doi:10.1016/j.jmb.2004.12.023 J. Mol. Biol. (2005) 346, 801–814

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

X-ray Structure of *Na*-ASP-2, a Pathogenesis-related-1 Protein from the Nematode Parasite, *Necator americanus*, and a Vaccine Antigen for Human Hookworm Infection

Oluwatoyin A. Asojo[1]*, Gaddam Goud[2], Kajari Dhar[1], Alex Loukas[2,3] Bin Zhan[2], Vehid Deumic[2], Sen Liu[2], Gloria E. O. Borgstahl[1] and Peter J. Hotez[2]

[1]*Eppley Institute for Research in Cancer and Allied Diseases 987696 Nebraska Medical Center, Omaha, NE 68198-7696 USA*

[2]*Department of Microbiology and Tropical Medicine, The George Washington University Medical Center, Washington DC 20037, USA*

[3]*Division of Infectious Diseases and Immunology, Queensland Institute of Medical Research Brisbane, Qld. 4006, Australia*

**Corresponding author*

Human hookworm infection is a major cause of anemia and malnutrition of adults and children in the developing world. As part of on-going efforts to control hookworm infection, The Human Hookworm Vaccine Initiative has identified candidate vaccine antigens from the infective L3 larval stages of the parasite, including a family of pathogenesis-related (PR) proteins known as the Ancylostoma-secreted proteins (ASPs). A novel crystal structure of *Na*-ASP-2, a PR-1 protein secreted by infective larvae of the human hookworm *Necator americanus*, has been solved to resolution limits of 1.68 Å and to an *R*-factor of 17% using the recombinant protein expressed in and secreted by *Pichia pastoris*. The overall fold of *Na*-ASP-2 is a three-layer αβα sandwich flanked by an N-terminal loop and a short, cysteine-rich C terminus. Our structure reveals a large central cavity that is flanked by His129 and Glu106, two residues that are well conserved in all parasitic nematode L3 ASPs. *Na*-ASP-2 has structural and charge similarities to chemokines, which suggests that *Na*-ASP-2 may be an extra-cellular ligand of an unknown receptor. *Na*-ASP-2 is a useful homology model for NIF, a natural antagonistic ligand of CR3 receptor. From these modeling studies, possible binding modes were predicted. In addition, this first structure of a PR-1 protein from parasitic helminths may shed light on the molecular basis of host–parasite interactions.

Keywords: crystal structure; pathogenesis-related protein; hookworm; ASP; *Necator americanus*

Introduction

Hookworm infection is one of the three major soil-transmitted helminth (STH) infections of humans, and a leading global cause of anemia and malnutrition.[1] An estimated 740 million people are infected with hookworm in the tropical and subtropical regions of the developing world.[2] Currently, the major approach to STH infection control worldwide relies on the frequent and periodic use of anthelminthic drugs in order to reduce the number of worms in the gastrointestinal tract of humans, especially children.[3] Based on this principle, a resolution was passed at the 2001 World Health Assembly, which advocates the regular administration of anthelminthic drugs to at least 75% of all school-age children at risk for morbidity caused by STH infections†. In time, this would become the largest health program ever attempted.[4]

Among the reasons why school-based anthelminthic chemotherapy programs may not control

Abbreviations used: SEC-MALS, size-exclusion chromatography–multi-angle laser light-scattering; STH, soil-transmitted helminth; L3, third-stage larvae; ASP, Ancylostoma-secreted protein; PR, pathogenesis-related; CRISP, cysteine rich secretory protein; NIF, neutrophil inhibitory factor; MIDAS, metal ion-dependent adhesion site.

E-mail address of the corresponding author: oasojo@unmc.edu

† www.who.int/wormcontrol

0022-2836/$ - see front matter

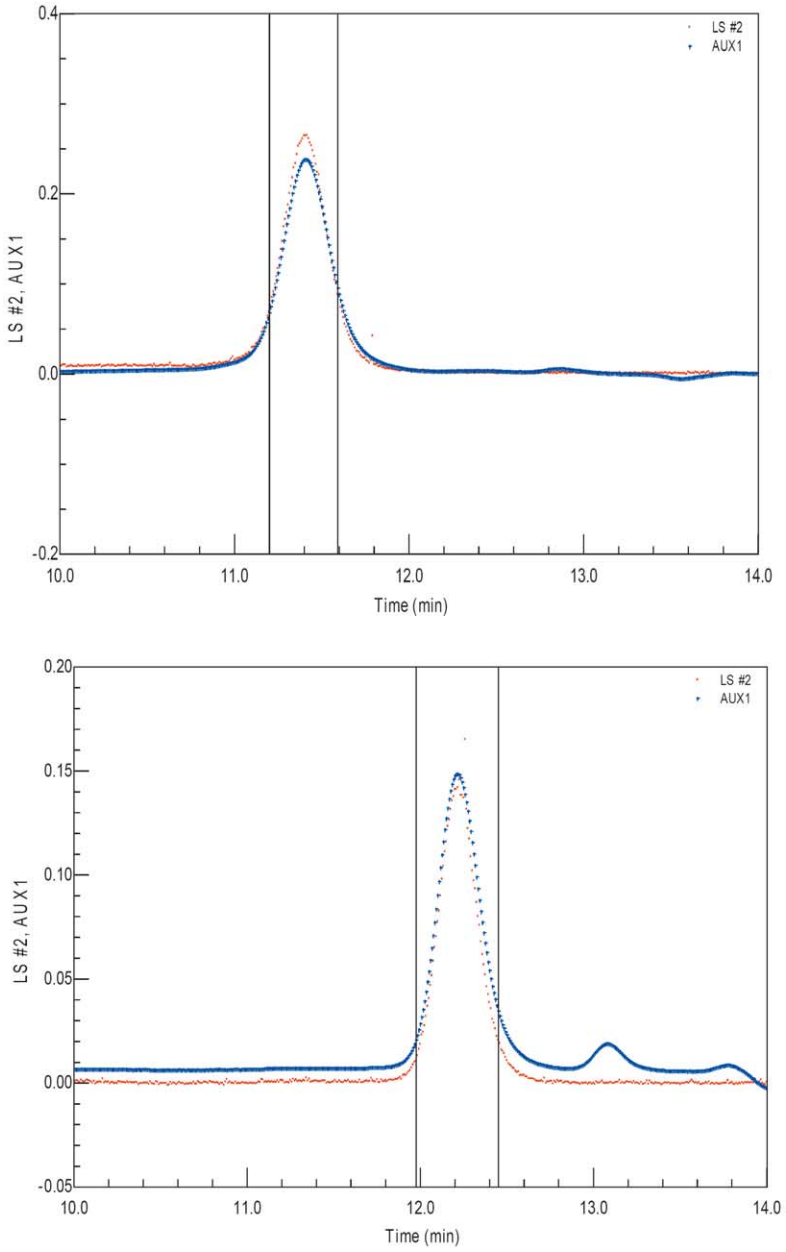

Figure 1. Representative SEC-MALS elution profiles for ASP-2. Light-scattering (red) and RI (blue) data are plotted against elution time for *Na*-ASP-2-His #1 (top) and *Na*-ASP-2 #2 (bottom). The numbers refer to the sample number and apply also to Table 1. Vertical lines represent the portion of the chromatogram analyzed by ASTRA4.0 software. The *Na*-ASP-2 protein samples eluted at 10–12 minutes.

hookworm infection effectively are: (1) the high rates of hookworm re-infection following drug treatment;[5] (2) the diminishing efficacy of the drug with repeated use,[6] possibly because of drug resistance;[7] and (3) the high prevalence and intensity of hookworm infection among adult populations, most notably women of reproductive age.[8]

As an alternative or complementary approach to hookworm control, an international effort known as The Human Hookworm Vaccine Initiative has been initiated to develop an anti-hookworm vaccine aimed at reducing worm burdens and intensity†.[9] The selection of antigens for hookworm vaccine development is based partly on the early observation that the infective stages of the parasite, third-stage larvae (L3), are immunogenic,[10] and that radiation-attenuated living L3 were used successfully as a veterinary vaccine for canine hookworm infection.[11] Therefore, a major goal of vaccine development for humans is to reproduce the effect of live attenuated L3 vaccines by substituting L3-derived chemically defined secreted antigens.[9] It was subsequently determined that the two major antigens secreted by L3 upon host entry, Ancylostoma-secreted protein 1 (ASP-1) and Ancylostoma-secreted protein 2 (ASP-2), are homologous to a subfamily of the pathogenesis-related (PR) protein superfamily.[9]

