Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Three-dimensional structure of a schistosome serpin revealing an unusual configuration of the helical subdomain

Joachim Granzin,[a] Ying Huang,[b] Celalettin Topbas,[b] Wenying Huang,[c]‡ Zhiping Wu,[b] Saurav Misra,[d] Stanley L. Hazen,[b] Ronald E. Blanton,[e] Xavier Lee[b] and Oliver H. Weiergräber[a]*

[a]Institute of Complex Systems, ICS-6: Structural Biochemistry, Forschungszentrum Jülich, 52425 Jülich, Germany, [b]Department of Cell Biology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA, [c]Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA, [d]Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA, and [e]Department of Infectious Diseases, Case Western Reserve University, Cleveland, OH 44190, USA

‡ Present address: GENEWIZ NJ, South Plainfield, NJ 07080, USA.

Correspondence e-mail: o.h.weiergraeber@fz-juelich.de

Received 30 January 2012
Accepted 24 February 2012

PDB Reference: ShSPI, 3sto.

Parasitic organisms are constantly challenged by the defence mechanisms of their respective hosts, which often depend on serine protease activities. Consequently, protease inhibitors such as those belonging to the serpin superfamily have emerged as protective elements that support the survival of the parasites. This report describes the crystal structure of ShSPI, a serpin from the trematode *Schistosoma haematobium*. The protein is exposed on the surface of invading cercaria as well as of adult worms, suggesting its involvement in the parasite–host interaction. While generally conforming to the well established serpin fold, the structure reveals several distinctive features, mostly concerning the helical subdomain of the protein. It is proposed that these peculiarities are related to the unique biological properties of a small serpin subfamily which is conserved among pathogenic schistosomes.

1. Introduction

Schistosomiasis is one of the most widespread and serious parasitic infections in humans and is caused by three species of the genus *Schistosoma* (Ross *et al.*, 2002; Steinmann *et al.*, 2006). The pathogen is a trematode, a complex organism with a seven-layered tegument, a gut and reproductive organs, as well as excretory and nervous systems. Paired male and female worms lodge inside the veins of the host, producing hundreds or thousands of eggs per day. Depending on the parasite species involved, the most apparent morbidity arising from schistosomiasis is either portal hypertension with life-threatening oesophogeal haemorrhage or bloody urine with impairment of renal function. Most infected individuals, however, suffer some degree of disease (King & Dangerfield-Cha, 2008). Adult schistosomes survive for years in the mesenteric veins draining the intestines or the vesicular veins draining the urinary bladder, where they do not stimulate clotting (Smith & Lichtenberg, 1974), are seemingly invisible to the complement system (McLaren & Incani, 1982; Novato-Silva *et al.*, 1980) and inhibit neutrophil attack (Caulfield *et al.*, 1980). How schistosomes manage to neutralize this wide range of activities is not known, but each of these defence-and-repair systems relies on the function of proteases. Therefore, the involvement of protease inhibitors is a logical conclusion.

The clotting cascade is usually activated through contact with foreign surfaces and those that are negatively charged provide the strongest stimulus. With its high content of free carboxylic groups (Stein & Lumsden, 1973), the schistosome tegument is negatively charged, yet living parasites are not observed to be surrounded by clot (Smith & Lichtenberg,

1974). Furthermore, schistosomes appear to actively inhibit clotting, since clots are only found around dead organisms examined *in situ* (Zussman & Bauman, 1971). Lysates or excretory–secretory products of schistosomes and nematodes inhibit clotting by specifically interfering with the activation or activity of factor XII, a serine protease also known as Hageman factor (Tsang & Damian, 1977; Foster *et al.*, 1992). The inhibitory activity of these preparations is neutralized by treatment with pronase and would thus appear to be mediated by a protein. However, the component responsible has never been isolated from schistosomes. Combined anticoagulant and complement-inhibitory activities have been demonstrated in several other helminths, suggesting a widespread phenomenon that is present even in parasites that inhabit the gut (Hammerberg *et al.*, 1980; Crawford *et al.*, 1982).

The infective stage of the parasite (cercariae) and the newly penetrated schistosomula are sensitive to killing by complement, but 48 h later they become resistant and remain so throughout the rest of their development. Sensitivity to complement can be restored in adult worms of *S. mansoni* by incubation with trypsin or pronase (Marikovsky *et al.*, 1990). This suggests that a surface protein mediates complement resistance. Protease inhibitors are unlikely to be the only defence mechanism of the parasite, nor are these nonspecific defences the host's only protection, but proteases and anti-proteases have not been as well studied as other components of the immune response.

The serpin (an acronym for serine protease inhibitor) superfamily constitutes the largest class of protease inhibitors, with more than 1500 members identified to date (Law *et al.*, 2006). They typically consist of a single large domain (of about 350–400 residues in length) with a highly conserved fold. In contrast to canonical protease inhibitors, serpins employ a unique conformational mechanism for inactivation of their target protease (Farady & Craik, 2010). A comprehensive phylogenetic analysis of several hundred serpin sequences resulted in a classification into 16 clades; remarkably, the invertebrate members are typically arranged according to speciation (with schistosome serpins forming clade m), while the vertebrate proteins group into diverse clusters (Irving *et al.*, 2000). Although most serpins are inhibitors of serine or cysteine proteases, several members have adopted different functions, acting, for example, as hormone transporters or molecular chaperones.

During screening of an *S. haematobium* expression library for species-specific antigens, we identified a putative serpin molecule; intriguingly, fractionation of worm lysates and immunolocalization suggested this protein to be membrane-associated (Blanton *et al.*, 1994). Given its unique location and the nature of the host immune response that it elicits (Li *et al.*, 1995), this serpin appears to participate in the parasite–host interaction.

The *S. haematobium* serpin (henceforth designated ShSPI, corresponding to UniProt entry Q26502) has been crystallized previously and data were collected to a resolution of 2.4 Å (Huang *et al.*, 1999). Here, we report the determination of its three-dimensional structure. Our data reveal a combination of structural features that have not been observed previously in the serpin superfamily. These properties are discussed in the context of the biological functions assigned to schistosome serpins.

Table 1
Data-collection and refinement statistics (PDB entry 3sto).

