

# Structure of a two-CAP-domain protein from the human hookworm parasite *Necator americanus*

**Oluwatoyin A. Asojo**Pathology and Microbiology Department,  
986495 Nebraska Medical Center, Omaha,  
NE 68198-6495, USA

Correspondence e-mail: oasojo@unmc.edu

Major proteins secreted by the infective larval stage hookworms upon host entry include *Ancylostoma* secreted proteins (ASPs), which are characterized by one or two CAP (cysteine-rich secretory protein/antigen 5/pathogenesis related-1) domains. The CAP domain has been reported in diverse phylogenetically unrelated proteins, but has no confirmed function. The first structure of a two-CAP-domain protein, *Na*-ASP-1, from the major human hookworm parasite *Necator americanus* was refined to a resolution limit of 2.2 Å. The structure was solved by molecular replacement (MR) using *Na*-ASP-2, a one-CAP-domain ASP, as the search model. The correct MR solution could only be obtained by truncating the polyalanine model of *Na*-ASP-2 and removing several loops. The structure reveals two CAP domains linked by an extended loop. Overall, the carboxyl-terminal CAP domain is more similar to *Na*-ASP-2 than to the amino-terminal CAP domain. A large central cavity extends from the amino-terminal CAP domain to the carboxyl-terminal CAP domain, encompassing the putative CAP-binding cavity. The putative CAP-binding cavity is a characteristic cavity in the carboxyl-terminal CAP domain that contains a His and Glu pair. These residues are conserved in all single-CAP-domain proteins, but are absent in the amino-terminal CAP domain. The conserved His residues are oriented such that they appear to be capable of directly coordinating a zinc ion as observed for CAP proteins from reptile venoms. This first structure of a two-CAP-domain ASP can serve as a template for homology modeling of other two-CAP-domain proteins.

Received 19 January 2011

Accepted 6 March 2011

**PDB Reference:** *Na*-ASP-1,  
3nt8.

## 1. Introduction

Hookworm infection is a major global health problem that holds communities in a lifelong cycle of poor health, poverty and underdevelopment (Musgrove & Hotez, 2009). The hookworms *Necator americanus* and *Ancylostoma duodenale* infect over 700 million of the world's poorest people (de Silva *et al.*, 2003), causing a disease burden of about 22 million disability-adjusted life years (Hotez, 2007). Owing to the absence of effective long-term control methods for hookworm infection, alternative approaches including experimental anti-hookworm vaccines are being investigated (Hotez *et al.*, 2010). The initial selection of experimental vaccine candidates aimed to reproduce the success of attenuated larval stage 3 (L3) worms as canine vaccines with recombinant secreted L3 proteins (Bethony *et al.*, 2005).

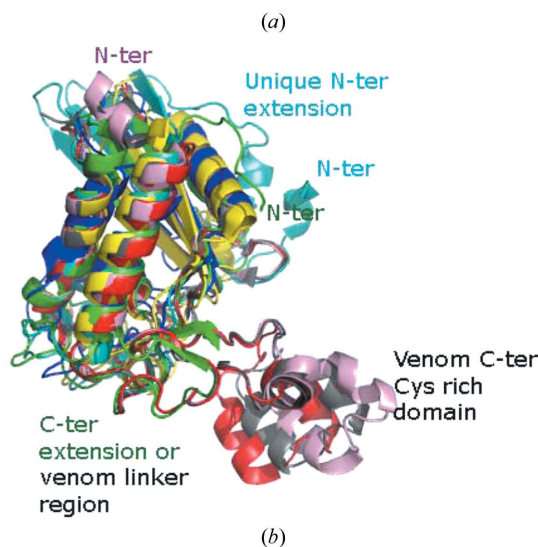
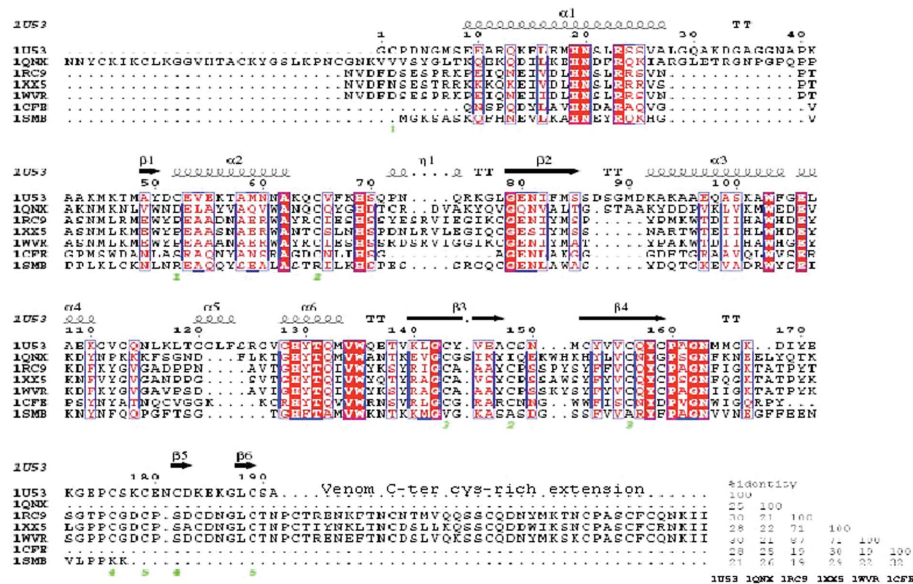
Two major proteins secreted by infective L3 hookworms upon host entry are *Ancylostoma* secreted protein 1 (ASP-1) and *Ancylostoma* secreted protein 2 (ASP-2), which have two and one CAP domains, respectively, and are exemplified by *Na*-ASP-1 and *Na*-ASP-2 from the most prevalent human

hookworm parasite *Necator americanus* (Goud *et al.*, 2005). Both *Na*-ASP-1 and *Na*-ASP-2 were initially investigated as experimental human hookworm vaccines to reproduce the effect of successful canine vaccines based on attenuated L3-stage worms (Bethony *et al.*, 2005). It was observed that recombinant full-length ASP-1 does not confer protective immunity to hookworm infection in permissive hosts such as hamsters (Goud *et al.*, 2004) and dogs (Hotez *et al.*, 2003). In contrast, ASP-2 shows similar protection as irradiated larvae in permissive mammalian hosts (Goud *et al.*, 2004), human vaccine studies (Bethony *et al.*, 2008) and endemic human populations (McSorley & Loukas, 2010). Additionally, truncating *Na*-ASP-1 by 96 amino-terminal residues increased its efficacy at reducing larval load compared with full-length *Na*-ASP-1 (Sen *et al.*, 2000; Ghosh *et al.*, 1996; Ghosh & Hotez, 1999).

