fraction exclusively as a dimer, but it is also extremely sensitive towards reducing agents, as revealed in reduction assays, in which dithiotreitol concentrations as low as 400 µ*M* sufficed to yield an approximately equivalent monomer:dimer ratio of native *Oo*-ASP-1, confirming the high accessibility of the key disulfide bond involved in dimer formation (Fig. 4*a*).

Single-domain ASPs from a number of different organisms have been reported to form dimers in solution (Asojo, 2011; Meyvis *et al.*, 2007). In our study, both native and recombinant *Oo*-ASP-1 demonstrated equivalent propensities towards dimerization. Whereas previous reports of specific dimerization interfaces have been based purely on predictions and modelling, by assuming maximization of an interaction interface area, usually through polar interactions and hydrophobic residue shielding, as the driving force in dimerization (Osman *et al.*, 2011; Serrano *et al.*, 2004), we here provide to our knowledge the first experimentally proven mechanism of CAP protein dimer assembly based on intermolecular disulfide bonding. With the exception of the *T. circumcincta* ASP (Supplementary Fig. S2), no other CAP protein has been shown to contain an equivalent cysteine in the hinge region or an additional free cysteine elsewhere in the sequence, thereby suggesting that this mechanism and the resulting dimer conformation is unique to the ASPs of these two species. Even so, *PISA* analyses of possible interfaces in the other reported ASP structures identify regions that have equally small interaction surfaces to that observed for *Oo*-ASP-1 (Table 2), although occurring between very different regions of the individual proteins. Comparatively, this suggests that these proteins consistently appear to maximize their solvent-exposed surface area, a property that is possibly necessary for function.

Interestingly, the parallels between the structure of dimeric *Oo*-ASP-1 and that of the double-domain *Na*-ASP-1 (PDB entry 3nt8; Asojo, 2011) are noteworthy. Whereas the *Oo*-ASP-1 dimer is based on a single high-rotational-freedom dithioether bond, the crystal structure of *Na*-ASP-1 reveals two CAP domains linked by means of a quasi-linear stretch of approximately 14 amino acids. As mentioned earlier, the interface surface between the two CAP domains in *Na*-ASP-1 is small (Table 2), suggesting that in solution the domain orientations may not be equivalent to that observed in the crystal.

Even though information concerning function and/or binding partners is still scarce for ASPs, the large solvent-exposed area at the dimer interface and the substantial inter-monomer rotational freedom highlighted by the *Oo*-ASP-1 structure may play a pivotal role in the true function of this protein. The exposure of key epitopes, in combination with inter-monomer flexibility, may govern ligand binding and/or conversion. From a functional point of view it is interesting to consider GAPR-1, a CAP-superfamily member that is found to partially dimerize in solution. This protein carries a number of catalytically active serine residues (Ser55 and Ser71) and other highly conserved residues (mainly His54, Glu65, Glu86 and His103) (Serrano *et al.*, 2004). From its predicted dimer structure it was postulated that upon dimerization a previously incomplete catalytic triad is restored, resulting in putative metalloprotease or serine protease activity (Serrano *et al.*, 2004). Whereas other ASPs may qualify as proteases, *Oo*-ASP-1 lacks the abovementioned key residues (Fig. 2*a*), thus rendering such activity highly unlikely. The true function or activity of *Oo*-ASP-1, and of activation-associated secreted proteins in general, remains elusive. Illustrating the tedious nature of the functional assessment of CAP proteins, to date Tex31 and tablysin-15, the former a substrate-specific endoprotease isolated from the cone snail *Conus textile* and the latter an integrin inhibitor and anti-inflammatory scavenger of fatty acids, are the only CAP-superfamily members that have had their functions fully characterized (Ma *et al.*, 2011; Milne *et al.*, 2003; Xu *et al.*, 2012).

