Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification and crystallization of a novel membrane-anchored protein: the *Schistosoma haematobium* serpin

Wenying Huang,[a] Thomas A. Haas,[b] John Biesterfeldt,[a] Lori Mankawsky,[a] Ronald E. Blanton[c] and Xavier Lee[a]*

[a]Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA, [b]Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA, and [c]Division of Geographic Medicine, School of Medicine Case Western Reserve University, Cleveland, OH, USA

Correspondence e-mail: lee@xtal.ri.ccf.org

Received 3 July 1997
Accepted 22 June 1998

A unique serine-protease inhibitor (serpin) of the blood fluke *S. haematobium* has been crystallized. It is an antitrypsin with an unusual residue (phenylalanine) at its reactive center. Unlike any known member of this gene family, it is a membrane-anchored protein on the surface of the parasite. The location of this serpin and immunological response to the protein indicate that it may play a important role in host–parasite interaction. The crystals belong to the trigonal space group $P3_221$ or $P3_121$ with unit-cell parameters $a = b = 64.7$, $c = 186.7$ Å, $\alpha = 90.0$, $\beta = 90.0$, $\gamma = 120.0^\circ$. There is one molecule per asymmetric unit and the crystals diffracted to 2.2 Å.

1. Introduction

Serine-protease inhibitors (serpins) are an abundant family of proteins with diverse function: most notably, the regulation of clotting and clot stability, control of complement activity and regulation of tissue development. Underlining their medical importance, serpins also regulate neutrophil function (Bucurenci *et al.*, 1992; Kilpatrick *et al.*, 1991; Kitagawa & Takaku, 1982) and are pathogenicity factors for pox viruses (Ray *et al.*, 1992). Sequence homology is diffuse among members of this gene family, but the basic serpin structure is well conserved (Whisstock *et al.*, 1998) and their general mechanism of action is known.

Serpins function by providing a substrate for the target serine protease. The protease binds and cleaves the inhibitor, but dissociation of the enzyme–inhibitor complex is very slow, being of the order of days. The reactive center of the inhibitor (P1 amino acid) is the site where the proteases cleave, and is located on an exposed stressed loop. Upon cleavage of this loop, the serpin undergoes a dramatic conformational change that inserts the free strand as the fourth member of a six-stranded β-pleated sheet. As a consequence, the P1 amino acid of the serpin moves 70 Å from its original position (Löebermann *et al.*, 1984; Potempa *et al.*, 1994; Wright & Scarsdale, 1995). The association between enzyme and cleaved inhibitor is sufficiently stable to resist boiling and SDS–PAGE under reducing conditions. Only four serpins have been crystallized in their uncleaved form and their structures resolved: ovalbumin (Stein & Lumsden, 1973), human antichymotrypsin (Wei *et al.*, 1994), antithrombin (Carrell *et al.*, 1994) and α_1-antitrypsin (Elliott *et al.*, 1996). While the overall structures are similar, it appears that human antichymotrypsin and antithrombin must be fitted to the enzymatic pocket, while α_1-antitrypsin exists in a canonical structure for interaction prior to protease exposure. The diversity of conformational states in the serpin family may reflect the variation of interaction between serpins and their cognate proteases. Many questions remain, therefore, about structure–function relationships for the family as a whole.

Serpins were among the first proteins whose structure was solved by X-ray diffraction. A survey of available serpin structures reveals that the crystal structures of active serpins are only available for human proteins, except for one each from bovine and equine sources. This hardly reflects their representation in nature, since serpins have also been found in viruses, insects and parasites. Among parasites, serpins were first identified in schistosomes. Schistosomes are parasitic trematodes that are a major source of human morbidity (World Health Organization, 1986). Mature worm pairs live in veins draining the urinary bladder or intestines. They survive the host's immune attack in part by silencing the non-specific repair and defense systems that depend on serine-protease activation, *i.e.* clotting (Foster *et al.*, 1992; Tsang & Damian, 1977), complement (McLaren & Incani, 1982; Novato-Silva *et al.*, 1980) and neutrophil activation (Caufield *et al.*, 1980). Serpins have been identified from the three major species of this parasite family, *S. haematobium*, *S. mansoni* and *S. japonicum* (Blanton *et al.*, 1993; Ghendler *et al.*, 1994), and their sequences have been reported (Blanton *et al.*, 1993). Several aspects of the *S. haematobium* serpin make its structure of particular interest. The target protease of the recombinant *S. haematobium* serpin is trypsin-like, a protease that usually cleaves after Lys or Arg. However, the P1 residue of *S. haematobium* serpin is phenylalanine instead of a charged

residue. The same observation for a human antitrypsin serpin, kalistatin was reported by Chao and his associates (Chai *et al.*, 1993). The *S. haematobium* serpin is also the only known member of this gene family that is anchored in a membrane. Lastly, until now, all known serpin structures are of proteins from humans, other mammals or chickens. A serpin from a distant phylum may reveal unanticipated conformations or confirm the current models. This will be a valuable addition to the accumulating knowledge about functional mechanisms and evolution within this gene family.