Values in parentheses are for the highest resolution shell.

Data set	Inflection point	Peak	Remote
Data collection			
Wavelength (Å)	0.9793	0.9788	0.9667
Space group	$P3_221$		
Unit-cell parameters (Å)	$a = b = 64.9$, $c = 187.0$		
Resolution range (Å)	21.58–2.40 (2.44–2.40)		
No. of reflections			
Measured	588775	542897	605659
Unique	18443	18641	18636
Completeness (%)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
Multiplicity	13.5 (14.0)	12.6 (13.0)	13.8 (14.2)
$R_{\rm merge}$ (%)	10.8 (41.8)	13.0 (66.5)	13.5 (90.9)
$\langle I/\sigma(I)\rangle$	36.6 (9.0)	29.2 (5.3)	25.8 (3.3)
Wilson B factor (Å^2)	28.7	36.9	41.5
Phasing statistics			
Mean figure of merit			
After *SHELXE*	0.573		
After *Parrot*	0.653		
Refinement statistics			
Resolution range (Å)	21.58–2.41		
$R_{\rm work}$ (%)	16.2		
$R_{\rm free}$ (%)	21.3		
No. of atoms	2948		
Protein residues	370		
Water molecules	79		
Average B factor (Å^2)			
Protein	33.3		
Solvent	33.0		
R.m.s. deviations			
Bond lengths (Å)	0.013		
Bond angles (°)	1.417		
Ramachandran plot (%)			
Favoured	98.6		
Additionally allowed	1.4		
Outliers	0		
ML coordinate error (Å)	0.26		

2. Materials and methods

2.1. Preparation and crystallization of ShSPI

The expression of histidine-tagged ShSPI in *Escherichia coli* and purification of the protein have been described previously (Huang *et al.*, 1999). ShSPI was crystallized by the vapour-diffusion method using a sitting-drop setup. Specifically, samples were prepared by mixing equal volumes of protein solution (9 mg ml^{-1}) and reservoir buffer [50 m*M* sodium acetate pH 5.0, 5%(*w*/*v*) PEG 6000, 1%(*v*/*v*) Triton X-100, 3%(*v*/*v*) dioxane]. Hexagonal crystals were obtained within 2–4 d. The integrity of the crystallized material was verified by SDS–PAGE and Western blotting (Huang *et al.*, 1999).

2.2. Data collection and structure determination

X-ray diffraction data were collected on the X4a beamline at the National Synchrotron Light Source (NSLS), Howard Hughes Medical Institute at Brookhaven National Laboratory

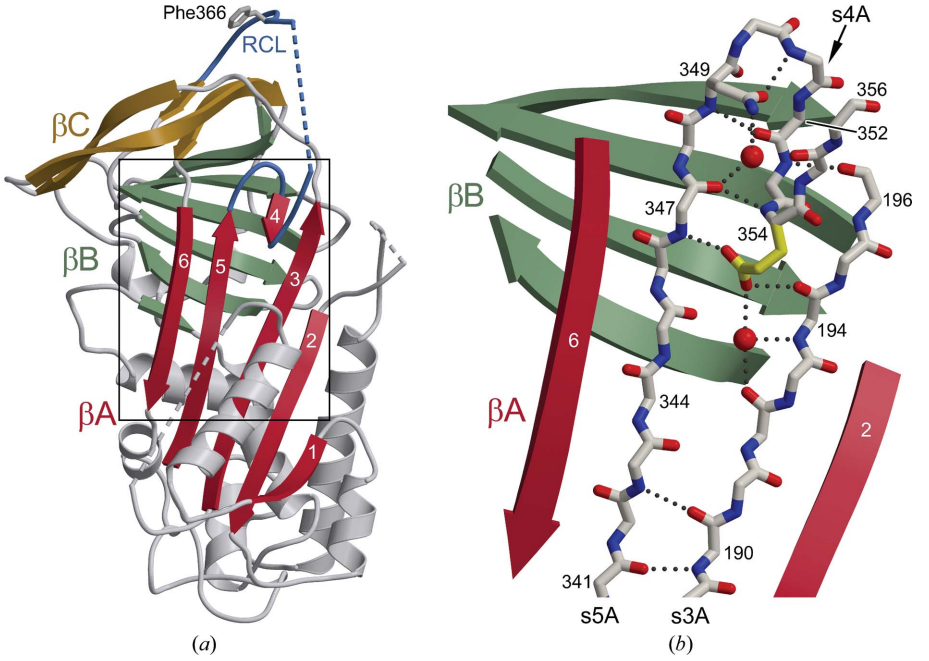

Figure 1
Crystal structure of ShSPI. (*a*) Ribbon model visualizing the typical serpin fold. The three β-sheets are distinguished by colour (βA, red; βB, green; βC, gold); in addition, strands are numbered for βA. Missing segments are represented by dashed connections. Note the RCL (blue) exposing a phenylalanine in the P1 position. (*b*) Close-up view of the area marked in (*a*), showing the hydrogen-bonding network in the breach region of βA. Strands 3A and 5A as well as the proximal hinge are drawn in stick mode; side chains are only included for Asn349 and Glu354 (yellow). Numbers in black indicate the sequence positions of selected residues.