PR proteins were initially isolated from plants and are so named because they are produced in elevated levels upon pathogen-induced injury or other stress.[12] PR proteins are further classified into subfamilies (PR-1, PR-2, through PR-14).[13] The

† www.sabin.org

Table 1. Masses of ASP-2 samples determined by SEC-MALS

Sample	M_n (g/mol)	M_w (g/mol)	Relative polydispersity[a]
Na-ASP-2-His #1	26,080 (5)	26,180 (5)	1.004
Na-ASP-2-His #2	25,370 (3)	25,430 (3)	1.002
Na-ASP-2 #1	41,140 (10)	45,150 (8)	1.097
Na-ASP-2 #2	34,770 (4)	39,930 (4)	1.148

M_w is the weight-average molecular mass measured directly by light-scattering, M_n is the number-average molecular mass. The percentage error is given in parentheses.
[a] The ratio M_w/M_n gives a relative measure of polydispersity.

numbering is based, in part, on the order of characterization of the first member of each subfamily. While there is no shared homology among the PR protein superfamily, there is structural homology and shared functions within subfamilies, and each has its own characteristic conserved domains.[12] ASPs belong to the PR-1 subfamily, which is characterized by the presence of at least one PR-1 domain. The PR-1 domain has been reported in a diversity of proteins unrelated by phylogeny and isolated from bacteria, plants, animals and viruses.[14] The CRISPs (cysteine rich secretory proteins), which are expressed in mammalian epididymis and granules, belong to the PR-1 subfamily,[15] as do a diversity of venom allergens.[16,17] Specific members of the PR-1 subfamily include mammalian testis-specific protein,[15] venom antigen 5 (Ves v 5) from the wasp *Vespula vulgaris*,[18] and the human brain tumor proteins, P25TI[19] and GliPR.[20]

Members of the PR-1 subfamily are characterized by a highly conserved cysteine-rich PR-1 domain of about 15–16 kDa, and have been implicated in conditions requiring cellular defense or proliferation, such as plant responses to pathogens and tumor growth.[21–23] Members of the PR-1 family are capable of surviving harsh environments and are resistant to proteolytic digestion.[12,24] The diverse sources of PR-1 proteins have resulted in many proposed functions. For example, the substrate-specific endoprotease, Tex31, from the cone snail *Conus textile*, is a novel serine protease with known substrates.[13] Tex31 shares 27% sequence identity and 42% sequence similarity with ASP-2 from the major human hookworm *Necator americanus* (*Na*-ASP-2). Another PR-1 protein, P25TI is a novel trypsin inhibitor that is highly over-expressed in human neuroblastoma and glioblastoma.[19]

ASPs have been found in all parasitic nematodes studied to date.[21–23,25] Two major types of ASPs have been isolated from adult and larval nematodes, each containing either a single or double PR-1 domain.[21–23,26–29] However, the functions of ASPs are unknown. Since the larval ASPs are released specifically when stimulated by host-like conditions, it is likely that these ASPs play a role in the transition to parasitism and infectivity.[28] Vaccination of laboratory animals with a yeast-expressed recombinant ASP-2 from the hookworm *Ancylostoma ceylanicum* or *Ancylostoma caninum* resulted in reduced worm burdens, lowered fecundity in female worms and decreased host blood loss relative to controls following larval challenge.[30] It was further shown that antiserum to ASP-2 inhibits hookworm larval invasion *in vitro* (J. Bethony *et al.*, unpublished results). The ASP-1 orthologue from *Onchocerca volvulus*, the parasitic nematode that causes river-blindness, was recently shown to be a protective antigen.[31] Furthermore, human immunological investigations identified a subset of individuals with anti-ASP-2 antibody responses that harbor low-intensity hookworm infections, implying that naturally hookworm-resistant individuals selectively recognize ASP-2 but not other L3 antigens.[32] On the basis of these results, *Na*-ASP-2, was selected as a hookworm vaccine antigen to undergo process development in anticipation of clinical trials. Structural studies were

Table 2. Statistics for data collection and model refinement

Data	*Na*-ASP-2
Space group	*P*21
Resolution (Å)	39–1.68 (1.76–1.68)
R_{merge}[a] (%)	2.9 (8.4)
Completeness (%)	94.5 (64.3)
Redundancy	3.1 (2.1)
$I/\sigma(I)$	27.6 (6.3)
Refinement	
Resolution (Å)	25–1.68
R-factor[b] (*R*-free[c])	0.17 (0.22)
Correlation coefficient	
F_o-F_c	0.967
F_o-F_c free	0.945
Rms deviation	
Bond lengths (Å)	0.017
Bond angles (°)	1.716
Mean *B*-factor	
Protein (Å^2)	18.1
All atoms (Å^2)	27.0
Ramachandran plot	
Most preferred (%)	90.9
Allowed (%)	9.1
Model composition	
Monomers	1
Residues	193
Water molecules	210

Values within parentheses are for the highest-resolution shell.
[a] $R_{merge}=(\Sigma|I-\langle I\rangle|)/\Sigma I$, where I is the observed intensity, and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.
[b] *R*-factor $=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o are F_c are observed and calculated structure factors, respectively.
[c] *R*-free set uses 5% of randomly chosen reflections.[62]

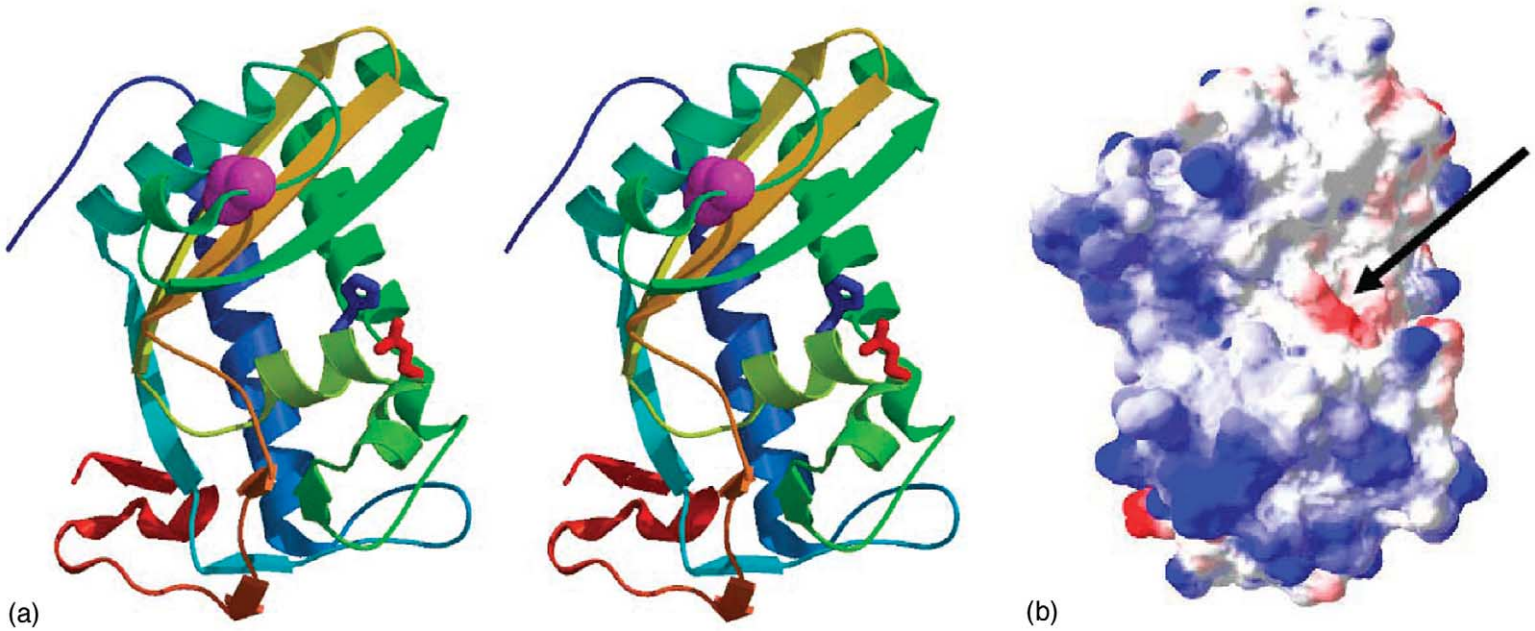

Figure 2. (a) Stereo-view ribbon diagram of *Na*-ASP-2 structure colored from blue (N terminus) to red (C terminus). *Na*-ASP-2 is folded as a typical three-layer αβα sandwich, consisting of a three-stranded antiparallel β sheet sandwiched between a layer with solitary α helix and a second layer composed of two parallel α helices. Ser70 (magenta) is incapable of interacting with the conserved His129 (blue) and Glu106 (red), to form the putative catalytic triad. (b) Charge distribution on the surface of *Na*-ASP-2 showing the solvent-exposed, negatively charged putative binding cavity (indicated by the black arrow), which is spanned by the conserved residues His129 and Glu106. (a) and (b) are in the same orientation.