The CAP [cysteine-rich secretory protein (CRISP)/antigen 5 (Ag-5)/pathogenesis related-1 (PR-1)] domain, alternatively referred to as sperm coating protein/Tpx-1/Ag-5/PR-1/Sc7 (SCP/TAPS; Tpx-1, testis specific protein) is characterized by the NCBI domain cd00168 or Pfam PF00188 and has been reported in diverse unrelated proteins from bacteria, plants, animals and viruses (Geer *et al.*, 2002; Gibbs *et al.*, 2008; Milne *et al.*, 2003; Cantacessi *et al.*, 2009; Ding *et al.*, 2000; Hawdon *et al.*, 1999; Zhan *et al.*, 2003; Gao *et al.*, 2001). This 15 kDa cysteine-rich domain has been implicated in conditions requiring cellular defense or proliferation such as plant responses to pathogens and human brain tumor growth (Ding *et al.*, 2000; Hawdon *et al.*, 1999; Zhan *et al.*, 2003; Gao *et al.*, 2001).

Several structures of proteins having a single CAP domain have been reported; however, only one of these is of an ASP, *Na*-ASP-2 (Asojo, Goud *et al.*, 2005). The three-dimensional structures of representative CAP proteins show significant structural homology in a conserved  $\alpha/\beta$ -barrel core (Fig. 1*a*). The sources of these representative CAP proteins are *Na*-ASP-2 from hookworm (PDB entry 1u53; Asojo, Goud *et al.*, 2005), Ves v 5 from yellow jacket (PDB entry 1qnx; Henriksen *et al.*, 2001), human Golgi-associated plant pathogenesis related protein 1 (GAPR-1; PDB entry 1smb; Serrano *et al.*, 2004),

tomato pathogenesis-related protein P14a (PDB entry 1cfe; Fernández *et al.*, 1997) and the snake-venom components natrin (PDB entry 1xx5; Wang *et al.*, 2005), triffin (PDB entry 1wvr; Shikamoto *et al.*, 2005) and sterisp (PDB entry 1rc9; Guo *et al.*, 2005). The main structural differences in the CAP domain include the lengths of the strands and helices as well as the length, orientation and location of loops (Fig. 1). Despite the significant structural homology of CAP proteins, they only



**Figure 1**

Structural elements of representative CAPs. (*a*) The conserved secondary-structure elements of the CAP core are based on *Na*-ASP-2 (PDB entry 1u53) and the CRISPs (PDB entries 1rc9, 1xx5 and 1wvr). This figure was generated with *ESPrpt* (Gouet *et al.*, 2003). Ves v 5 from yellow jacket (PDB entry 1qnx) has an amino-terminal extension, while both *Na*-ASP-2 and the CRISPs have carboxyl-terminal extensions. Tomato P14a (PDB entry 1cfe) and human Golgi-associated plant pathogenesis-related protein 1 (GAPR-1; PDB entry 1smb) do not have any extensions. The linker loops of the CRISPs share common sequence features with the carboxyl-terminal extension of 1u53. (*b*) Structural alignment of representative CAP proteins; the PDB codes are *Na*-ASP-2, 1u53 (green); GAPR-1, 1smb (yellow); PI14, 1cfe (blue); snake-venom CRISPs, 1xx5 (red), 1wvr (pink) and 1rc9 (gray). All core helices and strands overlay well from the mammalian (1smb), plant (1cfe), snake-venom CRISP (1xx5, 1wvr and 1rc9) and nematode (1u53) SCP domains. Snake-venom CRISPs have a unique C-terminal extension and 1qnx has a unique N-terminal extension. Despite the varied sequence homology, the three-dimensional structures of representative CAP proteins reveal significant structural homology in a conserved  $\alpha/\beta$ -barrel core.

share limited sequence identity to each other (Fig. 1*b*). According to the PROSITE database (<http://www.expasy.ch/prosite>), CAP proteins have two signature motifs, which are referred to as CRISP or CAP motifs (Fig. 1*b*). The consensus sequence for CAP1 is (GDER)(HR)(FYWH)(TVS)(QA)-(LIVM)(LIVMA)W<sub>xx</sub>(STN), while that for CAP2 is (LIVM-FYH)(LIVMFY)<sub>x</sub>C(NQRHS)Y<sub>x</sub>(PARH)<sub>x</sub>(GL)N(LIVNFYWDN). A PROSITE scan of *Na*-ASP-1 reveals only the carboxyl-terminal CAP1 motif consisting of residues 357-GHYTQMAWdtT-367. The carboxyl-terminal CAP2 motif 383-FgvCQYgPgGNY-394 is not recognized by PROSITE owing to the substitutions that are indicated as lower case letters. Two additional conserved regions (HN<sub>xx</sub>R) and [G(EQ)N(ILV)], referred to as the CAP3 and CAP4 motifs, respectively (Gibbs *et al.*, 2008), are also conserved in the carboxyl-terminal CAP domain. Only the CAP3 motif is present in the amino-terminal CAP domain of *Na*-ASP-1.

All one-CAP-domain proteins have a large central charged cavity containing key conserved charged residues from the CAP motifs; these correspond to His295, Glu306, Glu332, His358 and Ser296 in the carboxyl-terminal CAP domain. The CAP domain is a highly conserved cysteine-rich domain of approximately 15 kDa that has been implicated in conditions requiring cellular defense or proliferation such as plant responses to pathogens and human brain tumor growth (Ding *et al.*, 2000; Hawdon *et al.*, 1999; Zhan *et al.*, 2003; Gao *et al.*, 2001). There is no confirmed common function for CAP proteins; however, the substrate-specific endoprotease Tex31 from the cone snail *Conus textile* has been demonstrated to be a serine protease with known substrates (Milne *et al.*, 2003). The tertiary structure of Tex31 is undetermined but it has been speculated based on sequence similarity to Tex31 that CAP proteins may be substrate-specific Ser proteases (Gibbs *et al.*, 2008). Analysis of the reported structures of CAP proteins revealed that neither their monomers nor crystallographic dimers are capable of forming the catalytic triad of a conventional serine protease (Gibbs *et al.*, 2008). Since most one-domain CAPs assemble as dimers in solution it was important to identify if two-CAP-domain ASPs had interactions across the covalently linked CAP domains suggestive of CAP functions, such as forming a conventional Ser protease catalytic triad. To offer insights into the potential functions of ASPs and understand the structural basis of their vaccine potential, the first structure of a two-CAP-domain protein, *Na*-ASP-1, was determined.