using a Rigaku R-AXIS IV area detector. Specimens were flash-cooled in a nitrogen-gas stream at 100 K. A crystal of selenomethionyl protein diffracted X-rays to 2.4 Å resolution. It belonged to space group $P3_221$, with unit-cell parameters $a = b = 64.9$, $c = 187.0$ Å and one molecule in the asymmetric unit. Complete data sets were collected at peak, inflection-point and high-energy remote wavelengths of the Se absorption edge, exposing a different segment of the crystal at each wavelength. The diffraction data were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Local scaling, determination of Se-atom positions and initial phase calculations were performed with *REVISE*, which is part of the *CCP*4 suite (Winn *et al.*, 2011), and *SHELXD*/*E* (Sheldrick, 2008) and were followed by additional statistical phase improvement with *Parrot* (*CCP*4). Eight of the ten Se positions expected in the asymmetric unit could be determined and a partly interpretable electron-density map was obtained with an overall figure of merit of 0.653. An initial model containing 370 of 406 residues (91.1%) was built by combining *Buccaneer* (*CCP*4) and manual density interpretation using the inflection-point data set. Several cycles of model rebuilding and model refinement were performed using *O* (Jones *et al.*, 1991) and *PHENIX* (Adams *et al.*, 2002), assigning individual isotropic *B* factors to all atoms. 5% of the reflections were set aside as a random test set for cross-validation. During later stages of refinement, weights for both stereochemical and thermal parameter restraints were optimized with respect to $R_{\rm free}$. A total of 79 water molecules were picked from the difference Fourier map on the basis of peak heights and distance criteria. The final model had $R_{\rm work}$ and $R_{\rm free}$ values of 16.2% and 21.3%, respectively. Evaluation of model quality was performed with *RAMPAGE* (Lovell *et al.*, 2003): 98.6% of the residues were found in the favoured regions of the Ramachandran plot and the remaining 1.4% lay in additional allowed regions. Data-collection and refinement statistics are summarized in Table 1. The atomic coordinates and structure-factor amplitudes have been deposited in the PDB (http://www.pdb.org) with accession code 3sto.

2.3. Molecular graphics

Figures were generated with *MolScript* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997) using secondary-structure assignments provided by *DSSP* (Kabsch & Sander, 1983). Surface representations were prepared with *MSMS* (Sanner *et al.*, 1996). The structural alignment used in Fig. 3 was calculated with *LSQMAN* (Kleywegt & Jones, 1994) based on the C^{α} coordinates of β-strands 5A, 6A, 2B, 3B, 4B, 5B and 6B as well as helix B, corresponding to residues 339–349, 297–305, 242–247, 254–259, 381–385, 395–399, 38–42 and 44–52, respectively, of ShSPI. This superposition yielded a root-mean-square (r.m.s.) distance of 0.90 Å for the aligned stretches. Using all structurally equivalent segments (345 residues) as determined by *MUSTANG* (Konagurthu *et al.*, 2006), an overall r.m.s. distance of 3.05 Å was obtained.

2.4. Sequence alignment

A raw alignment of serpin sequences was generated by the *PRALINE* server (Simossis & Heringa, 2005) incorporating secondary-structure prediction by *PSIPRED* (Jones, 1999). This alignment was adjusted manually using the X-ray structures of human antithrombin III (ATIII) and ShSPI (PDB entries 1e05 and 3sto, respectively; McCoy *et al.*, 2003). During this process, gaps in the poorly conserved hC–hD loop were merged for the sake of clarity.

3. Results and discussion

3.1. Refinement and overall structure

The crystals of ShSPI belonged to space group $P3_221$ with one molecule per asymmetric unit. Using a crystal grown from selenomethionyl protein, three-wavelength diffraction data were collected at 100 K and processed to 2.4 Å resolution. The structure was determined by the MAD technique as described

in §2 and refined to an R factor of 16.5% using the inflection-point data set (Table 1). The protein adopts the classical serpin fold as exemplified by ATIII or α1-antitrypsin (representing serpin clades c and a, respectively). Indeed, evaluation of three-dimensional alignments using the *DALI* server (Holm & Rosenström, 2010) revealed human ATIII to be the closest relative of ShSPI in the PDB. Fig. 1(*a*) illustrates the major structural elements, including three β-sheets (βA, βB and βC) and nine α-helices (hA–hI; not labelled) assigned according to the conventional serpin nomenclature (Loebermann *et al.*, 1984). The reactive-centre loop (RCL), which essentially functions as a bait for the target protease, is largely exposed to the solvent. Several portions of the structure were not traceable owing to poor observed electron density; in addition to the extreme termini (residues 1–15 and 404–406, respectively), this also applies to loop segments 108–110 and 173–180. In accordance with previously determined serpin structures, the RCL also displays enhanced conformational freedom, with an overall temperature factor about 40% above the mean value of the core domain. Notably, the segment containing Val358–Val364 (P9–P3 according to the protease-substrate notation introduced by Schechter & Berger, 1967) could not be stably refined and was therefore excluded from the final model.

3.2. RCL and breach region

The structure of the RCL differs from the conformation found in most active inhibitory serpins such as α1-antitrypsin. Specifically, two residues of the N-terminal hinge, P15 (Gly352) and P14 (Ile353), are inserted into the upper part of βA (Fig. 1*a*), forming a very short strand 4A (s4A). A similar topology of this so-called breach region has been found in the structures of native ATIII (Carrell *et al.*, 1994) and antichymotrypsin (PDB entry 1yxa; A. J. Horvath, J. A. Irving, R. H. Law, J. Rossjohn, S. P. Bottomley, N. S. Quinsey, R. N. Pike, P. B. Coughlin & J. C. Whisstock, unpublished work). In contrast, the conformation of P13 (Glu354; the last residue in the conserved Glu-*X*-Gly-*X*-Glu serpin motif) is unique to ShSPI (Fig. 1*b*). The residue is located in a turn, with its amide N atom hydrogen-bonded to the carbonyl group of Lys347 in s5A, thus contributing one backbone hydrogen bond to βA. Most intriguingly, the side chain bends back into the plane of the β-sheet, with its carboxyl group interposed between strands 3A and 5A. It is fixed by a couple of strong hydrogen bonds (donor–acceptor distances 2.6–2.7 Å) involving backbone atoms of Lys347 (s5A) and Phe194 (s3A) as well as one nearby water molecule. It appears that, as far as hydrogen bonding of βA is concerned, the Glu354 side chain substitutes for the P13 carbonyl as well as the P12 amide N atom, thus mimicking a more extensive incorporation of s4A. In principle, the side-chain insertion described here should therefore promote further opening of the breach region at the top of βA. However, comparison of serpin structures with s4A insertions of different lengths reveals that this correlation is not very strict; in fact, excess space between the s3A and s5A main chains is often bridged by a variable number of water molecules. In ATIII, for instance, the s3A and s5A traces are very similar to those of ShSPI, despite its P13 residue being completely solvent-exposed.