2. Characterization of *S. haematobium* serpin and identification of its target protease

The DNA sequence, surface localization, characterization and expression of the *S. haematobium* (SH) serpin (Genbank accession number M99562) has been described previously (Blanton *et al.*, 1993). The size of the deduced protein is 45.9 kDa, but the recombinant protein was expressed with a 6 kDa N-terminal peptide derived from the expression vector. GeneWorks (Intelligenetics, Inc., 1991) alignment of the SH serpin protein sequence indicated 21% identity with human antithrombin and glial-derived nexin. Serpins from the related schistosome species *S. mansoni* (SM) and *S. japonicum* (SJ) have been subsequently cloned and sequenced (Genbank accession numbers L27100 and U11023, respectively).

Both immunofluorescent microscopy (Blanton *et al.*, 1993) and cell fractionation (Li *et al.*, 1995) indicated that SH serpin was almost entirely membrane associated. Treatment of an adult worm lysate with $CaCO_3$ and Triton X-114 indicate that the serpin partitions with the micellar fraction and is thus a membrane-associated protein (manuscript in preparation). The target specificity of the SH serpin has been identified by observing which of the three classes of serine proteases produces an inhibitory complex with the recombinant SH serpin. The SH serpin only forms an inhibitory complex when incubated with trypsin (Fig. 1). The likely target of a serpin of unknown function can often be predicted on the basis of the amino acid at its P1 site, *i.e.* the primary amino-acid sequence, since the structural elements of serpins can be aligned accurately from sequence data (Baumann *et al.*, 1991). The serpin P1 site is usually located 14–15 residues downstream of the strongly conserved hinge (Glu–*X*–Gly–*X*–Glu) motif and in most cases is followed by a serine residue (Carrell & Boswell, 1986). Trypsin-like proteases cleave after basic amino acids (Lys or Arg), elastase-like proteases cleave after amino acids bearing uncharged non-aromatic side chains (Ala, Val, Leu, Ile, Gly, Ser) and chymotrypsin-like proteases cleave after aromatic residues (Tyr, Phe, Trp). Thus, the SH serpins target specificity is unexpected.

3. Expression, purification and crystallization of SH serpin

SH serpin protein was expressed from the polyhistidine-tag plasmid expression vector pRSET (Invitrogen, San Diego, CA) and was purified by Ni^{2+}-affinity column chromatography. The His tag is at the N-terminus of the expressed protein and there is also a short sequence consisting of 35 amino-acid residues between the tag and the N-terminus of the SH serpin. The purified protein appeared to be 90–95% pure on Coomassie blue stained gels. After staining, there was one major band and a weaker band of slightly lower molecular weight (not shown). Mass spectroscopy showed that the major band corresponded to a molecular weight of 52.2 kDa (data not shown), which corresponded exactly to the calculated molecular weight of the serpin, based on the primary sequence including the histidine tag and two dithiothreitol (DTT) molecules. Based on our experience, DTT was required for the storage of serpin at 277 K. A Western blot of the serpin protein from a dissolved crystal indicated that only the major band was crystallized (Fig. 2). The lower molecular weight band represents the serpin cleaved at its C-terminus, since it reacts with an anti-serpin polyclonal antibody as well as monoclonal histidine-tag-specific antibody. Further purification by ion-exchange chromatography (SP Sepharose Fast Flow; Pharmacia, Uppsala, Sweden) removed this minor band. It was possible to grow crystals from the serpin protein obtained after purification by Ni-affinity chromatography alone which diffracted to 3.0 Å using a synchrotron radiation source. However, most of these crystals were twinned, judging from the split reflections produced. Further purifying the serpin protein by a second ion-exchange chromatographic step usually gave much better single crystals which diffracted to high resolution. This emphasises the importance of protein purity for successful crystallization. The concentration of recombinant protein used for crystallization was 9.0 mg ml^{-1}. Crystals were grown using the vapor-diffusion method in sitting drops at 295 K. The sitting drops contained 4.0 µl of protein solution and an equal volume of reservoir solution consisting of 50 m*M* acetate buffer at pH 5.0, 5% PEG 6000, 1% Triton X-100 and 3% dioxane. Hexagonal crystals appeared after approximately 2 d (Fig. 3). The largest crystals reached 0.15 × 0.15 × 0.10 mm in size and stopped growing after 4–5 d. Despite the small size of the crystals, they diffracted to 2.2 Å using the synchrotron radiation source.