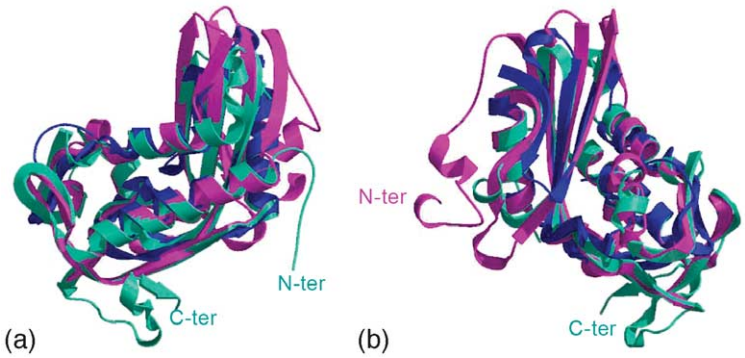

Figure 3. Two views of an overlay of the structure of *Na*-ASP-2 (aquamarine) with the structures of other known PR-1 proteins, Ves v5 (magenta) and p14a (blue), which reveals that the conserved PR-1 domain is a two α helix/three-stranded β sheet/one α helix sandwich. *Na*-ASP-2 is more helical than both structures and has a unique C-terminal extension. The right-hand side is rotated 180° to show the unique N-terminal extension of Ves v5.

initiated recently to clarify the roles of *Na*-ASP-2 as a functional vaccine. We present here the high-resolution structure of *Na*-ASP-2, which provides new insights into the role of a parasite PR-1 protein.

Results and Discussion

Molecular mass determination

The absolute molecular masses of samples of *Na*-ASP-2 and *Na*-ASP-2 with a C-terminal His-tag (*Na*-ASP-2-His) were measured by size-exclusion chromatography and multi-angle laser light-scattering (SEC-MALS). Both samples gave single peaks on the sizing column with different elution times (Figure 1). The light scattered by a protein is directly proportional to its weight-average molecular mass and its concentration. Clear differences were observed in the masses measured by light-scattering (Table 1). These results indicate that *Na*-ASP-2 forms a dimer in solution, whereas *Na*-ASP-2-His forms a monomer in solution.

Crystallization and structure determination

Initial crystallization screens were carried out on samples of *Na*-ASP-2-His. *Na*-ASP-2-His was very soluble and no crystal or crystalline precipitate was generated in any of the crystallization trials. However, crystals that diffracted well were obtained with *Na*-ASP-2 in a variety of conditions. Apparently, dimerization in solution may be important for the crystallization of *Na*-ASP-2.

Na-ASP-2 crystals were plate-like rods that were only 0.05 mm in the smallest dimension. Despite their small size, the best crystals had visible diffraction spots beyond 1.3 Å. Due to radiation damage, data to 1.56 Å were collected (Table 2). Most of the residues were clearly visible in a $2F_o - F_c$ omit electron density map phased by molecular replacement. However, because of the high resolution of the data, 180 residues of the model were built automatically using the program ARP-WARP.[33–40] The remaining residues and water molecules were built through iterative cycles of model building in O,[41] followed by structure refinement in REFMAC-5.[42–44] The structure is reported to 1.68 Å following conventions requiring completeness of at least 50% in the outermost shell. The final model has an *R*-factor of 17% (free *R*-factor of 22.0%) and includes 191 residues of the *Na*-ASP-2 and 209 water molecules. On the N terminus, two additional residues from the plasmid are visible. The main-chain and side-chain stereochemistry of the refined model is excellent, and 100% of main-chain φ, ψ angles lie within the allowed region of a Ramachandran plot (90.9% in the most-favored region and 9.1% in additionally allowed regions). No residue lies in the generously allowed region or the disallowed region. There is no bad contact or unusual fold in the model. More details of the quality of the structure as well as data collection are shown in Table 2.

Overall structure

This *Na*-ASP-2 structure offers the first three-dimensional atomic view of an ASP. *Na*-ASP-2 is a representative single-domain ASP, comprised of an N-terminal loop, a PR-1 domain, and a short, cysteine-rich C terminus (Figure 2). The tertiary structure of the PR-1 domain is a three-layer αβα sandwich, in which a three-stranded antiparallel β sheet (formed by residues 79–86, 140–148, and 151–159), lies between two layers of helices. One helical layer is composed of two parallel helices (formed by residues 9–31, and 92–115), while the other has a lone helix (residues 54–66). The short C terminus is folded into a loop structure, followed by helical turn, a short strand and terminating in a random coil (Figure 2).

Comparison with other PR-1 structures

The molecular replacement search model, Ves v 5 (PDB code 1QNX)[45] and the NMR structure of p14a, a PR-1 protein from tomato (PDB code 1CFE)[46] are the only other reported structures of proteins containing a PR-1 domain. The PR-1 domain of *Na*-ASP-2 aligns well with both and confirms that the core PR1-domain is a three-layer αβ sandwich composed of three helices and three strands (Figure 3). The rms deviation for the overlay of the main-chain atoms of the PR-1 domain of *Na*-ASP-2 (residues 8–161) with Ves v 5 (residues

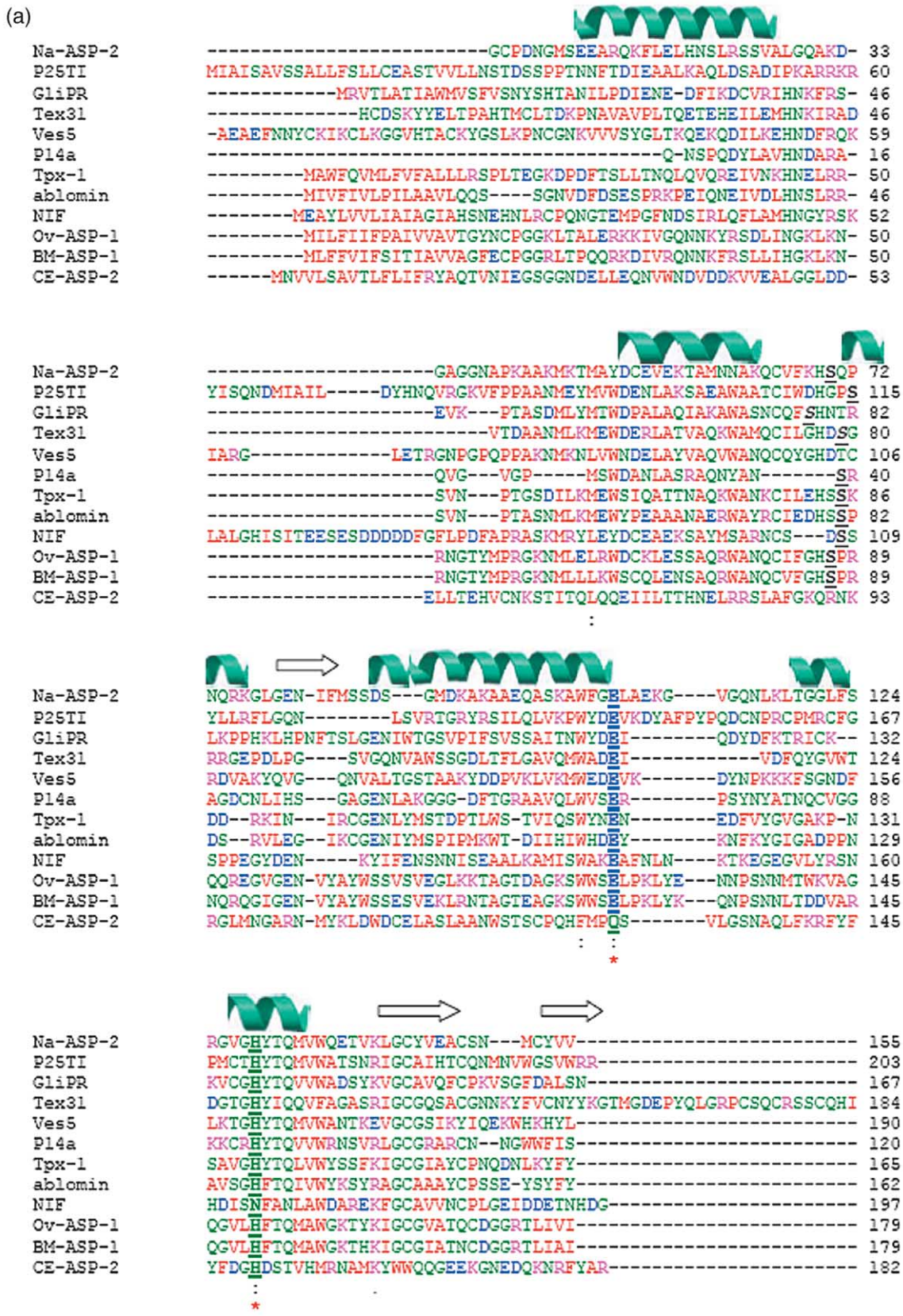

Figure 4 (*legend next page*)