It is important to note that the overall similarity to ATIII does not hold for the P1 residue defining the scissile bond: ShSPI has Phe366 at this position, in contrast to the classical AT-type serpins which expose a basic residue (typically Arg). The side chain of Phe366 partially points to the surface of the molecule; in the crystal lattice, however, this residue is largely shielded by a hydrophobic pocket formed by Tyr206 of the same molecule as well as Tyr65, Trp80 and the aliphatic portion of Arg62 of a symmetry equivalent. Additionally, the side-chain hydroxyl of Tyr206 forms a hydrogen bond to the backbone amide of Phe366. The well defined C-terminal part of the RCL is close to a crystallographic twofold axis and interacts with neighbouring molecules on either side. Owing to the considerable flexibility inherent to serpin RCLs, their conformations as observed in X-ray structures are expected to be susceptible to crystal-packing artifacts. Therefore, it is generally uncertain whether the various orientations and

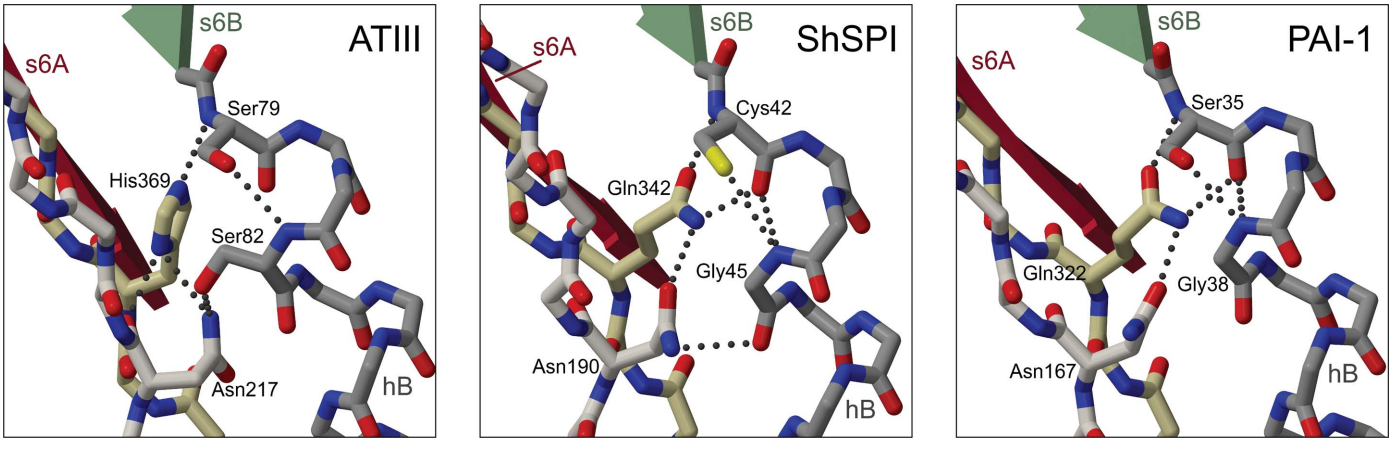

Figure 2
Comparison of the shutter regions of ShSPI (centre) and two typical inhibitory serpins representing clades c (ATIII; left) and e (PAI-1; right). Note the conservation of all but one hydrogen bond between ShSPI and PAI-1.

accessibilities found for the P1 side chains are representative of the structures in solution. On the other hand, interaction of the aromatic moieties of Phe366 and Tyr206 seems to be highly favourable in the present case since it provides the only obvious means of reducing the solvent exposure of these side chains.

3.3. Shutter region

In the serpin fold, the shutter region comprises the central portions of β-strands 3A and 5A as well as the underlying residues connecting strand 6B to helix B (Fig. 2). It was named owing to its role as a checkpoint on the path to full insertion of the RCL, since it controls the opening of the lower portion of βA (Hansen *et al.*, 2001). At the core of this mechanism is a hydrogen-bonding network which (in ATIII) chiefly involves His369 on s5A and Asn217 on s3A, together with Ser79 and Ser82. A sequence alignment of 219 serpin sequences (Irving *et al.*, 2000) revealed that these four positions are highly conserved, with the above-mentioned side chains present in 78, 85, 93 and 72% of cases, respectively. The functional significance of the shutter region is underscored by the discovery of naturally occurring mutations of these residues that lead to dysfunction and disease. Fig. 2 provides a close-up view of this region in our crystal structure of ShSPI (centre) as well as in ATIII (left) and plasminogen activator inhibitor 1 (PAI-1; right). Interestingly, of the four residues mentioned above only the asparagine (Asn190) is conserved in the schistosome protein, while the histidine is replaced by glutamine (Gln342) and the serine residues by cysteine (Cys42) and glycine (Gly45), respectively. Although this combination of exchanges (affecting three out of four positions) seems quite unusual, they are not unprecedented individually. The His-to-Gln substitution, for instance, is consistently found in clade e, which contains PAI-1, glia-derived nexin and a myxoma virus serpin, all three of which are well established inhibitory serpins. Another occurrence is found in SPN48 from the mealworm *Tenebrio molitor*, which also exhibits anti-proteinase activity (Jiang *et al.*, 2009). Finally, PAI-1 also contains the Ser-to-Gly exchange found in ShSPI, while in heparin cofactor II the two serines are replaced by alanine and glycine, respectively, without alterations at the remaining positions.

Despite the unusual composition of the shutter in ShSPI, the constituting side chains allow an extensive hydrogen-bonding network to be established, closely resembling that of clade e serpins (Fig. 2). We therefore assume that this region of ShSPI serves a similar gating function as in classical inhibitory members of the superfamily. Indeed, mutational studies with PAI-1 revealed that reversal of the shutter residues to the canonical residues resulted in significant acceleration of the latency transition, *i.e.* a destabilization of the active conformation. The delicate balance of conformational states, which is the result of global optimization for a specific task, implies that the tolerance of any individual structure towards exchanges in critical positions is more limited than the range of 'allowed' residues suggested by sequence comparisons.

For the incorporation of the RCL into βA to be favourable, it is believed that the energetic cost of breaking the above-mentioned hydrogen bonds (during the separation of s3A and s5A) needs to be offset by the interactions established by the newly formed s4A. Of particular importance in this respect is the P8 residue (Zhou *et al.*, 2003), which is a threonine in most inhibitory serpins (Thr359 in ShSPI).