## 2. Materials and methods

### 2.1. Crystallization and data collection

The crystallization of *Na*-ASP-1 has been described previously (Asojo, Loukas *et al.*, 2005). X-ray diffraction data were collected at the Eppley Institute crystallography core facility using an F-RE Superbright rotating-anode generator (Rigaku-MSD) operating at 45 kV and 45 mA with Osmic MicroMax optics and an R-AXIS IV<sup>++</sup> image-plate detector (Rigaku-MSD) and were processed and scaled as described in Asojo *et al.* (2010). Data statistics are listed in Table 1.

**Table 1**

Statistics for data collection and model refinement.

Values in parentheses are for the last shell.

Space group	<i>P</i> 2 <sub>1</sub>
Unit-cell parameters (Å, °)	<i>a</i> = 67.8, <i>b</i> = 74.6, <i>c</i> = 84.7, β = 112.1
Resolution limits (Å)	30.8–2.2 (2.3–2.2)
$\langle I/\sigma(I) \rangle$	15.4 (4.7)
No. of reflections	272397 (35542)
No. of unique reflections	37195 (5261)
Multiplicity	7.3 (6.8)
$R_{\text{merge}}^{\dagger}$ (%)	10.7 (55.7)
Completeness (%)	93.6 (91.4)
$R_{\text{cryst}}^{\ddagger}$	0.18 (0.19)
$R_{\text{free}}^{\S}$	0.24 (0.26)
Correlation coefficients	
$F_o - F_c$	0.946
$F_o - F_c$ (free)	0.911
R.m.s. deviations	
Bond lengths (Å)	0.020
Bond angles (°)	1.812
Mean <i>B</i> factors (Å <sup>2</sup> )	
Protein	16.1
All atoms	24.5
Model composition	
Monomers	2
Residues	798
Water molecules	211

<sup>†</sup>  $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  and  $\langle I(hkl) \rangle$  are the intensity of measurement *i* of *I* and the mean intensity of the reflection with indices *hkl*, respectively. <sup>‡</sup>  $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  are observed and  $F_{\text{calc}}$  are calculated structure-factor amplitudes. <sup>§</sup> The  $R_{\text{free}}$  set used 5% of randomly chosen reflections.

### 2.2. Structure determination and refinement

Attempts at molecular replacement (MR) using residues 12–180 of *Na*-ASP-2 yielded one monomer of *Na*-ASP-1 in the asymmetric unit with a reasonable solvent content (Asojo, Loukas *et al.*, 2005). However,  $R_{\text{free}}$  stayed in the high 40% and the electron-density maps were not consistent with the resolution limits of the data set. Thus, MR was repeated with the program *Phaser* (McCoy *et al.*, 2007) using a truncated polyalanine search model consisting of the core CAP region of *Na*-ASP-2 (residues 20–160) with all the long loops removed. This approach provided an initial MR solution with four identical polyalanine search models per asymmetric unit (Fig. 2*a*). When the loops were not removed from the search model the MR result was a monomer in the asymmetric unit. The variation in the length and the orientation of the removed loops proved vital in generating the correct MR phases for *Na*-ASP-1.

The correct MR solution implied a dimer of *Na*-ASP-1 per asymmetric unit with a Matthews coefficient (Kantardjiev & Rupp, 2003; Matthews, 1968) of 2.24 Å<sup>3</sup> Da<sup>-1</sup> and a solvent content of 45.1%. Owing to the high quality of the initial  $2F_o - F_c$ ,  $\varphi_c$  electron-density maps calculated from MR phases, the carboxyl-terminal and amino-terminal CAP domains could be visually distinguished. MR molecules *A* and *C* could be assigned as amino-terminal CAP domains, while molecules *B* and *D* were assigned as carboxyl-terminal CAP domains (Fig. 2*a*). In addition, 160 amino-acid residues of the *B* and *D* molecules as well as 150 amino-acid residues of the two amino-terminal domains were manually built through iterative

cycles of model building in the program *O* (Jones *et al.*, 1991), followed by structure refinement in *REFMAC5* (Murshudov *et al.*, 2011). Owing to the proximity of the four CAP domains there were two options to covalently link the domains. Option 1 has monomer 1 with CAP domain *A* linked to CAP domain *B* and monomer 2 with CAP domain *C* linked to CAP domain *D*, while option 2 has monomer 1 with CAP domain *A* linked to CAP domain *D* and monomer 2 with CAP domains *C* and *B* linked (Fig. 2*a*). The amino-acid residues were renumbered to generate models for each option. Both models were subsequently automatically built with the *CCP4* statistical protein chain-tracing program *Buccaneer* (Cowtan, 2006). The  $R_{\text{free}}$  for option 1 remained unreasonably high (47%), while that for