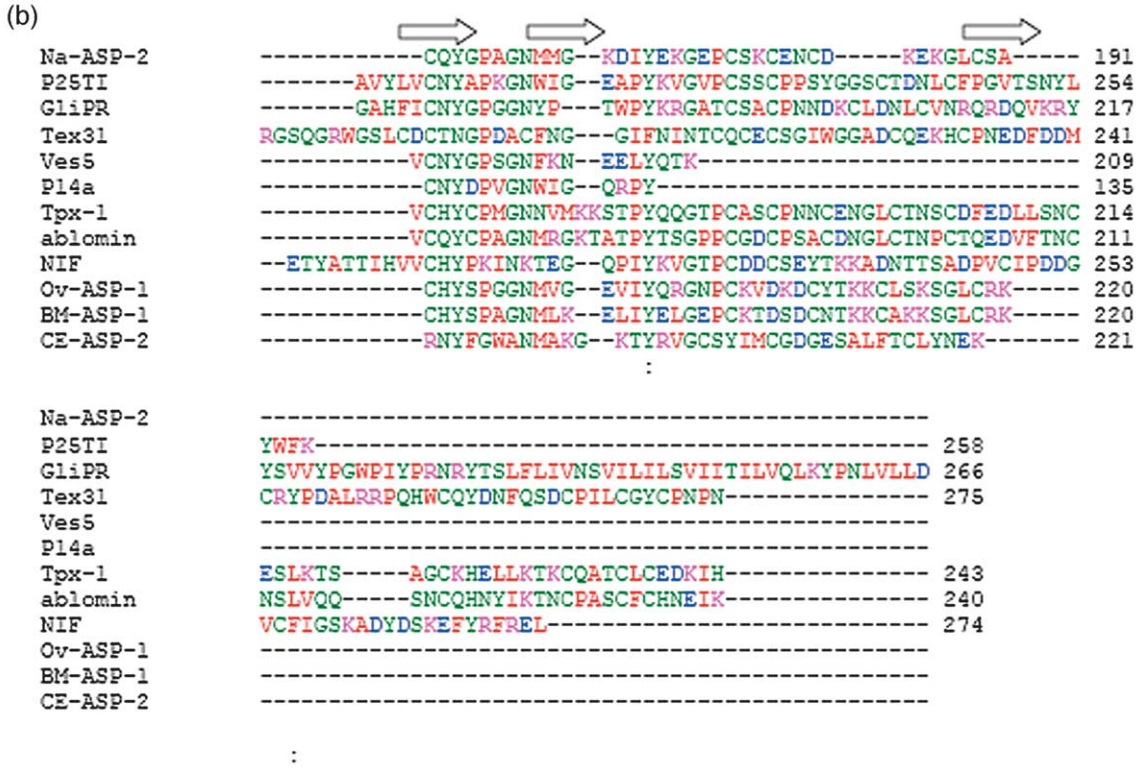

Figure 4. Clustal_W alignment of PR-1 proteins with *Na*-ASP-2. The only ASP that is poorly conserved in both key sites (bold, underlined and red asterisk) is of the free-living *Caenorhabditis elegans* (CE-ASP-2), which does not transition to parasitism. Also shown is the somewhat conserved proposed catalytic Ser (bold, underlined and black font). Tex31; serine protease from the cone snail (CAD36507); Ves5, antigen Ves v5 from *Vespula vilgaris* (Q05110); P14a from cherry tomato (ID code P04284); NIF, neutrophil inhibitory factor from *A. caninum* (ID code AAA27789.1); GliPR from brain tumor (ID code NP 006841); ablomin from snake venom (Q8J140); Tpx-1 from human testis (P16563); as well as the following ASPs; Ov-ASP-1, BM-ASP-1 and CE-ASP-2, *Onchocerca volvulus* (AAP06732), *Brugia malayi* (AAK12274) and *C. elegans* (NP_509803). PUBMED protein ID codes are shown in parentheses (www.pubmed.org). Amino acid residues are colored as follows: red for small + hydrophobic (AVFPMILW); blue for acidic (DE); magenta for basic (RHK); and green for hydroxyl + amine + basic (STYHCNGQ). Key: * indicate sites in all aligned sequences with identical residues, while . and : indicate conservatively mutated positions, there is a higher level of conservation for the latter. Regions of secondary structure are shown for *Na*-ASP-2, with helices in green and strands as arrows.

(37–192) and p14a (residues 4–125) are 1.03 Å and 0.98 Å, respectively.

There are distinct differences between the three structures. *Na*-ASP-2 has a unique C-terminal extension of 20 residues (Figure 3). *Na*-ASP-2 and p14a have a β sheet composed of three strands; however, Ves v 5 has a four-stranded β sheet (Figure 3(a)). The fourth strand of Ves v 5 forms an integral part of a much longer N terminus, which includes a unique helix (Figure 3(b)). The N terminus of *Na*-ASP-2 has a short hydrophobic loop following the conserved helix. This helix is comparable in length to that found in Ves v 5 but two turns longer than that in p14a. In addition, *Na*-ASP-2 has strands that are shorter than those of Ves v 5 but longer than those of p14a. Overall, *Na*-ASP-2 has a more helical structure, composed of longer conserved helices and additional helices (Figure 3). Despite these differences, the core of all three structures has a wide, putative binding cavity spanned by two conserved residues, Glu106 and His129 (Figure 2).

Structure–function analysis of *Na*-ASP-2

Putative binding cavity

Our structure represents the uncomplexed or "free" conformation of *Na*-ASP-2. A series of conserved residues that are characteristic of PR-1 proteins cluster around an electronegative cavity (Figure 2(b)). This putative binding cavity has been speculated to play a role in PR-1 function.[13] A closer examination of the cavity reveals a network of interactions involving the conserved residues. There is a 2.54 Å hydrogen bond from Glu106($O^{\varepsilon 2}$)to His129($N^{\delta 1}$). His129($N^{\varepsilon 2}$) also forms a 2.60 Å hydrogen bond with a proximal water molecule. This water molecule also has a 2.62 Å interaction with the well conserved His69($N^{\varepsilon 2}$). In addition, His69($N^{\varepsilon 2}$) has a 2.80 Å hydrogen bond with Glu80($O^{\varepsilon 2}$). The interactions in the putative binding cavity are not limited to the preceding intra-molecular interactions. There are two notable inter-molecular interactions with His69 from

Glu180($O^{\varepsilon 1}$) and Asn181($O^{\delta 1}$) at distances of 4.11 Å and 3.11 Å, respectively. These interactions between monomers may be the basis for dimerization in solution. The extensive and strong interactions may stabilize the putative binding cavity in a free conformation.

Putative proteolytic activity

While it is accepted that *Na*-ASP-2 is involved in the transition to parasitism and infectivity,[21] what role it plays remains unclear. Structural homology to Tex31 from cone snails led Milne and colleagues to suggest that PR-1 proteins may be substrate-specific proteases, and that this activity involves the conserved putative binding cavity.[13] The three-dimensional structure of Tex31 is unknown, and sequence alignments and comparisons reveal that the residues in the putative catalytic triad correspond to His129, Glu106, and Ser70 in *Na*-ASP-2 (Figure 4). His129 and Glu106 are conserved in many ASPs from infective nematodes, including *O. volvulus*[31,47] and *Brugia malayi*.[48] In fact, the only ASP that is poorly substituted at both sites is that of the free-living *Caenorhabditis elegans* (Figure 4). In the *Na*-ASP-2 structure, Ser70 is too far away and pointing away from the other putative members of the catalytic triad (Figure 2). Ser70(O^{γ}) is over 12.00 Å from His129($N^{\varepsilon 2}$), instead forming a hydrogen bond (at 2.62 Å distance) with Glu80($O^{\varepsilon 2}$). In addition, the backbone is held in this position by a hydrogen bond between Arg75($N^{\eta 2}$) and Ser70(O) (at 2.91 Å distance). There is a second hydrogen bond between Glu80($O^{\varepsilon 1}$) and Ser70(N) (at 2.81 Å distance). For *Na*-ASP-2 to behave as a Ser protease, extensive conformational changes must occur upon binding to bring Ser70 closer to His106 and Glu129, to form a typical serine protease catalytic triad. Ser70 lies on a loop that may allow some conformational flexibility. However, it is unlikely that the main chains of Glu106 and His129 will move appreciably, as they both lie in the middle of helices. Furthermore, Ser70 lies in a variable loop region, outside the structurally conserved PR domain. All other Ser residues in the structure can be dismissed with similar arguments.