3.4. Helical subdomain

The canonical serpin fold contains nine α-helices, most of which are clustered laterally close to one pole of the molecule (bottom right in Fig. 1*a*). In functional terms this region is very interesting since it undergoes conformational changes which are coupled to the state of RCL insertion into βA. During evolution of the superfamily, this connection has been exploited for highly specialized functions. First of all, the binding of ligands to the helical subdomain can allosterically modify the equilibrium between different RCL conformations. This has been particularly well established for ATIII, which is activated by a sulfated pentasaccharide found in heparin and related glycosaminoglycans on the vascular endothelium. Binding of this cofactor in the helix D region leads to the expulsion of the partially inserted RCL, turning the serpin into an efficient inhibitor of factor Xa (Jin *et al.*, 1997). Conversely, the state of the RCL affects the conformations of ligand-binding sites and thus their affinities. In ATIII, for instance, full loop insertion in the covalent protease complex reduces the affinity for heparin, thus releasing the protein into the circulation. Another example is the non-

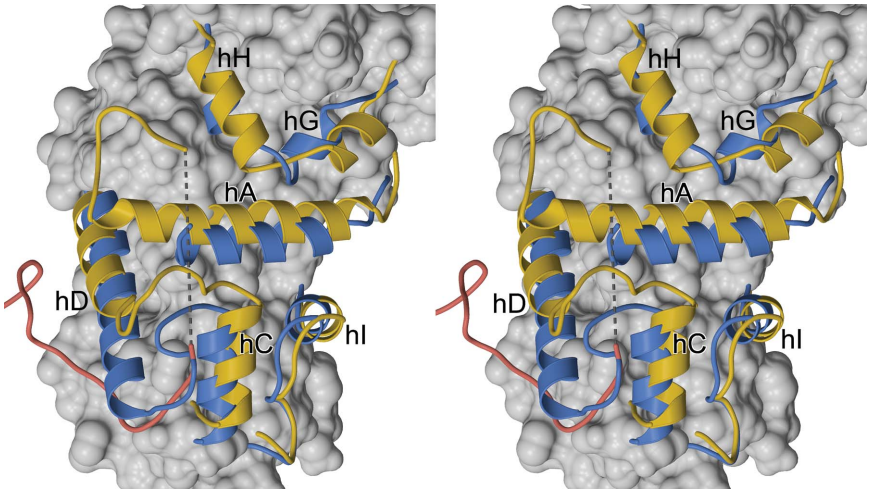

Figure 3
Unusual features of the ShSPI helical subdomain (stereoview). Selected helices and connecting loops are drawn in ribbon mode for ShSPI (blue) and ATIII (yellow), with the remainder of ShSPI represented as a molecular surface. The unique N-terminal extension of ATIII is shown in red. See text for details.

inhibitory serpin thyroxin-binding globulin (TBG), which serves to transport thyroid hormones in the blood. In this case, ligand binding occurs in a hydrophobic pocket formed by helices A and H, as well as strands 4B and 5B, and the protein is believed to cycle between high-affinity and low-affinity states with different extents of RCL insertion (Zhou *et al.*, 2006).

Comparing the helical subdomains of ShSPI and other members of the serpin superfamily reveals several unique properties of the schistosome protein, which are mostly centred on the hC–hD region (Fig. 3). The most intriguing observation is a large N-terminal elongation of helix D, which comprises 20 residues compared with 12 in ATIII. The connection between helices C and D is established by an S-shaped loop which contains a type I β-turn and is additionally stabilized by side-chain hydrogen bonds. Helix C itself is offset towards the lower pole of the molecule. This correlates with a significant displacement of helix A, which in turn affects the position of helix G and the conformation of the hG–hH loop. Overall, most of the structural differences described here appear to be secondary to the striking elongation of helix D. Comparison of a large number of serpin sequences has revealed huge variations in the length of the hC–hD loop (Irving *et al.*, 2000), with large insertions found primarily in intracellular members (clade b), such as myeloid and erythroid nuclear termination stage-specific protein and PAI-2.

To the best of our knowledge, the combination of features seen in ShSPI (the N-terminal elongation of helix D and the displacement of helices A, C and G) has not previously been observed in three-dimensional serpin structures; in fact, sequence comparison indicates that it is probably restricted to a distinct subfamily of schistosome serpins (discussed below).

It is interesting to note, however, that the ATIII molecule contains a unique N-terminal extension which overlaps the elongated helix D and the adjacent hC–hD loop in ShSPI (Fig. 3). Compared with conventional serpin structures, the additional polypeptide stretches present in ShSPI and ATIII are likely to restrict the mobility of helices C and D relative to the subjacent helix B. In this context, the side chain of Trp80 in the hC–hD loop of ShSPI is particularly interesting because it simultaneously interacts with Leu54 (helix B), Tyr65 (helix C), the aliphatic portions of Thr76 and Ser82 (hC–hD loop) and Ala85 (helix D). Remarkably, the antithrombin N-terminal segment is tightly anchored to the body of the molecule in the same region by a disulfide bridge as well as hydrophobic contacts. The functional implications of this analogy are unclear at present. In ATIII, residues in the N-terminal stretch have been shown to play an indirect role in heparin binding by modulating the ligand-induced conformational change (Schedin-Weiss *et al.*, 2004).

Does the helical subdomain of ShSPI display ligand-binding properties, similar to ATIII or TBG? Our crystal structure does not give any clear indication of such an activity. The hydrophobic ligand-binding pocket described for TBG attains a closed conformation in the ShSPI structure. The situation regarding helix D is more ambiguous. The surface properties in this region differ from those in antithrombin by the absence of basic clusters, making interaction with sulfated glycosaminoglycans unlikely. The same is true for alternative heparin-binding sites centred on helix H and strand 2C for protein C inhibitor (Huntington *et al.*, 2003) and on helix I for the beetle serpin SPN48 (Park *et al.*, 2011). Nevertheless, the striking extension of helix D and the preceding loop observed in ShSPI are likely to reflect a specific function of this schistosome serpin, such as association with as yet unknown protein binding partners.