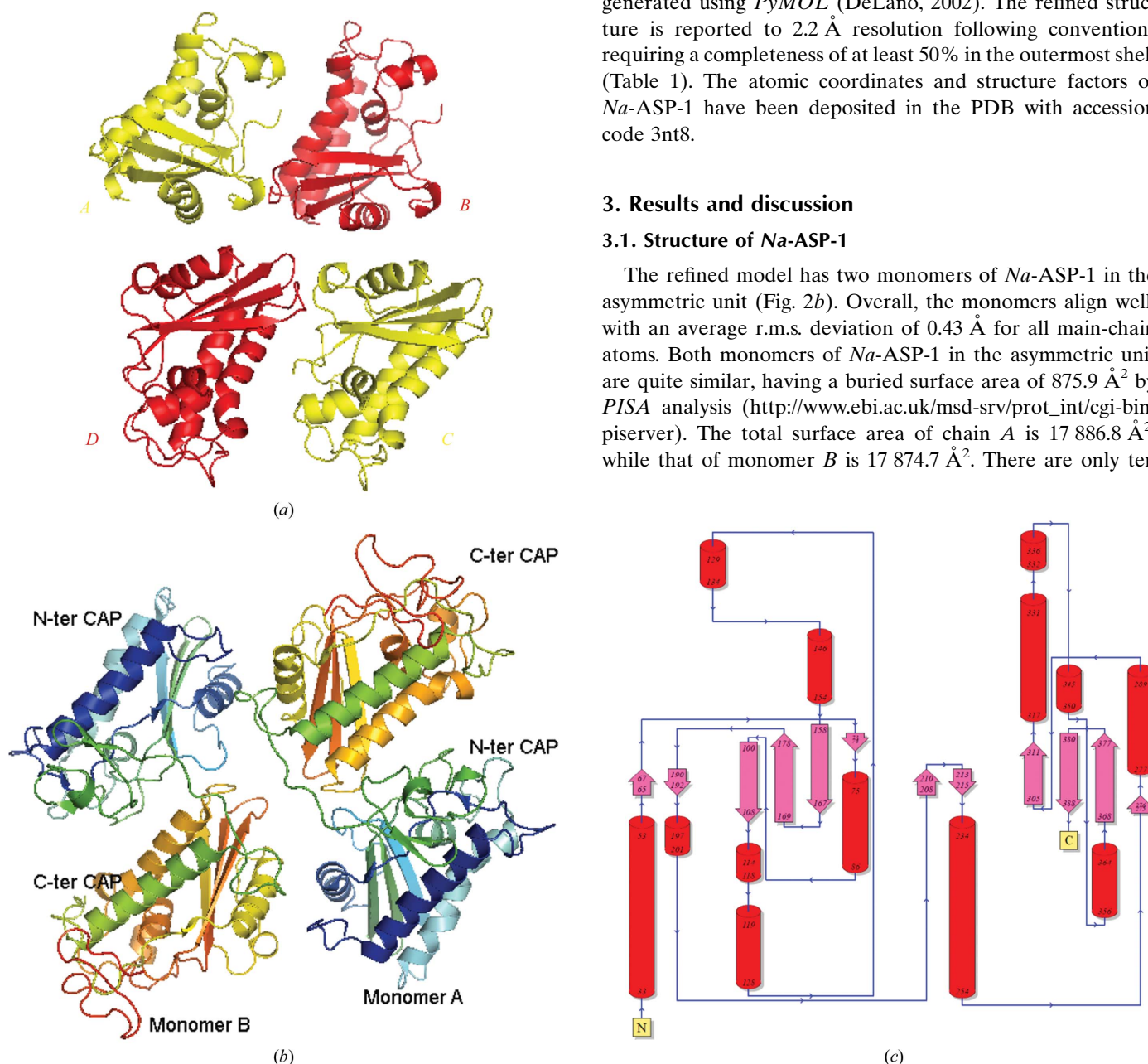
option 2 fell to 27% after ten cycles of automatic building with iterative refinement using *REFMAC5*. The main chains and side chains of 395 residues for both monomers as well as all the linker residues were also automatically built for option 2, which was not the case for option 1.

The model was improved through manual model-building cycles using *Coot* (Emsley & Cowtan, 2004) followed by *REFMAC5* refinement using a maximum-likelihood refinement procedure with Engh and Huber geometric parameters (Engh & Huber, 1991) and  $R_{\text{free}}$  (Brünger, 1992) to yield a crystallographic dimer of two *Na*-ASP-1 monomers (Fig. 2*b*). The refined structure was validated using *PROCHECK* (Laskowski *et al.*, 1993). Unless otherwise noted, figures were generated using *PyMOL* (DeLano, 2002). The refined structure is reported to 2.2 Å resolution following conventions requiring a completeness of at least 50% in the outermost shell (Table 1). The atomic coordinates and structure factors of *Na*-ASP-1 have been deposited in the PDB with accession code 3nt8.

### 3. Results and discussion

#### 3.1. Structure of *Na*-ASP-1

The refined model has two monomers of *Na*-ASP-1 in the asymmetric unit (Fig. 2*b*). Overall, the monomers align well, with an average r.m.s. deviation of 0.43 Å for all main-chain atoms. Both monomers of *Na*-ASP-1 in the asymmetric unit are quite similar, having a buried surface area of 875.9 Å<sup>2</sup> by *PISA* analysis ([http://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver)). The total surface area of chain *A* is 17 886.8 Å<sup>2</sup>, while that of monomer *B* is 17 874.7 Å<sup>2</sup>. There are only ten



**Figure 2** Structure of *Na*-ASP-1. (a) A ribbon diagram of the molecular-replacement solution shows the proximity of the four CAP domains and illuminates the difficulty in assigning the connectivity of the domains. (b) Ribbon diagram of crystallographic dimer of *Na*-ASP-1 rainbow colored from blue (N-terminus) to red (C-terminus). (c) Topology plot of *Na*-ASP-1 generated with *PDBSum* (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>).

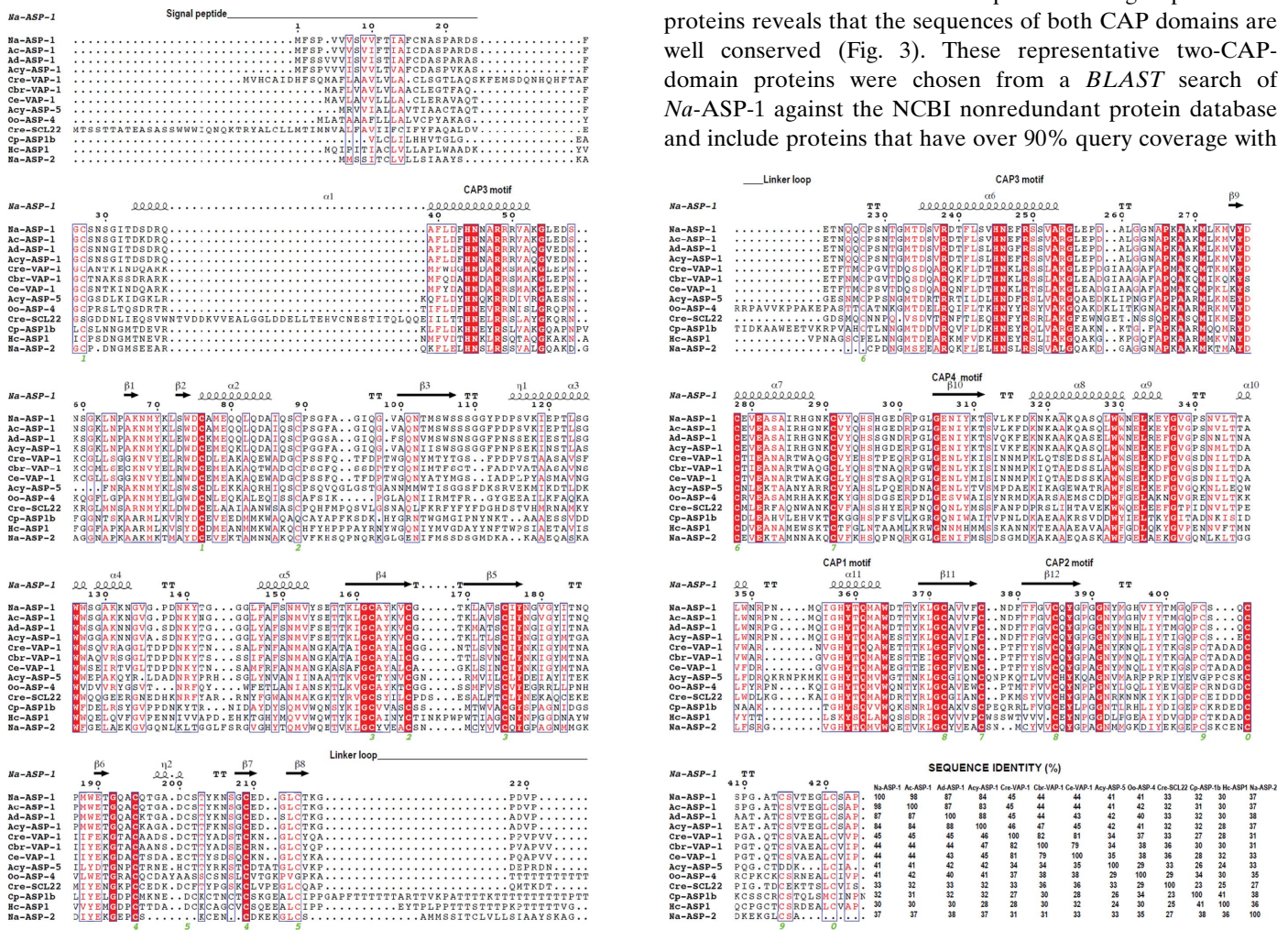
atoms that interact across the crystallographic dimer interface at a distance of 3.5 Å or less (Table 2).