The PR-1 proteins as Ser protease hypothesis is further weakened by the observation *N*a-ASP-2 has no proteolytic activity as measured by gelatin and casein assays. We conducted additional tests for proteolytic activity including screening of a positional scanning combinatorial synthetic peptide library (C. Craik & Y. Choe, University of California San Francisco, unpublished results). Recombinant *Na*-ASP-2 was screened against this library using fixed P1 residues. These residues (amongst others) were chosen on the basis of the results of studies by Milne and colleagues, which showed that Tex31 has serine protease activity, despite not having the usual architecture associated with typical Ser proteases.[13] Libraries were screened with Arg or Lys at the P1 position under conditions that were optimal for Tex31; i.e. neutral pH, some salts and divalent cations. The lack of proteolytic activity suggests that if *Na*-ASP-2 is a protease, it is a highly substrate-specific protease with specificity different from that of Tex31. It is clear that *Na*-ASP-2 has an available cavity to bind substrates; however, it appears unlikely that it would cleave bound substrates using known catalytic mechanisms.

Chemotaxin mimicry

It is possible that PR-1 proteins are secreted extracellular ligands or inhibitors of G-protein receptors. This behavior has been observed from PR-1 toxins such as ablomin, helothermine, triflin and latisemin that block voltage-gated calcium and potassium channels, and ryanodine receptors.[16,17] There are no known structures of these proteins. G-protein receptor ligands are typically characterized by similar charge distributions and possibly regional structural similarities. Peptide size and fold is apparently not as important to ligand recognition. An examination reveals structural and charge similarities of *Na*-ASP-2 with CC-chemokines.

Chemokines are secreted G protein-coupled chemotaxins that cause the influx of blood cells, like T and B lymphocytes, monocytes, neutrophils, eosinophils and basophils, in allergic and other inflammatory conditions. CC-chemokines have a characteristic sequential Cys-Cys motif. CC-chemokines have a conserved structure of three-stranded antiparallel β sheets flanked by a helix. This is quite similar to the central sheet and one of the helices of *Na*-ASP-2 (Figure 5). The large number of cysteine residues and disulfide bonds is also a common thread between the two families. A comparison of the charge distribution of *Na*-ASP-2 with a representative chemokine, TARC, revealed striking similarities. Most notably, the almost complete separation of the positive and negative charges, where one side is almost completely electronegative, while the other is almost completely electropositive (Figure 5). Many CC-chemokines have a similar charge distribution.[49] These similarities suggest that *Na*-ASP-2 may be a chemokine mimic, possibly competing with these chemotaxins for receptor binding. If confirmed, this observation could help explain some of the immuno-modulating properties of hookworm parasites.[50,51] Chemokine mimics have been identified from pathogenic viruses, where these mimics aid in the evasion of immune responses by the viruses.[52,53] It is plausible that nematode ASPs secreted upon host entry perform a similar role.

An ASP as an antagonistic ligand of CR3 receptor

Helminths have co-evolved with their hosts; consequently, they possess unique mechanisms to sustain their infectivity while evading detection by the host's immune system. One mechanism involves the binding of the antagonistic ligands to the CR3 receptor, which prevents the binding of

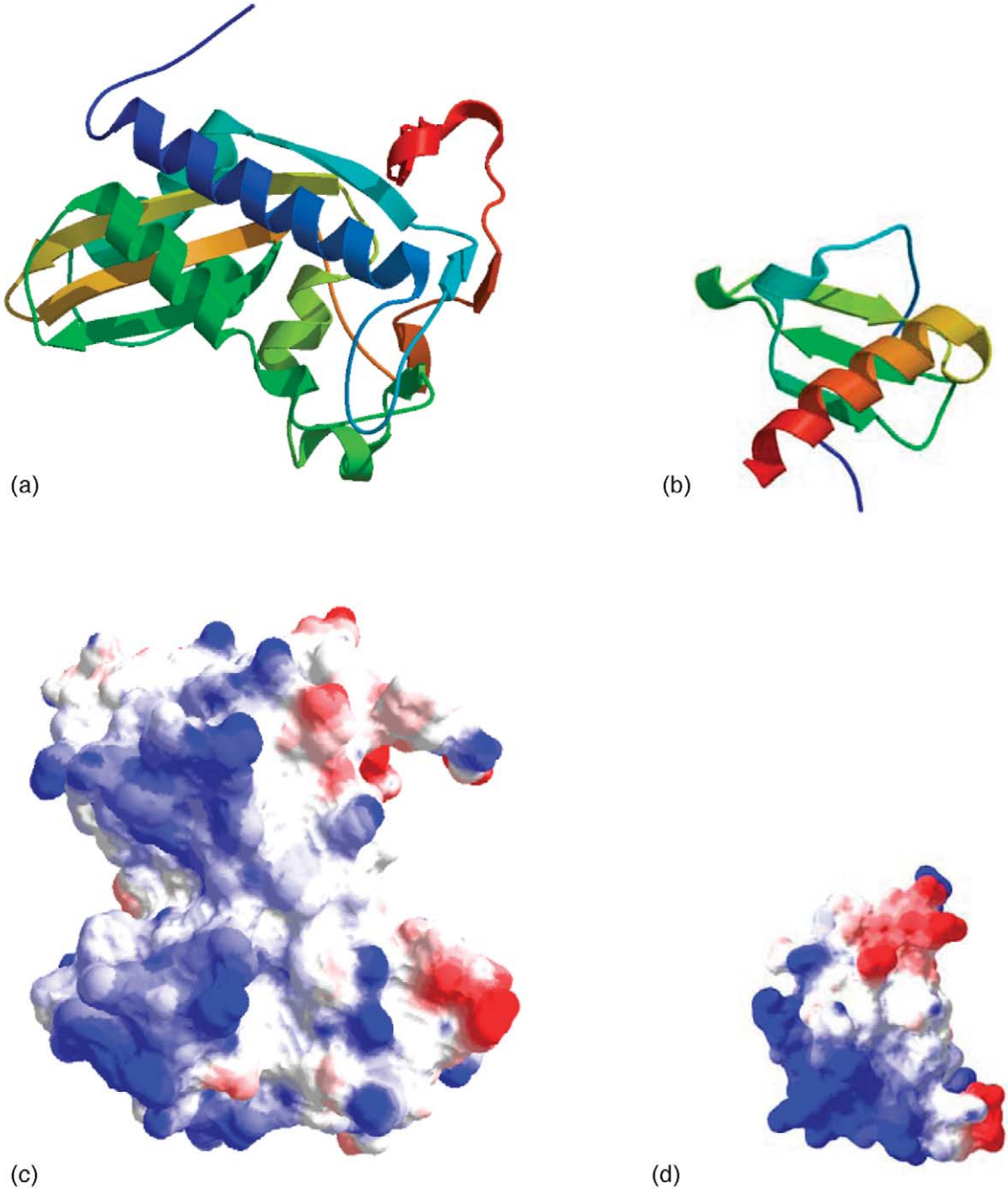

Figure 5. Structural and charge mimicry of CC-chemokines. Comparison of *Na*-ASP-2 ((a) and (c)) with a representative CC-chemokine, TARC ((b) and (d)). The ribbon diagram reveals a similar fold of a three-stranded antiparallel sheet underneath a helical region, or an open αβ sandwich. Both have a similar separation of charges, where one side is almost completely electronegative while the other is almost completely electropositive.

chemotaxins, consequently altering the immune cascade.[54] The ASP-like protein from the dog hookworm, *A. caninum*, neutrophil inhibitory factor (NIF), is an antagonistic ligand that binds on the metal-ion dependent adhesion site (MIDAS) of the extracellular A domain of the CR3 receptor.[55] Essentially, NIF is a secreted extracellular ligand of the CR3 receptor that affects the migration of monocytes, thus providing a mechanism for immune evasion. We have data that suggests that *Na*-ASP-2 binds to monocytes (A.L. & P.J.H., unpublished results). Even if *Na*-ASP-2 does not bind to CR3, understanding how NIF binds will elucidate modes of receptor binding by a PR-1 protein. NIF has 26% sequence identity (41% sequence homology) with *Na*-ASP-2. Although the three-dimensional structure of NIF is unknown, the structure of *Na*-ASP-2 will provide insights into NIF function, through homology modeling and other analyses.