3.5. Biological role of ShSPI

The serpin from *S. haematobium* investigated in this study is distinct from mammalian serpins in several respects. First of all, it is associated with the tegument of the schistosome and is at least partly anchored to membranes, whereas classical serpins are recognized to be soluble proteins. Secondly, its reactive-centre loop has features consistent with anti-elastase activity (see below), although its overall amino-acid sequence as well as its tertiary structure are more similar to those of coagulation inhibitors of the AT family. Understanding the biological role of ShSPI requires these issues to be addressed.

The crystallographic structure of ShSPI does not reveal exposed hydrophobic segments which could be inserted into a lipid bilayer, and sequence signatures coding for lipid modifications are also absent. These observations suggest that membrane association is mediated by interaction with another macromolecule. Indeed, we have shown previously that polyclonal antisera raised against ShSPI recognize additional proteins of higher apparent molecular mass in worm lysate, possibly corresponding to covalent complexes with other molecules (Blanton *et al.*, 1994). At the same time, studies in *S. mansoni* revealed a strikingly similar constellation for a serpin termed Smpi56 (Ghendler *et al.*, 1994). The latter was found to be covalently linked to a 28 kDa serine protease on the surface of the parasite. Notably, biochemical characterization indicated that the protease may be anchored to the membrane *via* a glycosylphosphatidylinositol moiety (Ghendler *et al.*, 1996).

Is ShSPI the *S. haematobium* counterpart of Smpi56? Sequencing of the *S. mansoni* genome has recently been completed (Berriman *et al.*, 2009) and a database search indicated the presence of eight different serpin genes in this organism. Unfortunately, the Smpi56 sequence has to our knowledge never been published, but the authors later noted that the protein displays greater than 70% sequence identity to ShSPI (Fishelson, 1995). From this statement, we conclude that Smpi56 is identical to UniProt entry C4QLC6 (GeneDB name Smp_090080).

The genome of *S. japonicum*, the third major species involved in human schistosomiasis, has also been sequenced (*Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009). This organism possesses three serpin genes and again one of these encodes a protein very similar (approximately 65% identical) to ShSPI. Akin to ShSPI, the *S. japonicum* protein (Sj serpin; UniProt entry

Q967L9; GeneDB name Sjp_0085750) was originally identified in a search for potential vaccination targets and immunofluorescence revealed that it also localizes to the tegument of the parasite (Yan *et al.*, 2005).

Taken together, these data strongly indicate that ShSPI, Smpi56 and Sj serpin are orthologues and thus are likely to share most, if not all, of their functional properties. In support of the latter assumption, we note that the amino-acid sequences of these three serpins are very similar to each other, but are divergent from other schistosome serpins, in solvent-exposed regions which are likely to be functionally significant, such as the RCL and the hC–hD extension (Fig. 4).

Several lines of evidence support the view that these proteins are inhibitors of elastase-like enzymes. As mentioned above, Smpi56 has been identified in a complex with a serine protease (designated m28) on the surface of adult worms (Ghendler *et al.*, 1994). This protease is believed to be closely related to a soluble form (s28) which is shed by the acetabular glands of invading cercariae and has been ascribed elastase activity. Specificity analysis using a number of mammalian serine proteases revealed that Smpi56 forms complexes with elastases of pancreatic as well as neutrophil origin (Ghendler *et al.*, 1994). It is interesting to note that the P1 residues in the RCL of ShSPI (phenylalanine) and Smpi56/Sj serpin (isoleucine) appear to be more appropriate for targeting chymotrypsin-like proteases, whereas most elastases prefer smaller uncharged residues at this position. Indeed, the elastinolytic cercarial proteases are inhibited by succinyl-Ala-Ala-Pro-Phe chloromethyl ketone, which is well known as a chymotrypsin inhibitor (Ghendler *et al.*, 1996). For the dominant cercarial elastases of *S. mansoni* (CE-1a and CE-1b), sequence preferences at substrate positions P4–P1 have been analyzed using combinatorial peptide libraries (Salter *et al.*, 2002). The RCL sequences of ShSPI and its orthologues are in reasonable agreement with the profile reported by these authors. Specifically, phenylalanine is the second-ranking residue at P1, while the proline residue in P2 is by far the most favoured side chain at this position. Even among vertebrate serpins, a P2 proline is typically found in so-called antitrypsins which in fact target elastases.

According to these considerations, the surface-tethered serpins might represent a protective measure against either