Each monomer of *Na*-ASP-1 has two CAP domains connected by an extended 'linker loop' region. The amino-terminal CAP domain includes residues 24–215, the carboxyl-terminal CAP domain spans residues 234–424 and the linker loop includes residues 216–233 (Fig. 2c). The tertiary structure of the core of each CAP domain of *Na*-ASP-1 is a three-layer  $\alpha$ - $\beta$ - $\alpha$  sandwich in which a three-stranded antiparallel  $\beta$ -sheet lies between two layers of  $\alpha$ -helices. One  $\alpha$ -helical layer is composed of two parallel  $\alpha$ -helices, while the other has a solitary  $\alpha$ -helix. The r.m.s. deviation between the amino-terminal CAP domains is 0.24 Å and that for the carboxyl-terminal CAP domains is 0.39 Å. Overall, the

carboxyl-terminal CAP domain of *Na*-ASP-1 has greater tertiary structure similarity to *Na*-ASP-2 than the amino-terminal domain of *Na*-ASP-1. The r.m.s. deviation for the alignment of the carboxyl-terminal CAP domain with *Na*-ASP-2 is 0.485 Å, whereas it is 1.043 Å with the amino-terminal CAP domain. For the former, 152 of 190 (80%) amino-acid residues align, while only 137 of 191 (71%) amino-acid residues align in the latter.

### 3.2. *Na*-ASP-1 as a template for homology modeling of two-CAP-domain proteins

Two-CAP-domain proteins have only been identified in nematodes. Sequence alignment of *Na*-ASP-1 followed by secondary-structure assignment using the secondary-structure features of *Na*-ASP-1 with a representative group of these proteins reveals that the sequences of both CAP domains are well conserved (Fig. 3). These representative two-CAP-domain proteins were chosen from a *BLAST* search of *Na*-ASP-1 against the NCBI nonredundant protein database and include proteins that have over 90% query coverage with



**Figure 3** Comparison of *Na*-ASP-1 with selected representative two-CAP-domain proteins. Sequence alignment of selected two-CAP-domain proteins reveals the sequence conservation across both domains. The highest variability is in the signal peptide and loop regions. This figure was generated with *ESPrpt*. The different secondary-structure elements shown are  $\alpha$ -helices as large squiggles labelled  $\alpha$ ,  $3_{10}$ -helices as small squiggles labelled  $\eta$ ,  $\beta$ -strands as arrows labelled  $\beta$  and  $\beta$ -turns labelled TT. Identical residues are shown on a red background, conserved residues are shown in red and conserved regions are shown in blue boxes. *Na*-ASP-2 (AAP41952.1) is included as a representative one-CAP-domain ASP. Aligned proteins are *Na*-ASP-1, *N. americanus* ASP-1 (AAD13340.1); *Ac*-ASP-1, *Anclostoma caninum* ASP-1 (AAD318391); *Ad*-ASP-1, *A. duodenale* ASP-1 (AAD13339.1); *Acy*-ASP-1, *A. ceylanicum* ASP-1 (AAN11402.1); *Cre*-ASP-1, *Caenorhabditis remanei* two-domain activation-associated secreted protein (XP\_003106746.1); *Cbr*-VAP-1, *C. briggsae* VAP-1 (CAP3467.2); *Ce*-VAP-1, *C. elegans* VAP-1 (NM\_001029382.1); *Cre*-SCL22, *C. remanei* two-domain activation-associated secreted protein (XP\_003109641.1); *Cp*-ASP1b, *Cooperia punctata* two-domain activation-associated secreted protein-like (AAK35199.1); *Acy*-ASP-5, *A. ceylanicum* ASP-5 (ABB53347.1); *Oo*-VAP-4, *Ostertagia ostertagi* VAP-4 (CA000417); *Hc*-ASP-1, *Haemonchus contortus* putative secretory protein precursor (AAC03562). GenBank accession numbers are given in parentheses. The sequence identity of each protein is also listed.

**Table 2**

Contacts at the *Na*-ASP-1 crystallographic dimer interface (at 3.5 Å cutoff).

Source atom	Target atom	Distance (Å)
Arg286B NH1 (N)	Ser28A CB (C)	3.30
Arg286B CD (C)	Ser28A OG (O)	3.30
Arg286B NE (N)	Ser28A OG (O)	3.22
Arg286B CZ (C)	Ser28A OG (O)	3.15
Arg286B NH1 (N)	Ser28A OG (O)	3.05
Ala77B CA (C)	Ser229A O (O)	3.36
Ser28B CB (C)	Glu279A OE2 (O)	3.41
Ser28B OG (O)	Glu279A OE2 (O)	3.47
Asp84B OD2 (O)	Arg286A CG (C)	3.37
Asp84B OD1 (O)	Arg286A NH1 (N)	3.40

*Na*-ASP-1. The sequence identity among these diverse two-CAP-domain proteins varies from 23 to 98% (Fig. 3). For the selected sequences, the Cys residues that form the ten disulfide bonds either align or are off by one or two amino-acid positions, suggesting that the disulfide bridges are conserved. The carboxyl-terminal CAP domain is more conserved than the amino-terminal CAP domain. In the carboxyl-terminal CAP domain amino-acid residues that lie in helices or strands are well conserved and the regions of highest sequence variation are in the loops. In the amino-terminal CAP domain there appears to be some additional variations across species, with variations in the number of amino acids that make up the helices and  $\beta$ -strands. An additional structural variation for two-CAP-domain proteins is observed in the linker region, with canine and human hookworm ASPs having shorter linker loops than the others. The typical CAP motifs exemplified by *Na*-ASP-2 are only conserved in the carboxyl-terminal CAP domain. Instead, the amino-terminal CAP domain has its own conserved residues (Fig. 3). Owing to the sequence conservation for two-CAP-domain ASPs, the structure of *Na*-ASP-1 provides the first template for homology modeling of two-CAP-domain proteins.