Figure 6. Superposition of the structure of *Na*-ASP-2 (aquamarine) with the homology model of NIF (magenta). The conserved Glu142 of NIF (red) is more exposed and lies along the edge of a longer helix than Glu109 of *Na*-ASP-2 (blue). This exposure and helical extension may facilitate a tighter fit with CR3 integrin.

The structure of NIF was homology modeled using the MOE package (CCG. MOE, version 2004.3. Chemical Computing Group, Montreal, Canada), with the structure of *Na*-ASP-2 as a starting model. The model of NIF, as expected, retains the core PR-1 domain (Figure 6). The core PR-1 domain includes a series of highly conserved residues from all parasitic ASPs. The key binding motif of all known integrin ligands is an acidic residue.[56] Furthermore, the crystal structure of extracellular CR3 domain A reveals a glutamate side-chain (Glu314) from a second domain A acting possibly as a ligand mimetic,[55] and it is possible that this role is played by the highly conserved Glu142, which corresponds to Glu106 of *Na*-ASP-2 (Figure 7). There is ample space for the helix containing Glu142 to fit into the MIDAS. Alignment of the helix containing Glu142 with the corresponding helix from the symmetrically related model containing Glu314 reveals a possible mode of binding by NIF with the CR3 domain in which the helix is completely enveloped by the MIDAS of CR3 and there is extensive buried surface between both, which is indicative of a strong and preferred interaction (Figure 8). This mode of binding maintains the binding regions predicted from non-structural data.[57] It is important to point out that *Na*-ASP-2 will not "fit" or interact structurally as well as NIF with the CR3 receptor. The apparent advantage of NIF is because Glu142 of NIF lies on the edge of a longer helix and is more exposed on the protein surface than Glu106 of *Na*-ASP-2. This suggests that different ASPs may modulate immune responses by binding to different receptors. This hypothesis is supported by the fact that some PR-1 toxins (e.g. ablomin) bind to other receptors.[16,17]

Concluding remarks

The three-dimensional structure of the vaccine antigen *Na*-ASP-2 offers key insights into its possible functions. This is a novel structure of the ASPs, a family of proteins that are important in the transition to parasitism and infectivity of hookworm and other parasitic nematodes, and serves as a representative structure for homology modeling and computational studies of other ASPs. The three-dimensional structure indicates that *Na*-ASP-2 in the free form is not a typical Ser protease. The structural and charge mimicry of chemotaxins seen with ASP-2 might explain why these proteins are proving to be efficacious anti-nematode vaccines; vaccination might induce antibodies that block or interfere with host–parasite ligand–receptor interactions and thereby interrupt the ability of the parasite to modulate the host immune response, making it more vulnerable to immune attack.

Experimental Procedures

Molecular cloning and expression

A cDNA library of infective third-stage larvae (L3) of *N. americanus* was constructed as described.[29,30] cDNA encoding *Na*-ASP-2 was isolated by immunoscreening of an expression cDNA library using antiserum against *Ac*-ASP-2 from *A. caninum* followed by cDNA cloning and sequencing.[29] The entire coding sequence minus the N-terminal signal peptides of *Na*-ASP-2 was PCR amplified from the first strand cDNA of *N. americanus* L3 with *Na*-ASP-2 gene-specific primers. The primers Na-ASP2-F1 (GGG AAT TCG GTT GTC CTG ACA ATG GAA T) and Na-ASP2-R1 (TGT CTA GAG CAC TGC AGA GTC CCT TCT C) were used to clone *Na*-ASP-2 with a His-tag at the C terminus; while primers Na-ASP-2-F1 and Na-ASP-2-R2 (TGT CTA GAT CAA GCA CTG CAG AGT CCC TTC TC) were used to clone *Na*-ASP-2 without a His-tag at the C terminus. The PCR products were sub-cloned into the pichia expression vector pPIC-Zα (Invitrogen) *via* the EcoRI and XbaI sites. The correct insert and right reading frame were confirmed by double-strand sequencing of the recombinant plasmid using flanking vector primer: α-factor and 3′ AOX1. Fermentation and expression were carried out

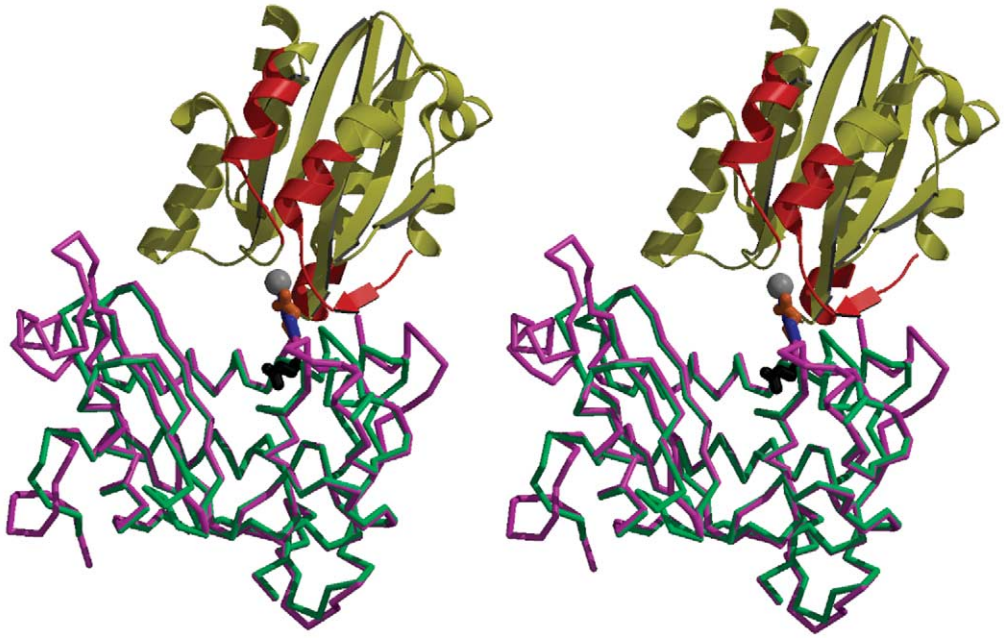

Figure 7. Stereo-view of a proposed mode of ASP interaction with the CR3 A-domain (PDB code1IDO), which is shown as gold ribbon, with the predicted NIF-binding region shown in red. *Na*-ASP-2 and NIF are shown as aquamarine and magenta line-trace, respectively. Glu142 of NIF (orange) overlays well with the ligand mimic Glu314 from the symmetry-related CR3 molecule (blue). Glu109 of *Na*-ASP-2 (black) will not interact with CR3 but instead points away from the Mg co-factor (gray sphere).

according to G.G. *et al.* (unpublished) and the protocols described for *Ay*-ASPs.[30]

Size-exclusion chromatography and multi-angle light-scattering

The SEC-MALS experiments were performed by loading 200 µg of protein sample onto a Shodex KW-803 column (JM Science, Grand Island, NY) at flow-rate of 0.5 ml min^{-1} using an Agilent HPLC. The column buffer was 10 mM Hepes (pH 7.8), 100 mM KCl, 0.25% (w/v) inositol, 1 mM EDTA. A UV detector (Agilent), a miniDAWN triple-angle light-scattering detector (Wyatt Technology) and an Optilab DSP interferometric refractometer (Wyatt Technology) were connected in series downstream from the column. The refractometer provided a continuous index of protein concentration. A d_n/d_c (refractive index increment) value of 0.185 ml mg^{-1} was used. Bovine serum albumin was used as an isotropic scatterer for detector normalization. The light scattered by a protein is directly proportional to its weight-average molecular mass and concentration. Therefore, molecular masses were