3.5. ASP classification system

The sequence diversity observed among CAP-superfamily members is staggering. From a functional point of view, this is reflected in their involvement in numerous processes that include, but are not limited to, reproduction, development, immune function and cancer (Gibbs *et al.*, 2008). At the primary-structure level this degeneracy is even observed for the CAP motifs commonly used to identify members of this protein family, thus rendering software-driven protein classification and function prediction unreliable. However, the fact of the matter remains that all CAP-superfamily proteins, as shown by *Oo*-ASP-1, share a common α-β-α sandwich fold in their tertiary structure (Fernández *et al.*, 1997). Apart from this specific trait, activation-associated secreted proteins are known to contain two or three conserved disulfide bridges that stabilize the CAP domain (disulfide bridges 1–3; Figs. 1*c* and 2*a*), as well as two others that are found in the hinge region (disulfide bridges 4 and 5). Osman *et al.* (2011) recently introduced an ASP classification system consisting of three different groups based on the presence of the aforementioned disulfide bridges and the occurrence of two conserved histidine residues in the CAP domain. As a rule of thumb ASPs from groups 1 and 3 all bear the five signature disulfide bridges, whereas those belonging to group 2 lack dithioether bond 2. Distinction between groups 1 and 3, which are

phylogenetically closely related (Cantacessi *et al.*, 2009), is made by means of the presence of two conserved histidines, one prior to the CAP4 motif and the other part of CAP1 (Fig. 2*a*), whereas group 3 members are devoid of these residues. These two conserved histidines have been found to orient in a way that allows divalent metal binding (Asojo, 2011), upon which the structure may resemble a protease active site (Henriksen *et al.*, 2001; Shikamoto *et al.*, 2005); however, experimental proof of such enzymatic activity is still lacking. Interestingly, and for reasons of completeness, it should be noted that group 2 ASPs also lack these two histidines together with an additional stretch of amino acids in the loop preceding the CAP1 motif (Osman *et al.*, 2011). Fig. 2(*a*) shows an alignment per group based on secondary-structural data derived from currently available ASP structures, particularly *Na*-ASP-2 (PDB entry 1u53; Asojo *et al.*, 2005), the C-terminal moiety of the double-domain *Na*-ASP-1 (PDB entry 3nt8; Asojo, 2011), *Ac*-ASP-7 (PDB entry 3s6s; Osman *et al.*, 2011) and *Oo*-ASP-1 (PDB entry 4g2u; this work). Owing to the lack of experimentally obtained structural data, group 3 is not represented in this figure. However, a software-driven (*PSIPRED*) secondary structure-based alignment using a wider panel of ASPs and other CAP-superfamily proteins from all groups is provided in Supplementary Fig. S2.

Transposing the currently established group-specific characteristics to the *Oo*-ASP-1 crystal structure, we note the absence of the conserved histidine residues, thereby excluding it from group 1. Consequently, since *Oo*-ASP-1 does in fact bear dithioether bond 2 it would presumably be classified as a group 3 CAP protein. However, in addition to the five hallmark dithioether bonds, an additional dithioether bond was revealed to link its N- and C-termini, yielding a quasi-cyclic molecule (dithioether bond 6; Figs. 1*b* and 2*a*). Since the absence of a disulfide bridge in ASPs justifies the introduction of a separate group, we believe that the addition of a fourth group would accommodate the peculiar features of *Oo*-ASP-1, as discussed in this paper, and possibly those of *T. circumcincta* ASP (*Tc*-ASP; GenBank accession No. CBJ15404.1), which demonstrates the same cysteine pattern and displays 75% sequence identity to *Oo*-ASP-1 (Supplementary Fig. S2).

4. Conclusions and future perspectives

CAP-superfamily members are commonly recognized as cysteine-rich secretory proteins in which disulfide bonds play an important role in the tertiary structure. For *Oo*-ASP-1 these dithioether bonds are also essential for quaternary assembly and possibly proteolytic stability. Whereas to date reports on the dimerization interface of CAP proteins have been based on predictions and modelling, we have experimentally demonstrated that the *Oo*-ASP-1 dimer is based on a single intermolecular disulfide bridge, in agreement with the crystal structure (Fig. 4*a*). Additionally, the structure of *Oo*-ASP-1 reveals an intramolecular disulfide bridge that links the N- and C-termini, resulting in a quasi-cyclic conformation for each monomer that may render it more resistant to proteolysis. The cysteines involved in these two events only appear to be conserved in one other ASP protein (*T. circumcincta* ASP), suggesting that they confer specific functional properties to these two proteins that differ from other members of the ASP family. Besides exploring the peculiar properties of *Oo*-ASP-1, such as its dimerization, further structural investigation of the currently available ASP sequences led us to propose an expansion of the currently used ASP classification system, which is mainly based on the presence of cysteine and other conserved residues at particular positions in the primary sequence (Osman *et al.*, 2011).