### 3.3. The CAP cavity

As observed in all previously reported CAP structures, each CAP domain of *Na*-ASP-1 has a large central cavity. Both central cavities of the covalently linked CAP domains of *Na*-ASP-1 extend into each other to form a single large cavity (Fig. 4). The CAP cavity is an exposed central cavity that is observed in all CAP protein structures. The conserved amino-acid residues found in the CAP cavity of single-CAP-domain structures are only present in the carboxyl-terminal CAP domain and they all align well with those in *Na*-ASP-2 (Fig. 5a). While the putative cavity in *Na*-ASP-1 extends from the amino-terminal CAP domain to the carboxyl-terminal CAP domain, there is no proximal Ser that could form a Ser protease catalytic triad Ser/His/Glu as previously hypothesized (Milne *et al.*, 2003). Essentially, the network of interactions formed by the putative catalytic site residues in the two-domain ASP (*Na*-ASP-1) is virtually identical to that of the one-domain ASP (*Na*-ASP-2). Thus, as was the case with *Na*-ASP-2, it is unlikely that *Na*-ASP-1 is a typical serine protease.

**Table 3**

Contact of the carboxyl-terminal CAP domain with the linker loop and the amino-terminal CAP domain (at 3.5 Å cutoff).

Source atom	Target atom	Distance (Å)
Met233A C (C)	Thr234A N (N)	1.33
Met233A O (O)	Thr234A N (N)	2.26
Met233A CA (C)	Thr234A N (N)	2.40
Met233A CB (C)	Thr234A N (N)	3.15
Thr223A O (O)	Thr234A CA (C)	3.32
Met233A C (C)	Thr234A CA (C)	2.45
Met233A O (O)	Thr234A CA (C)	2.84
Thr223A O (O)	Thr234A CB (C)	3.40
Met233A C (C)	Thr234A CB (C)	3.48
Glu222A OE1 (O)	Thr234A CB (C)	3.43
Glu222A OE1 (O)	Thr234A OG1 (O)	3.01
Thr223A O (O)	Asp235A N (N)	2.81
Gln225A N (N)	Asp235A CG (C)	3.48
Thr223A O (O)	Asp235A CG (C)	3.40
Gln226A CB (C)	Asp235A OD1 (O)	3.27
Gln226A CG (C)	Asp235A OD1 (O)	3.15
Gln225A N (N)	Asp235A OD1 (O)	3.35
Gln226A N (N)	Asp235A OD1 (O)	3.08
Gln225A CG (C)	Asp235A OD2 (O)	3.04
Asn224A N (N)	Asp235A OD2 (O)	3.37
Gln225A N (N)	Asp235A OD2 (O)	2.82
Thr223A O (O)	Asp235A OD2 (O)	3.32
Thr223A OG1 (O)	Asp235AOD2 (O)	2.90
Thr223A C (C)	Asp235A OD2 (O)	3.31
Glu222A OE1 (O)	Ser236A CB (C)	3.42
Glu222A CD (C)	Ser236A OG (O)	3.49
Glu222A OE1 (O)	Ser236A OG (O)	2.80
Glu222A OE2 (O)	Ser236A OG (O)	3.42
Asn224A OD1 (C)	Arg238A CZ (C)	3.14
Asn224A OD1 (C)	Arg238A NH1 (N)	2.67
Cys227A CB (C)	Arg238A NH2 (N)	3.27
Asn224A OD1 (C)	Arg238A NH2 (N)	2.80
Cys227A O (C)	Arg238A NH2 (N)	3.47
Gln226A NE2 (N)	Asp239A OD1 (C)	3.12
Cys227A SG (S)	Cys238A CA (C)	3.39
Asn230A ND2 (N)	Cys278A O (C)	3.01
Cys227A SG (S)	Cys278A CB (C)	3.11
Gln226A O (O)	Cys278A SG (S)	3.40
Cys227A CA (C)	Cys278A SG (S)	3.44
Cys227A CB (C)	Cys278A SG (S)	3.03
Cys227A SG (S)	Cys278A SG (S)	2.03
Asn230A ND2 (C)	Glu281A OE1 (O)	2.93
Asn230A CB (C)	Glu281A OE1 (O)	3.43
Asn230A OD1 (O)	Ala282A CB (C)	3.40
Gly98A N (N)	Leu314A O (O)	2.55
Gly98A CA (C)	Leu314A O (O)	3.25
Gln97A C (C)	Leu314A O (O)	3.40
Gln97A CB (C)	Leu314A O (O)	3.45
Gly232A O (O)	Lys318A CE (C)	3.46
Gly232A O (O)	Lys318A NZ (N)	3.34
Tyr204A OH (O)	Lys320A CE (C)	3.40
Tyr204A OH (O)	Lys320A NZ (N)	3.45
Met233A CE (C)	Cys372A O (O)	3.35

The CAP cavity was revealed to be the major Zn<sup>2+</sup> binding site in the Zn<sup>2+</sup> and heparin-sulfate dependent mechanisms of inflammatory modulation in studies on the cobra CRISP natrin (Wang *et al.*, 2010). The structure of cobra CRISP natrin in complex with Zn<sup>2+</sup> (PDB entry 3mz8) revealed that the two conserved histidines directly coordinate the Zn<sup>2+</sup>. Likewise, the Zn<sup>2+</sup> complex of another CRISP, pseudocin (PDB entry 2epf), also shows similar Zn<sup>2+</sup> binding (Suzuki *et al.*, 2008). Owing to the highly conserved cavity, it is possible that a Zn<sup>2+</sup> ion will bind in a very similar fashion in *Na*-ASP-1 since the residues in the carboxyl-terminal CAP cavity superpose well with those of natrin (Fig. 5b). Since ASPs, like CRISPs, are

produced in conditions involving host immune responses, it is plausible that their conserved open central conserved cavities bind  $Zn^{2+}$  for similar processes, for example modulation of inflammation; however, more studies are required to verify if and how ASPs bind  $Zn^{2+}$ .