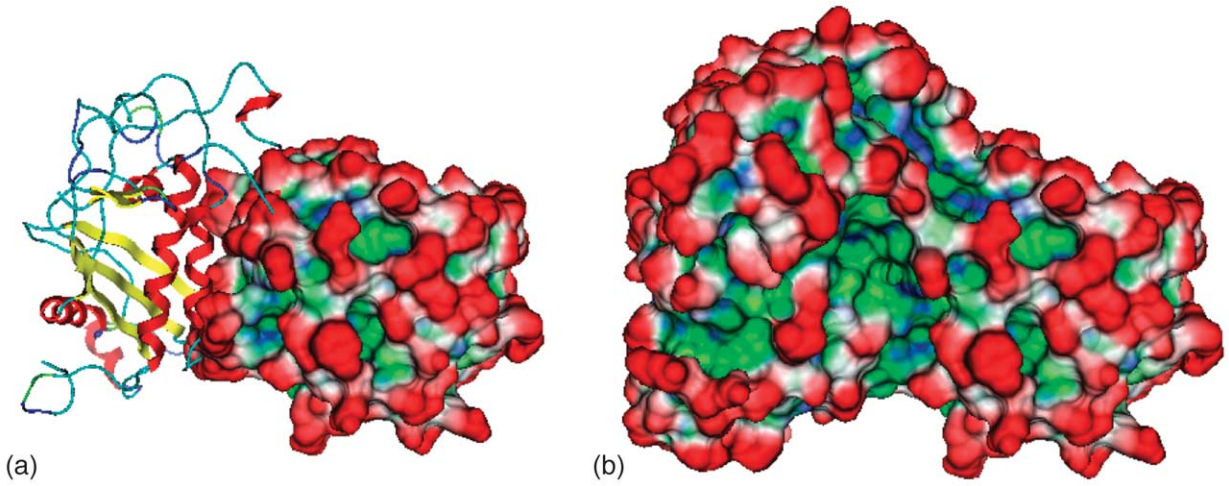

Figure 8. Possible mode of interaction of NIF with CR3. (a) The helix containing Glu142 is engulfed completely by a hollow cavity on the surface of CR3, which results in (b) extensive overlap between the surfaces of NIF and CR3. Exposed surfaces are colored red, while hydrophobic and hydrophilic pockets are colored green and blue, respectively.

calculated from the light-scattering and interferometric refractometer data using ASTRA 4.0 software.

Crystallization and structure determination

Crystals were grown at 20 °C by vapor-diffusion in sitting drops. All crystals grew in high concentrations (30–45%, v/v) of PEG200, PEG400 or PEG600 in the presence of a low concentration of basic buffer (50–150 mM Tris, Ches, imidazole, or Hepes at pH 7.5–10.5). The crystals that diffracted best were obtained by vapor-diffusion, from equilibrating either a sitting drop or a hanging drop containing a mixture of 3 µl (7.5 mg ml^{-1}) protein and 1.5 µl of reservoir solution with a reservoir containing 40% (v/v) PEG400, 100 mM imidazole (pH 8.5). Plate-like rod crystals of dimensions 1.0 mm × 0.3 mm × 0.05 mm grew overnight.

Crystals were cryo-cooled in a stream of nitrogen gas at −160 °C (X-stream 2000 low-temperature system, RigakuMSC) before data collection. The X-ray system consisted of an FRE-superbright rotating anode generator (RigakuMSC) operating at 45 kV and 45 mA, with Osmic Micromax optics and an R-axis IV++ image plate detector (RigakuMSC). A complete data set was collected from a single crystal using a crystal-to-detector distance of 100 mm and exposure times of two minutes for 0.5° oscillations. All X-ray data sets were collected and processed using the Crystal Clear (d*trek) package.[58] The space group was $P2_12_12_1$ with cell constants approximately $a = 37$ Å, $b = 51$ Å, $c = 43$ Å and a monomer in the crystallographic asymmetric unit. The structure was solved by molecular replacement using AMoRe,[59] using a truncated structure of Ves v5 (PDB code 1QNX)[45] as the search model. The model was built using 200 cycles of ARP/WARP version 5 in the automatic build and slow mode,[40] with free-*R*,[60] followed by iterative cycles of manual model building with the program O,[41] and structure refinement with REFMAC5,[43,44] using a maximum likelihood refinement procedure with Engh & Huber geometric parameters and keeping the same *R*-free set. The refined model statistics are shown in Table 2.

The superposition of models and rms deviations were calculated using the program Swiss-PdbViewer version 3.7.[61,62] Homology modeling was performed using the MOE package (version 2004.03, homology model). All Figures were generated using Raster3D,[63] BOBSCRIPT[64] and MOLSCRIPT,[65] except surfaces that were generated using GRASP[66] and MOE.

Data Bank accession numbers

The atomic coordinates and structure factors of *Na*-ASP-2 have been deposited with the RCSB Protein Data Bank with accession code 1U53. The *Na*-ASP-2 cDNA sequence has been deposited in GenBank under accession number AY288089.

Acknowledgements

We thank Dr Youngchool Choe and Dr Charles S. Craik, at the University of California San Francisco, for screening *Na*-ASP-2 for proteolytic activity against an extensive peptide library. We thank Mr Gilson Baia, Mr Oluwatosin Louis A. Asojo, Ms Abigail Henderson, Dr Nicole La-Rhonde-LeBlanc and Dr Istvan Botos for their support, useful discussion, assistance and encouragement. This research was supported, in part, by the Nebraska Tobacco Settlement Biomedical Research Development Fund (GEOB), which funds the X-ray facility. The Human Hookworm Vaccine Initiative is supported by a grant from The Bill and Melinda Gates Foundation awarded to the Sabin Vaccine Institute and George Washington University Medical Center.

References

1. Hotez, P. J., Brooker, S., Bethony, J. M., Bottazzi, M. E., Loukas, A. & Xiao, S. (2004). Hookworm infection. *N. Engl. J. Med.* **351**, 799–807.
2. de Silva, N. R., Brooker, S., Hotez, P. J., Montresor, A., Engels, D. & Savioli, L. (2003). Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol.* **19**, 547–551.
3. WHO. (2002). Prevention control of schistosomiasis and soil-transmitted helminthiasis. *World Health Organ. Tech. Rep. Ser.* **912**, 1–57.
4. Horton, J. (2003). Global anthelmintic chemotherapy programs: learning from history. *Trends Parasitol.* **19**, 405–409.
5. Albonico, M., Smith, P. G., Ercole, E., Hall, A., Chwaya, H. M., Alawi, K. S. & Savioli, L. (1995). Rate of reinfection with intestinal nematodes after treatment of children with mebendazole or albendazole in a highly endemic area. *Trans. Roy. Soc. Trop. Med. Hyg.* **89**, 538–541.
6. Albonico, M., Bickle, Q., Ramsan, M., Montresor, A., Savioli, L. & Taylor, M. (2003). Efficacy of mebendazole and levamisole alone or in combination against intestinal nematode infections after repeated targeted mebendazole treatment in Zanzibar. *Bull. World Health Organ.* **81**, 343–352.
7. Albonico, M. (2003). Methods to sustain drug efficacy in helminth control programmes. *Acta Trop.* **86**, 233–242.
8. Bundy, D. A., Chan, M. S. & Savioli, L. (1995). Hookworm infection in pregnancy. *Trans. Roy. Soc. Trop. Med. Hyg.* **89**, 521–522.
9. Hotez, P. J., Zhan, B., Bethony, J. M., Loukas, A., Williamson, A., Goud, G. N. *et al.* (2003). Progress in the development of a recombinant vaccine for human hookworm disease: the Human Hookworm Vaccine Initiative. *Int. J. Parasitol.* **33**, 1245–1258.
10. Miller, T. A. (1971). Vaccination against the canine hookworm diseases. *Advan. Parasitol.* **9**, 153–183.
11. Miller, T. A. (1978). Industrial development and field use of the canine hookworm vaccine. *Advan. Parasitol.* **16**, 333–342.
12. Van Loon, L. C. & Van Strien, E. A. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* **55**, 85–97.
13. Milne, T. J., Abbenante, G., Tyndall, J. D., Halliday, J.