Given the unlikely protease activity for *Oo*-ASP-1 and the vast overall diversity observed in CAP-superfamily members, it is clear that a thorough functional characterization of *Oo*-ASP-1, although remaining a high priority, will be a demanding task. Nevertheless, the currently available data on CAP proteins and ASPs, combined with the findings reported here, will allow us to guide future experiments in terms of (i) ligand identification and interaction profiling and (ii) the use of this information in engineering sequence-optimized CAP proteins, possibly in a multivalent format, yielding an elevated immune response of the host towards the parasite. Such efforts, which are currently in progress in our laboratory, may help to set the stage for a better understanding of CAP-superfamily proteins and their use as vaccines, which is a novel and highly necessary route in parasitic infection treatment given the current issues caused by anthelmintic resistance (Anziani *et al.*, 2004; Condi *et al.*, 2009; Demeler *et al.*, 2009; Waghorn *et al.*, 2006).

This work was partially funded through PARAVAC, an EU Seventh Framework Programme (project reference 265862 – wbs B/12252/01). SDW acknowledges the support of a Marie-Curie Reintegration Grant. Crystallographic X-ray data collection was made possible by the generous support of the staff of the PROXIMA 1 beamline at the SOLEIL Synchrotron, Saint-Aubin, France.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Anziani, O. S., Suarez, V., Guglielmone, A. A., Warnke, O., Grande, H. & Coles, G. C. (2004). *Vet. Parasitol.* **122**, 303–306.
Asojo, O. A. (2011). *Acta Cryst.* D**67**, 455–462.
Asojo, O. A., Goud, G., Dhar, K., Loukas, A., Zhan, B., Deumic, V., Liu, S., Borgstahl, G. E. & Hotez, P. J. (2005). *J. Mol. Biol.* **346**, 801–814.
Cantacessi, C., Campbell, B. E., Visser, A., Geldhof, P., Nolan, M. J., Nisbet, A. J., Matthews, J. B., Loukas, A., Hofmann, A., Otranto, D., Sternberg, P. W. & Gasser, R. B. (2009). *Biotechnol. Adv.* **27**, 376–388.
Condi, G. K., Soutello, R. G. & Amarante, A. F. (2009). *Vet. Parasitol.* **161**, 213–217.
Conlan, B. F., Gillon, A. D., Craik, D. J. & Anderson, M. A. (2010). *Biopolymers*, **94**, 573–583.
Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). *Genome Res.* **14**, 1188–1190.
Datu, B. J., Gasser, R. B., Nagaraj, S. H., Ong, E. K., O'Donoghue, P., McInnes, R., Ranganathan, S. & Loukas, A. (2008). *PLoS Negl. Trop. Dis.* **2**, e130.
Del Valle, A., Jones, B. F., Harrison, L. M., Chadderdon, R. C. & Cappello, M. (2003). *Mol. Biochem. Parasitol.* **129**, 167–177.