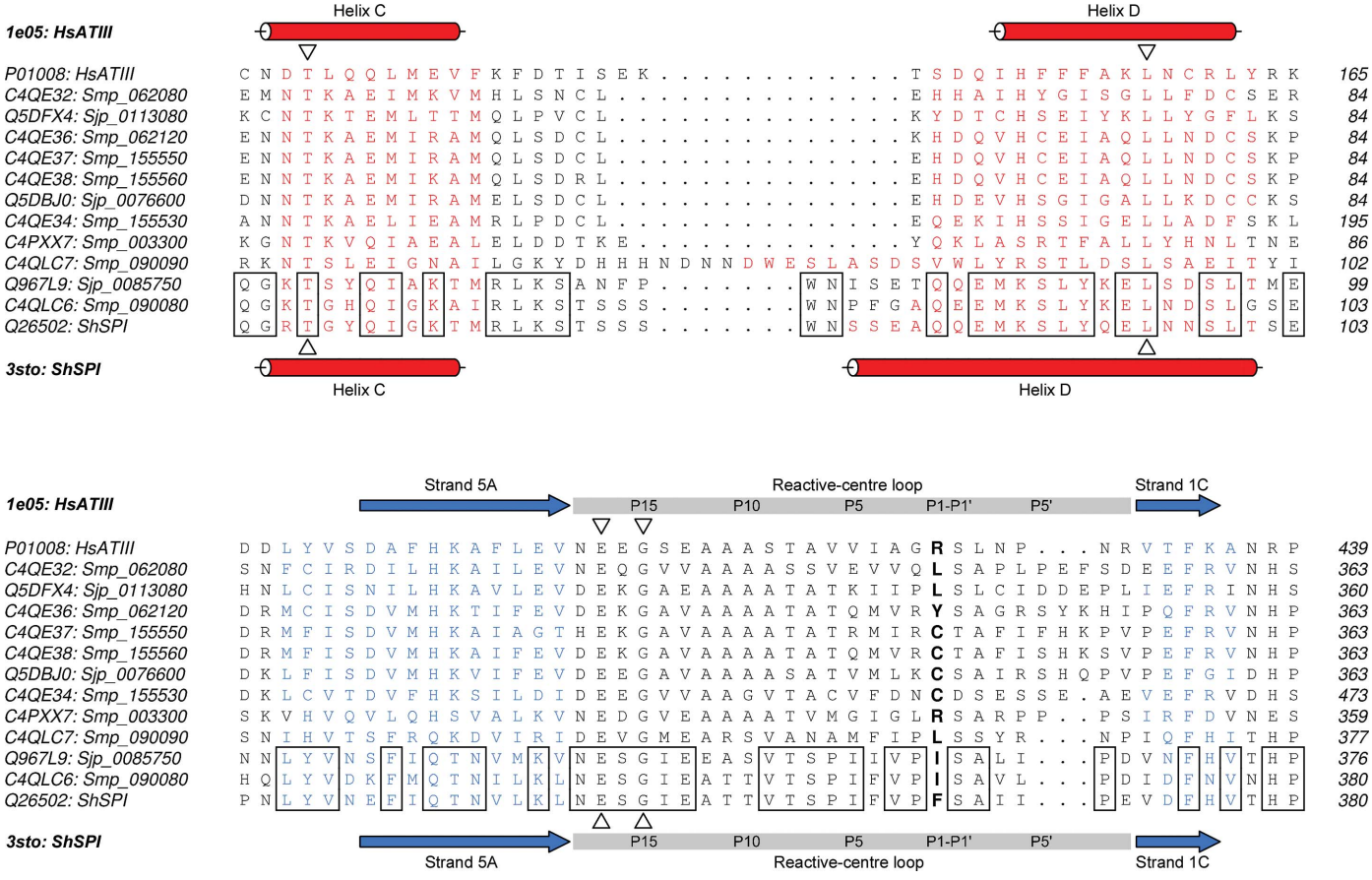

Figure 4
Structure-assisted sequence alignment comparing all known schistosome serpins with human ATIII (top panel, hC–hD region; bottom panel, RCL and adjacent β-strands). Sequences are identified by UniProt ID, together with the GeneDB name or (for ATIII and ShSPI) the customary abbreviation. Bold signifies the P1 residue. In all sequences, *PSIPRED* predictions for helix and strand are indicated by red and blue colours, respectively, whereas pictograms represent secondary-structure confirmed by X-ray crystallography in ATIII (PDB entry 1e05) and ShSPI (PDB entry 3sto; this study). Residues conserved between ShSPI and its proposed orthologues are boxed; the high degree of sequence identity suggests that the length of helix D found in ShSPI should apply to all three proteins. Residues in the RCL region (grey) are numbered according to the conventional protease-substrate notation. Triangles indicate invariant positions.

the parasite's own secreted elastase or, alternatively, elastase released by host phagocytes such as neutrophils. While the precise mechanisms of schistosome killing by these cells is not known, elastase has been shown to be toxic *in vitro* to different developmental stages of *S. mansoni* (Freudenstein-Dan *et al.*, 2003).

A recent investigation of blood feeding by *S. mansoni* suggested an unexpected role for serpins in the parasite gut. During a proteomic analysis of worm vomitus, the authors detected a schistosome serpin identified with gene Smp_090080 corresponding to Smpi56/C4QLC6. Within this environment, the protein was hypothesized to function as an inhibitor of coagulation of ingested blood (Hall *et al.*, 2011). Indeed, immunofluorescence studies of Sj serpin using sliced *S. japonicum* samples revealed distinct staining of intestinal epithelia in addition to the tegument (Yan *et al.*, 2005) and, retrospectively, our previous studies on ShSPI using intact parasites are compatible with a similar distribution in *S. haematobium* (Blanton *et al.*, 1994). In light of these results, we cannot exclude that in addition to being elastase inhibitors, ShSPI and its orthologues might target additional host proteases. This speculation is tempting since it could account for a large body of evidence supporting the presence of an anticoagulant and complement-inhibitory protein on the parasite surface. Further experiments will be required to clarify whether ShSPI is involved in these activities. While sequencing of the *S. haematobium* genome is not complete, the very different numbers of serpin genes in *S. mansoni* (eight) and *S. japonicum* (three) may indicate different extents of adaptation to the host organism (Snyder & Loker, 2000; Quezada & McKerrow, 2011). The presence of ShSPI/Smpi56/Sj serpin in all three species, together with their complex tissue distribution, supports the conclusion that this may represent a primordial schistosome serpin designed to fulfil a broad range of functions. Consequently, the diversification of serpin genes in *S. mansoni* would mirror the extremely sophisticated mechanisms of innate immunity and inflammation found in humans, which largely rely on proteolytic cascades.

4. Conclusion

In this paper, we have presented the first three-dimensional structure of a serpin from a human endoparasite. Since ShSPI and its orthologues are exposed on the schistosome tegument they have been recognized as obvious targets for immunization, but preliminary results have been inconsistent. We anticipate that the distinctive structural features of ShSPI, as observed in the current study, may be exploited for the design of more efficient vaccines. Despite such advances, several aspects of serpin biochemistry in schistosomes remain to be elucidated, including the identity of the target proteases, which cannot be predicted with accuracy from the RCL sequences alone, as well as their mode of surface attachment. Further investigation of these issues will help in understanding the strategies of disguise that are used so successfully by intravascular schistosomes to evade the defence machinery of their hosts.