& Lewis, R. J. (2003). Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J. Biol. Chem.* **278**, 31105–31110.
14. Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. (2002). CDART: protein homology by domain architecture. *Genome Res.* **12**, 1619–1623.
15. Kasahara, M., Gutknecht, J., Brew, K., Spurr, N. & Goodfellow, P. N. (1989). Cloning and mapping of a testis-specific gene with sequence similarity to a sperm-coating glycoprotein gene. *Genomics*, **5**, 527–534.
16. Yamazaki, Y., Hyodo, F. & Morita, T. (2003). Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. *Arch. Biochem. Biophys.* **412**, 133–141.
17. Yamazaki, Y., Koike, H., Sugiyama, Y., Motoyoshi, K., Wada, T., Hishinuma, S. *et al*. (2002). Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. *Eur. J. Biochem.* **269**, 2708–2715.
18. Lu, G., Villalba, M., Coscia, M. R., Hoffman, D. R. & King, T. P. (1993). Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. *J. Immunol.* **150**, 2823–2830.
19. Yamakawa, T., Miyata, S., Ogawa, N., Koshikawa, N., Yasumitsu, H., Kanamori, T. & Miyazaki, K. (1998). cDNA cloning of a novel trypsin inhibitor with similarity to pathogenesis-related proteins, and its frequent expression in human brain cancer cells. *Biochim. Biophys. Acta*, **1395**, 202–208.
20. Szyperski, T., Fernandez, C., Mumenthaler, C. & Wuthrich, K. (1998). Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system. *Proc. Natl Acad Sci. USA*, **95**, 2262–2266.
21. Hawdon, J. M., Narasimhan, S. & Hotez, P. J. (1999). Ancylostoma secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. *Mol. Biochem. Parasitol.* **99**, 149–165.
22. Zhan, B., Liu, Y., Badamchian, M., Williamson, A., Feng, J., Loukas, A. *et al*. (2003). Molecular characterisation of the Ancylostoma-secreted protein family from the adult stage of *Ancylostoma caninum*. *Int. J. Parasitol.* **33**, 897–907.
23. Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J. & Hussey, R. S. (2001). Molecular characterisation and expression of two venom allergen-like protein genes in Heterodera glycines. *Int. J. Parasitol.* **31**, 1617–1625.
24. Hidalgo, P., Garreton, V., Berrios, C. G., Ojeda, H., Jordana, X. & Holuigue, L. (2001). A nuclear casein kinase 2 activity is involved in early events of transcriptional activation induced by salicylic acid in tobacco. *Plant Physiol.* **125**, 396–405.
25. Ding, X., Shields, J., Allen, R. & Hussey, R. S. (2000). Molecular cloning and characterisation of a venom allergen AG5-like cDNA from *Meloidogyne incognita*. *Int. J. Parasitol.* **30**, 77–81.
26. Hawdon, J. M., Jones, B. F. & Hotez, P. J. (1995). Cloning and characterization of a cDNA encoding the catalytic subunit of a cAMP-dependent protein kinase from *Ancylostoma caninum* third-stage infective larvae. *Mol. Biochem. Parasitol.* **69**, 127–130.
27. Hawdon, J. M. & Hotez, P. J. (1996). Hookworm: developmental biology of the infectious process. *Curr. Opin. Genet. Dev.* **6**, 618–623.
28. Hawdon, J. M., Jones, B. F., Hoffman, D. R. & Hotez, P. J. (1996). Cloning and characterization of Ancylostoma-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. *J. Biol. Chem.* **271**, 6672–6678.
29. Zhan, B., Hawdon, J., Shan, Q., Ren, H., Qiang, H., Xiao, S. H. *et al*. (2000). Construction and analysis of cDNA library of *Necator americanus* third stage larvae. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi*, **18**, 26–28.
30. Goud, G. N., Zhan, B., Ghosh, K., Loukas, A., Hawdon, J., Dobardzic, A. *et al*. (2004). Cloning, yeast expression, isolation, and vaccine testing of recombinant Ancylostoma-secreted protein (ASP)-1 and ASP-2 from *Ancylostoma ceylanicum*. *J. Infect. Dis.* **189**, 919–929.
31. MacDonald, A. J., Tawe, W., Leon, O., Cao, L., Liu, J., Oksov, Y. *et al*. (2004). Ov-ASP-1, the *Onchocerca volvulus* homologue of the activation associated secreted protein family is immunostimulatory and can induce protective anti-larval immunity. *Parasite Immunol.* **26**, 53–62.
32. Bethony, J. M., Loukas, A., Smout, M. J., Mendez, S., Wang, Y., Bottazzi, M. E. *et al*. (2004). Antibodies against a secreted protein from hookworm larvae reduce the intensity of infection in humans and vaccinated laboratory animals. *Nature Med*. In the press.
33. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244.
34. Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O. *et al*. (2004). Breaking good resolutions with ARP/wARP. *J. Synchrotron Radiat.* **11**, 56–59.
35. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallog. sect. D*, **58**, 968–975.
36. Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). ARP/wARP and molecular replacement. *Acta Crystallog. sect. D*, **57**, 1445–1450.
37. Lamzin, V. S. & Perrakis, A. (2000). Current state of automated crystallographic data analysis. *Nature Struct. Biol.* **7**, 978–981.
38. Lamzin, V. S., Perrakis, A., Bricogne, G., Jiang, J., Swaminathan, S. & Sussman, J. L. (2000). Apotheosis, not apocalypse: methods in protein crystallography. *Acta Crystallog. sect. D*, **56**, 1510–1511.
39. Perrakis, A., Antoniadou-Vyza, E., Tsitsa, P., Lamzin, V. S., Wilson, K. S. & Hamodrakas, S. J. (1999). Molecular, crystal and solution structure of a beta-cyclodextrin complex with the bromide salt of 2-(3-dimethylaminopropyl)tricyclo[3.3.1.1(3,7)]decan-2-ol, a potent antimicrobial drug. *Carbohydr. Res.* **317**, 19–28.
40. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
41. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, G. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
42. Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read,

R. J. (1998). Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallog. sect. D*, **54**, 1285–1294.
43. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallog. sect. D*, **55**, 247–255.
44. Murshudov, G. N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
45. Henriksen, A., King, T. P., Mirza, O., Monsalve, R. I., Meno, K., Ipsen, H. *et al*. (2001). Major venom allergen of yellow jackets, Ves v 5: structural characterization of a pathogenesis-related protein superfamily. *Proteins: Struct. Funct. Genet.* **45**, 438–448.
46. Fernandez, C., Szyperski, T., Bruyere, T., Ramage, P., Mosinger, E. & Wuthrich, K. (1997). NMR solution structure of the pathogenesis-related protein P14a. *J. Mol. Biol.* **266**, 576–593.
47. Higazi, T. B., Pearlman, E., Whikehart, D. R. & Unnasch, T. R. (2003). Angiogenic activity of an *Onchocerca volvulus* Ancylostoma secreted protein homologue. *Mol. Biochem. Parasitol.* **129**, 61–68.
48. Murray, J., Gregory, W. F., Gomez-Escobar, N., Atmadja, A. K. & Maizels, R. M. (2001). Expression and immune recognition of *Brugia malayi* VAL-1, a homologue of vespid venom allergens and Ancylostoma secreted proteins. *Mol. Biochem. Parasitol.* **118**, 89–96.
49. Asojo, O. A., Boulegue, C., Hoover, D. M., Lu, W. & Lubkowski, J. (2003). Structures of thymus and activation-regulated chemokine (TARC). *Acta Crystallog. sect. D*, **59**, 1165–1173.
50. Olatunde, B. O. & Onyemelukwe, G. C. (1994). Immunosuppression in Nigerians with hookworm infection. *Afr. J. Med. Med. Sci.* **23**, 221–225.
51. Loukas, A. & Prociv, P. (2001). Immune responses in hookworm infections. *Clin. Microbiol. Rev.* **14**, 689–703.
52. Alcami, A., Ghazal, P. & Yewdell, J. W. (2002). Viruses in control of the immune system. Workshop on molecular mechanisms of immune modulation: lessons from viruses. *EMBO Rep.* **3**, 927–932.
53. Murphy, P. M. (2001). Viral exploitation and subversion of the immune system through chemokine mimicry. *Nature Immunol.* **2**, 116–122.
54. Ehlers, M. R. (2000). CR3: a general purpose adhesion–recognition receptor essential for innate immunity. *Microb. Infect.* **2**, 289–294.
55. Rieu, P., Sugimori, T., Griffith, D. L. & Arnaout, M. A. (1996). Solvent-accessible residues on the metal ion-dependent adhesion site face of integrin CR3 mediate its binding to the neutrophil inhibitory factor. *J. Biol. Chem.* **271**, 15858–15861.
56. Haas, T. A. & Plow, E. F. (1994). Integrin–ligand interactions: a year in review. *Curr. Opin. Cell Biol.* **6**, 656–662.
57. Zhang, L. & Plow, E. F. (1997). Identification and reconstruction of the binding site within alphaMbeta2 for a specific and high affinity ligand, NIF. *J. Biol. Chem.* **272**, 17558–17564.
58. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallog. sect. D*, **55**, 1718–1725.
59. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
60. Brunger, A. T. (1992). The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
61. Kopp, J. & Schwede, T. (2004). The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. *Nucl. Acids Res.* **32**, D230–D234.
62. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucl. Acids Res.* **31**, 3381–3385.
63. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
64. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.
65. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
66. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Edited by I. Wilson

(Received 23 October 2004; received in revised form 8 December 2004; accepted 9 December 2004)