Demeler, J., Van Zeveren, A. M., Kleinschmidt, N., Vercruysse, J., Höglund, J., Koopmann, R., Cabaret, J., Claerebout, E., Areskog, M. & von Samson-Himmelstjerna, G. (2009). *Vet. Parasitol.* **160**, 109–115.
DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*, **473**, 540–543.
Eberle, H. B., Serrano, R. L., Füllekrug, J., Schlosser, A., Lehmann, W. D., Lottspeich, F., Kaloyanova, D., Wieland, F. T. & Helms, J. B. (2002). *J. Cell Sci.* **115**, 827–838.
Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fernández, C., Szyperski, T., Bruyère, T., Ramage, P., Mösinger, E. & Wüthrich, K. (1997). *J. Mol. Biol.* **266**, 576–593.
Flocco, M. M. & Mowbray, S. L. (1994). *J. Mol. Biol.* **235**, 709–717.
Geldhof, P., Vercauteren, I., Gevaert, K., Staes, A., Knox, D. P., Vandekerckhove, J., Vercruysse, J. & Claerebout, E. (2003). *Mol. Biochem. Parasitol.* **128**, 111–114.
Gibbs, G. M., Roelants, K. & O'Bryan, M. K. (2008). *Endocr. Rev.* **29**, 865–897.
Guo, M., Teng, M., Niu, L., Liu, Q., Huang, Q. & Hao, Q. (2005). *J. Biol. Chem.* **280**, 12405–12412.
Hawdon, J. M., Jones, B. F., Hoffman, D. R. & Hotez, P. J. (1996). *J. Biol. Chem.* **271**, 6672–6678.
Henriksen, A., King, T. P., Mirza, O., Monsalve, R. I., Meno, K., Ipsen, H., Larsen, J. N., Gajhede, M. & Spangfort, M. D. (2001). *Proteins*, **45**, 438–448.
Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R. & Callewaert, N. (2009). *Nature Protoc.* **4**, 58–70.
Jones, D. T. (1999). *J. Mol. Biol.* **292**, 195–202.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Ma, D., Xu, X., An, S., Liu, H., Yang, X., Andersen, J. F., Wang, Y., Tokumasu, F., Ribeiro, J. M., Francischetti, I. M. & Lai, R. (2011). *Thromb. Haemost.* **105**, 1032–1045.
Meyvis, Y., Callewaert, N., Gevaert, K., Timmerman, E., Van Durme, J., Schymkowitz, J., Rousseau, F., Vercruysse, J., Claerebout, E. & Geldhof, P. (2008). *Mol. Biochem. Parasitol.* **161**, 67–71.
Meyvis, Y., Geldhof, P., Gevaert, K., Timmerman, E., Vercruysse, J. & Claerebout, E. (2007). *Vet. Parasitol.* **149**, 239–245.
Milne, T. J., Abbenante, G., Tyndall, J. D., Halliday, J. & Lewis, R. J. (2003). *J. Biol. Chem.* **278**, 31105–31110.
Moser, J. M., Freitas, T., Arasu, P. & Gibson, G. (2005). *Mol. Biochem. Parasitol.* **143**, 39–48.
Moyle, M., Foster, D. L., McGrath, D. E., Brown, S. M., Laroche, Y., De Meutter, J., Stanssens, P., Bogowitz, C. A., Fried, V. A., Ely, J. A., Soule, H. R. & Vlasuk, G. P. (1994). *J. Biol. Chem.* **269**, 10008–10015.
Neves, M. A., Yeager, M. & Abagyan, R. (2012). *J. Phys. Chem. B*, **116**, 7006–7013.
Osman, A., Wang, C. K., Winter, A., Loukas, A., Tribolet, L., Gasser, R. B. & Hofmann, A. (2011). *Biotechnol. Adv.* **30**, 652–657.
Punta, M. *et al.* (2012). *Nucleic Acids Res.* **40**, 290–301.
Serrano, R. L., Kuhn, A., Hendricks, A., Helms, J. B., Sinning, I. & Groves, M. R. (2004). *J. Mol. Biol.* **339**, 173–183.
Shikamoto, Y., Suto, K., Yamazaki, Y., Morita, T. & Mizuno, H. (2005). *J. Mol. Biol.* **350**, 735–743.
Sigrist, C. J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A. & Hulo, N. (2010). *Nucleic Acids Res.* **38**, D161–D166.
Vanderschaeghe, D., Szekrényes, A., Wenz, C., Gassmann, M., Naik, N., Bynum, M., Yin, H., Delanghe, J., Guttman, A. & Callewaert, N. (2010). *Anal. Chem.* **82**, 7408–7415.
Waghorn, T. S., Leathwick, D. M., Rhodes, A. P., Jackson, R., Pomroy, W. E., West, D. M. & Moffat, J. R. (2006). *N. Z. Vet. J.* **54**, 278–282.
Wang, J., Shen, B., Guo, M., Lou, X., Duan, Y., Cheng, X. P., Teng, M., Niu, L., Liu, Q., Huang, Q. & Hao, Q. (2005). *Biochemistry*, **44**, 10145–10152.
Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
Xu, X., Francischetti, I. M., Lai, R., Ribeiro, J. M. & Andersen, J. F. (2012). *J. Biol. Chem.* **287**, 10967–10976.