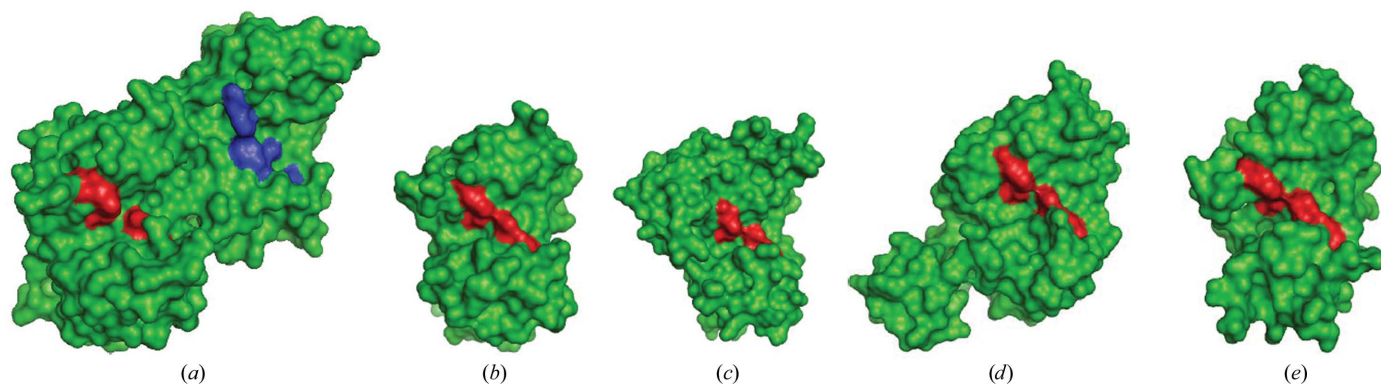
### 3.4. Structural insights into vaccine efficacy

A measure of an effective hookworm vaccine is its ability to reduce larval load in permissive hosts at comparable levels to irradiated L3 (Hotez *et al.*, 2010). By this measure *Na*-ASP-2 is an effective experimental vaccine, whereas *Na*-ASP-1 is not (Ghosh *et al.*, 1996; Hotez *et al.*, 1996, 2003; Goud *et al.*, 2004; Bethony *et al.*, 2008). The structure of *Na*-ASP-1 offers insights into why this is the case. The presence of the amino-terminal CAP domain as well as the linker loop blocks an entire face as well as portions of the carboxyl-terminal CAP domain which are exposed in *Na*-ASP-2 (Fig. 6). The blocking includes 53 contacts of less than 3.5 Å between the carboxyl-terminal CAP domain, the amino-terminal CAP domain and the linker loop (Table 3). The blockage of conserved portions of the carboxyl-terminal CAP domain may affect how *Na*-ASP-1 interacts with other molecules, including host anti-

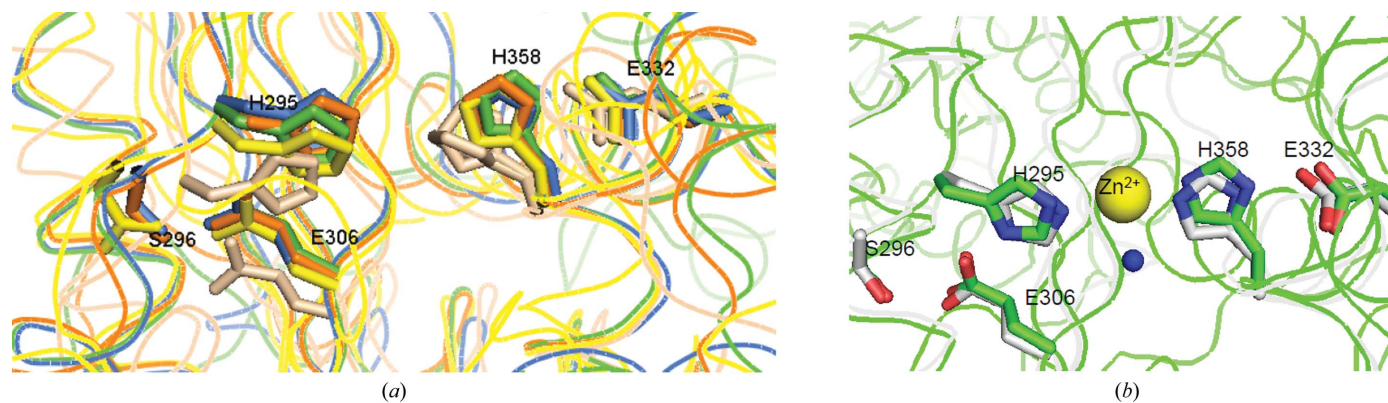
bodies, suggesting that the binding patterns of the two-domain ASPs are different from that of the one-domain ASPs. Sufficient regions of the carboxyl-terminal CAP domain are blocked to render *Na*-ASP-1 less effective at eliciting protective antibodies and reducing larval loads when compared with *Na*-ASP-2.

### 4. Conclusions

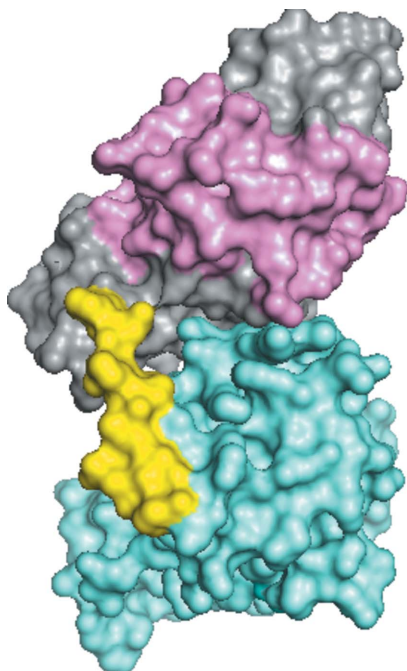
The structure-determination process for *Na*-ASP-1 reiterates the importance of using appropriate molecular-replacement search models and removing flexible regions that may yield incorrect molecular-replacement solutions. This structure of *Na*-ASP-1 is the first structure of a two-CAP-domain protein and reveals a central cavity that extends across both CAP domains. Unlike the amino-terminal CAP cavity, the carboxyl-terminal CAP cavity contains key residues that are virtually identical to those found in that of *Na*-ASP-2. The presence of the amino-terminal CAP domain and linker loop sufficiently blocks parts of the carboxyl-terminal CAP domain to render *Na*-ASP-1 an ineffective experimental vaccine, unlike *Na*-ASP-2.