4. Crystal characterization and data collection

Data collection was performed using the beamline X4a of the National Synchrotron Radiation Source (NSLS), Howard Hughes Medical Institute at Brookhaven National Laboratory. The size of the crystal was 0.15 × 0.15 × 0.10 mm and it was frozen at 100 K. The cryogenic solution was the same as the reservoir solution with the addition of 15% ethylene glycol. An R-AXIS IV area detector was used for data collection and *DENZO* and *SCALEPACK* were used for data processing (Otwinowski, 1993). The space group of the crystals was $P3_221$ or $P3_121$ with $a = b = 64.7$ and $c = 186.7$ Å. A data set to 2.5 Å was collected, and the R_{sym} of the data was 4.4% with an overall completeness of 77.2%.[1] The completeness of the highest shell (2.50–2.59 Å) is 61.2%. There was only one molecule per asymmetric unit. In fact, the crystal diffracted beyond 2.2 Å but, owing to the unfavorable orientation of the crystal in the mounting loop, serious overlap occurred at higher resolution reflections. Therefore, 2.5 Å resolution was the limit for this data collection.

5. Future direction

We have successfully purified and crystallized the recombinant SH serpin and have collected a data set to 2.5 Å for the resultant crystals. The crystals diffract well and are candidates for high-resolution crystallographic analysis. The molecular-replacement technique is being attempted to determine the SH serpin structure. As backup, crystals of trimethyl lead acetate derivative have been found. Multiple-wavelength anomalous diffraction will be used for the determination of the SH serpin if the molecular-replacement method fails to give

[1] $R_{sym} = \sum_h \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i I_{ih}$, where I_{ih} is the ith observation of reflection h and $\langle I_h \rangle$ is the average intensity obtained from the same reflection observed i times.

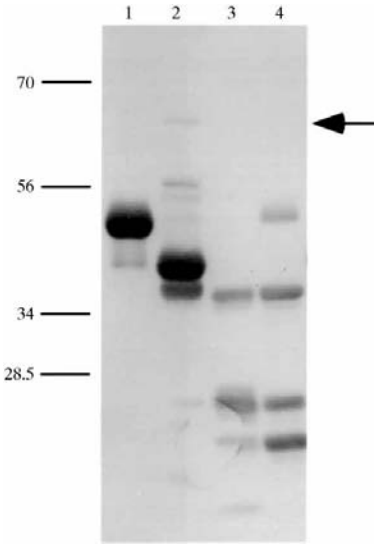

Figure 1
Western blot analysis of recombinant SH serpin–protease complex formation. rSH serpin (25 µg) (lane 1) was incubated at a 5:1 molar ratio (I:E) with bovine trypsin (lane 2), elastase (lane 3) and chymotrypsin (lane 4). After SDS–PAGE and transfer to a nitrocellulose membrane, the blot was incubated with rabbit anti-rSH serpin antibody and binding indicated by horse-radish peroxidase conjugated protein A. Only incubation with trypsin resulted in the appearance of a novel band of 68 kDa, corresponding to the predicted size of a rSH serpin–trypsin complex. The lower immunoreactive bands represent degradation of the recombinant serpin or serpin–protease complex. This result was reproduced in more than five consecutive experiments.

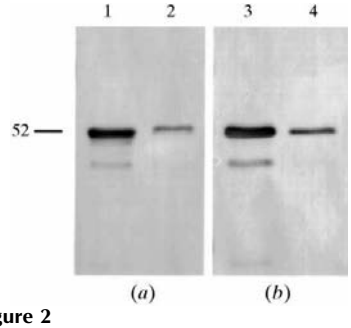

Figure 2
Western blot analysis of the purified SH serpin (lanes 1 and 3) and protein from a dissolved SH serpin crystal (lanes 2 and 4). (*a*) Western blot analysis using specific polyclonal rabbit antibodies. (*b*) Western blot analysis using antibody against the 6-histidine tag (Clontech, Palo Alto, CA).

a good solution. Three important questions can be addressed by examining the structure of the SH serpin. Firstly, does the structure of the SH serpin's stressed loop conform to either of the two current models for inhibitor–enzyme interaction? Since these

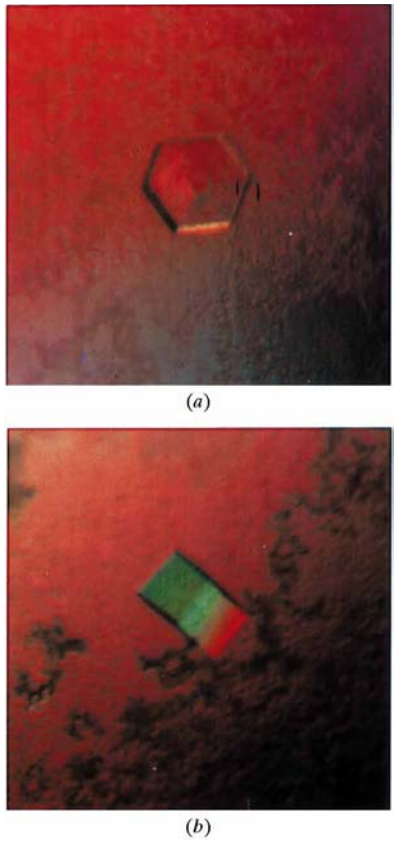

Figure 3
Photomicrographs of an SH serpin crystal. (*a*) Hexagonal form of the crystal (0.15 × 0.15 × 0.10 mm). (*b*) Thickness of the crystal used for data collection. Both are shown at the same magnification.