JG and OHW are grateful to Georg Büldt and Dieter Willbold for continuous generous support. XL and SH are supported by NIH grant P01HL076491. Furthermore, excellent assistance by Craig Ogata, the manager of the X4a beamline, and his staff at NSLS is greatly appreciated.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Berriman, M. *et al.* (2009). *Nature (London)*, **460**, 352–358.
Blanton, R. E., Licate, L. S. & Aman, R. A. (1994). *Mol. Biochem. Parasitol.* **63**, 1–11.
Carrell, R. W., Stein, P. E., Fermi, G. & Wardell, M. R. (1994). *Structure*, **2**, 257–270.
Caulfield, J. P., Korman, G., Butterworth, A. E., Hogan, M. & David, J. R. (1980). *J. Cell Biol.* **86**, 46–63.
Crawford, G. P., Howse, D. J. & Grove, D. I. (1982). *J. Parasitol.* **68**, 1044–1047.
Farady, C. J. & Craik, C. S. (2010). *Chembiochem*, **11**, 2341–2346.
Fishelson, Z. (1995). *Mem. Inst. Oswaldo Cruz*, **90**, 289–292.
Foster, C. B., Flanigan, T. P., DeStigter, K. K., Blanton, R., Dumenco, L. L., Gallagher, C. & Ratnoff, O. D. (1992). *J. Lab. Clin. Med.* **120**, 735–739.
Freudenstein-Dan, A., Gold, D. & Fishelson, Z. (2003). *J. Parasitol.* **89**, 1129–1135.
Ghendler, Y., Arnon, R. & Fishelson, Z. (1994). *Exp. Parasitol.* **78**, 121–131.
Ghendler, Y., Parizade, M., Arnon, R., McKerrow, J. H. & Fishelson, Z. (1996). *Exp. Parasitol.* **83**, 73–82.
Hall, S. L., Braschi, S., Truscott, M., Mathieson, W., Cesari, I. M. & Wilson, R. A. (2011). *Mol. Biochem. Parasitol.* **179**, 18–29.
Hammerberg, B., Dangler, C. & Williams, J. F. (1980). *J. Parasitol.* **66**, 569–576.
Hansen, M., Busse, M. N. & Andreasen, P. A. (2001). *Eur. J. Biochem.* **268**, 6274–6283.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Huang, W., Haas, T. A., Biesterfeldt, J., Mankawsky, L., Blanton, R. E. & Lee, X. (1999). *Acta Cryst.* D**55**, 350–352.
Huntington, J. A., Kjellberg, M. & Stenflo, J. (2003). *Structure*, **11**, 205–215.
Irving, J. A., Pike, R. N., Lesk, A. M. & Whisstock, J. C. (2000). *Genome Res.* **10**, 1845–1864.
Jiang, R., Kim, E.-H., Gong, J.-H., Kwon, H.-M., Kim, C.-H., Ryu, K.-H., Park, J.-W., Kurokawa, K., Zhang, J., Gubb, D. & Lee, B.-L. (2009). *J. Biol. Chem.* **284**, 35652–35658.
Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N. & Carrell, R. W. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 14683–14688.
Jones, D. T. (1999). *J. Mol. Biol.* **292**, 195–202.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
King, C. H. & Dangerfield-Cha, M. (2008). *Chronic Illn.* **4**, 65–79.
Kleywegt, G. J. & Jones, T. A. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 9–14.
Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. (2006). *Proteins*, **64**, 559–574.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Law, R. H., Zhang, Q., McGowan, S., Buckle, A. M., Silverman, G. A., Wong, W., Rosado, C. J., Langendorf, C. G., Pike, R. N., Bird, P. I. & Whisstock, J. C. (2006). *Genome Biol.* **7**, 216.
Li, Z., King, C. L., Ogundipe, J. O., Licate, L. S. & Blanton, R. E. (1995). *J. Infect. Dis.* **171**, 416–422.
Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). *J. Mol. Biol.* **177**, 531–557.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Marikovsky, M., Parizade, M., Arnon, R. & Fishelson, Z. (1990). *Eur. J. Immunol.* **20**, 221–227.
McCoy, A. J., Pei, X. Y., Skinner, R., Abrahams, J.-P. & Carrell, R. W. (2003). *J. Mol. Biol.* **326**, 823–833.
McLaren, D. J. & Incani, R. N. (1982). *Exp. Parasitol.* **53**, 285–298.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Novato-Silva, E., Nogueira-Machado, J. A. & Gazzinelli, G. (1980). *Am. J. Trop. Med. Hyg.* **29**, 1263–1267.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, S.-H., Jiang, R., Piao, S., Zhang, B., Kim, E.-H., Kwon, H.-M., Jin, X.-L., Lee, B.-L. & Ha, N.-C. (2011). *J. Biol. Chem.* **286**, 1567–1575.
Quezada, L. A. & McKerrow, J. H. (2011). *An. Acad. Bras. Cienc.* **83**, 663–672.
Ross, A. G., Bartley, P. B., Sleigh, A. C., Olds, G. R., Li, Y., Williams, G. M. & McManus, D. P. (2002). *N. Engl. J. Med.* **346**, 1212–1220.
Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C. S. & McKerrow, J. H. (2002). *J. Biol. Chem.* **277**, 24618–24624.
Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). *Biopolymers*, **38**, 305–320.
Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
Schedin-Weiss, S., Desai, U. R., Bock, S. C., Olson, S. T. & Björk, I. (2004). *Biochemistry*, **43**, 675–683.
Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium (2009). *Nature (London)*, **460**, 345–351.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Simossis, V. A. & Heringa, J. (2005). *Nucleic Acids Res.* **33**, W289–W294.
Smith, J. H. & von Lichtenberg, F. (1974). *Am. J. Trop. Med. Hyg.* **23**, 71–77.
Snyder, S. D. & Loker, E. S. (2000). *J. Parasitol.* **86**, 283–288.
Stein, P. C. & Lumsden, R. D. (1973). *Exp. Parasitol.* **33**, 499–514.
Steinmann, P., Keiser, J., Bos, R., Tanner, M. & Utzinger, J. (2006). *Lancet Infect. Dis.* **6**, 411–425.
Tsang, V. C. W. & Damian, R. T. (1977). *Blood*, **49**, 619–633.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yan, Y., Liu, S., Song, G., Xu, Y. & Dissous, C. (2005). *Vet. Parasitol.* **131**, 53–60.
Zhou, A., Stein, P. E., Huntington, J. A. & Carrell, R. W. (2003). *J. Biol. Chem.* **278**, 15116–15122.
Zhou, A., Wei, Z., Read, R. J. & Carrell, R. W. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 13321–13326.
Zussman, R. A. & Bauman, P. M. (1971). *J. Parasitol.* **57**, 233–234.