**Figure 4** Surface plots of representative CAP structures in green reveal a conserved putative binding cavity. The plots of (a) *Na*-ASP-1, (b) *Na*-ASP-2 (1u53), (c) Ves v 5 (1qnx), (d) snake-venom CRISP (1xx5) and (e) GAPR-1 (1smb) are shown the same orientation and the location of the two conserved Glu and His residues are shown in red. The amino-terminal CAP cavity of *Na*-ASP-1 is colored blue. The extension of the cavity through the entire monomer of *Na*-ASP-1 is reminiscent of the CRISPs in which the cavity extends from the amino-terminal CAP to the carboxyl-terminal cysteine-rich domain.



**Figure 5** Superposed conserved central cavity of CAPs. (a) The superposed cavities reveal that key residues corresponding to His295, Glu306, Glu332, His358 and Ser296 superimpose well in representative CAP structures. CAP structures are colored as follows: *Na*-ASP-1, green; *Na*-ASP-2, orange; Ves v 5, yellow; GAPR-1, brown; snake-venom CRISP (1xx5), blue. (b) These same residues superpose in *Na*-ASP-1 (green), with the cobra CRISP natrin (3mz8) in complex with  $Zn^{2+}$  (gray).  $Zn^{2+}$  is shown in yellow and a water molecule that is coordinated to the  $Zn^{2+}$  is shown in blue.



**Figure 6**  
Domain association of covalently linked CAP domains of *Na*-ASP-1. The C-terminal CAP domain of *Na*-ASP-1 (shown in aquamarine) is blocked by the N-terminal CAP domain (gray and pink) and the linker loop (yellow). Also shown are the first 96 residues of the N-terminal CAP domain (pink) that when truncated expose more of the C-terminal CAP domain and result in better antibody binding.

This work was supported by the National Institutes of Health (R03AI065990). Thanks to Drs Peter Hotez and Raymond Koski for all their assistance, discussions and comments.

**References**

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.  
 Asojo, O. A., Loukas, A., Inan, M., Barent, R., Huang, J., Plantz, B., Swanson, A., Gouthro, M., Meagher, M. M. & Hotez, P. J. (2005). *Acta Cryst.* **F61**, 391–394.  
 Asojo, O. A., Loukas, A., Inan, M., Barent, R., Huang, J., Plantz, B., Swanson, A., Gouthro, M., Meagher, M. M. & Hotez, P. J. (2010). *Acta Cryst.* **F66**, 1549.  
 Bethony, J., Loukas, A., Smout, M., Brooker, S., Mendez, S., Plieskatt, J., Goud, G., Bottazzi, M. E., Zhan, B., Wang, Y., Williamson, A., Lustigman, S., Correa-Oliveira, R., Xiao, S. & Hotez, P. J. (2005). *FASEB J.* **19**, 1743–1745.  
 Bethony, J. M., Simon, G., Diemert, D. J., Parenti, D., Desrosiers, A., Schuck, S., Fujiwara, R., Santiago, H. & Hotez, P. J. (2008). *Vaccine*, **26**, 2408–2417.  
 Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.  
 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.  
 Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.  
 Delano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.  
 Ding, X., Shields, J., Allen, R. & Hussey, R. S. (2000). *Int. J. Parasitol.* **30**, 77–81.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.  
 Eng, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.  
 Fernández, C., Szyperski, T., Bruyère, T., Ramage, P., Mössinger, E. & Wüthrich, K. (1997). *J. Mol. Biol.* **266**, 576–593.  
 Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J. & Hussey, R. S. (2001). *Int. J. Parasitol.* **31**, 1617–1625.  
 Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. (2002). *Genome Res.* **12**, 1619–1623.  
 Ghosh, K., Hawdon, J. & Hotez, P. (1996). *J. Infect. Dis.* **174**, 1380–1383.  
 Ghosh, K. & Hotez, P. J. (1999). *J. Infect. Dis.* **180**, 1674–1681.  
 Gibbs, G. M., Roelants, K. & O’Byrne, M. K. (2008). *Endocr. Rev.* **29**, 865–897.  
 Goud, G. N. *et al.* (2004). *J. Infect. Dis.* **189**, 919–929.  
 Goud, G. N. *et al.* (2005). *Vaccine*, **23**, 4754–4764.  
 Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.  
 Guo, M., Teng, M., Niu, L., Liu, Q., Huang, Q. & Hao, Q. (2005). *J. Biol. Chem.* **280**, 12405–12412.  
 Hawdon, J. M., Narasimhan, S. & Hotez, P. J. (1999). *Mol. Biochem. Parasitol.* **99**, 149–165.  
 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.  
 Hotez, P. J. (2007). *PLoS Negl. Trop. Dis.* **1**, e149.  
 Hotez, P. J. *et al.* (2003). *J. Parasitol.* **89**, 853–855.  
 Hotez, P. J., Bethony, J. M., Diemert, D. J., Pearson, M. & Loukas, A. (2010). *Nature Rev. Microbiol.* **8**, 814–826.  
 Hotez, P. J., Hawdon, J. M., Cappello, M., Jones, B. F., Ghosh, K., Volvovitz, F. & Xiao, S. H. (1996). *Pediatr. Res.* **40**, 515–521.  
 Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.  
 Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.  
 Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.  
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.  
 McSorley, H. J. & Loukas, A. (2010). *Parasite Immunol.* **32**, 549–559.  
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.  
 Milne, T. J., Abbenante, G., Tyndall, J. D., Halliday, J. & Lewis, R. J. (2003). *J. Biol. Chem.* **278**, 31105–31110.  
 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.  
 Musgrove, P. & Hotez, P. J. (2009). *Health Aff. (Millwood)*, **28**, 1691–1706.  
 Sen, L., Ghosh, K., Bin, Z., Qiang, S., Thompson, M. G., Hawdon, J. M., Koski, R. A., Shuhua, X. & Hotez, P. J. (2000). *Vaccine*, **18**, 1096–1102.  
 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.  
 Silva, N. R. de, Brooker, S., Hotez, P. J., Montresor, A., Engels, D. & Savioli, L. (2003). *Trends Parasitol.* **19**, 547–551.  
 Suzuki, N., Yamazaki, Y., Brown, R. L., Fujimoto, Z., Morita, T. & Mizuno, H. (2008). *Acta Cryst.* **D64**, 1034–1042.  
 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.  
 Wang, Y.-L., Kuo, J.-H., Lee, S.-C., Liu, J.-S., Hsieh, Y.-C., Shih, Y.-T., Chen, C.-J., Chiu, J.-J. & Wu, W. (2010). *J. Biol. Chem.* **285**, 37872–37883.  
 Zhan, B., Liu, Y., Badamchian, M., Williamson, A., Feng, J., Loukas, A., Hawdon, J. M. & Hotez, P. J. (2003). *Int. J. Parasitol.* **33**, 897–907.