models are based on only three human serpins, the structure of a serpin from a distant phylum will either validate the current models or reveal an unanticipated structure in this region. Secondly, what are the structural determinants of specificity for this inhibitor given the fact that Phe is the P1 residue? Thirdly, how does membrane association modify the SH serpin structure relative to that of the other serpins? The determination of the structure of this serpin may help us to understand how the schistosome is able to persist in the bloodstream.

The authors wish to thank James Lang for his assistance with preparation of the figures for publication and Mr Michael Bushey for his computer assistance in our laboratory. Drs Bruce Carpick and Jiacheng Yang offered helpful discussions and assistance with protein purification.

References

Baumann, U., Huber, R., Bode, W., Grosse, D., Lesjak, M. & Laurell, C. B. (1991). *J. Mol. Biol.* **218**, 595–606.
Blanton, R. E., Licate, L. S. & Aman, R. A. (1993). *Mol. Biochem. Parasitol.* **63**(1), 1–11.
Bucurenci, N., Blake, D. R., Chidwick, K. & Winyard, P. G. (1992). *FEBS Lett.* **300**, 21–24.
Carrell, R. W. & Boswell, D. R. (1986). *Proteinase Inhibitors*, edited by A. Barrett and G. Salvesen, pp. 400–420. Amsterdam: Elsevier.
Carrell, R. W., Stein, P. E., Fermi, T. & Wardell, M. R. (1994). *Structure*, **2**, 257–270.
Caufield, J. P., Korman, G., Butterworth, A. E., Hogan, M. & David, J. R. (1980). *J. Cell Biol.* **86**, 46–63.
Chai, K., Chen, L., Chao, J. & Chao, L. (1993). *J. Biol. Chem.* **268**, 24498–24505.
Elliott, P. R., Lomas, D. A., Carrell, R. W. & Abrahams, J. P. (1996). *Nature Struct. Biol.* **3**, 676–681.
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.
Ghendler, Y., Arnon, R. & Fishelson, Z. (1994). *Exp. Parasitol.* **78**, 121–131.
Intelligenetics, Inc. (1991). *GeneWorks, Version 2.0.* Mountain View, CA, USA.
Kilpatrick, L., Johnson, J. L., Nickbarg, E. B., Wang, Z.-M., Clifford, T. F., Banach, M., Cooperman, B. S., Douglas, S. D. & Rubin, H. (1991). *J. Immunol.* **146**, 2388–2393.
Kitagawa, S. & Takaku, F. (1982). *Adv. Exp. Med. Biol.* **141**, 441–451.
Li, Z., King, C. L., Ogundipe, J., Licate, L. S. & Blanton, R. E. (1995). *J. Infect. Dis.* **171**, 416–422.
Löebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). *J. Mol. Biol.* **177**, 531–556.
McLaren, D. J. & Incani, R. N. (1982). *Exp. Parasitol.* **53**, 285–298.
Novato-Silva, E., Machado, J. A. N. & Gazinelli, G. (1980). *Am. J. Trop. Med. Hyg.* **29**, 1263–1267.
Otwinowski, Z. (1993). *Data Collection and Processing. Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Potempa, J., Korzus, E. & Travis, J. (1994). *J. Biol. Chem.* **269**, 15957–15960.
Ray, C. A., Black, R. A., Kronheim, S. R., Greenstreet, T. A., Sleath, P. R., Salvesen, G. S. & Pickup, D. J. (1992). *Cell*, **69**, 597–604.
Stein, P. C. & Lumsden, R. D. (1973). *Exp. Parasitol.* **33**, 499–514.
Tsang, V. C. W. & Damian, R. T. (1977). *Blood*, **49**, 619–633.
Wei, A., Rubin, H., Cooperman, B. S. & Christianson, D. W. (1994). *Nature Struct. Biol.* **1**, 251–258.
Whisstock, J., Skinner, R. & Lesk, A. M. (1998). *Trends Biochem. Sci.* **23**, 63–67.
World Health Organization (1986). *W. H. O. Statist. Quart.* **39**, 145–160.
Wright, H. T. & Scarsdale, J. N. (1995). *Proteins*, **22**, 210–225.