

# Domains of invasion organelle proteins from apicomplexan parasites are homologous with the Apple domains of blood coagulation factor XI and plasma pre-kallikrein and are members of the PAN module superfamily

Philip J. Brown<sup>a</sup>, Andrew C. Gill<sup>a</sup>, Philip G. Nugent<sup>b,1</sup>, John H. McVey<sup>b</sup>,  
Fiona M. Tomley<sup>a,\*</sup>

<sup>a</sup>*Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire RG20 7NN, UK*

<sup>b</sup>*Haemostasis Group, MRC Clinical Sciences Centre, ICSM, London W12 0NN, UK*

Received 4 April 2001; accepted 13 April 2001

First published online 3 May 2001

Edited by Thomas L. James

**Abstract** Micronemes are specialised organelles, found in all apicomplexan parasites, which secrete molecules that are essential for parasite attachment to and invasion of host cells. Regions of several microneme proteins have sequence similarity to the Apple domains (A-domains) of blood coagulation factor XI (FXI) and plasma pre-kallikrein (PK). We have used mass spectrometry on a recombinant-expressed, putative A-domain from the microneme protein EtMIC5 from *Eimeria tenella*, to demonstrate that three intramolecular disulphide bridges are formed. These bridges are analogous to those that stabilise A-domains in FXI and PK. The data confirm that the apicomplexan domains are structural homologues of A-domains and are therefore novel members of the PAN module superfamily, which also includes the N-terminal domains of members of the plasminogen/hepatocyte growth factor family. The role of A-domains/PAN modules in apicomplexan parasites is not known, but their presence in the microneme suggests that they may be important for mediating protein–protein or protein–carbohydrate interactions during parasite attachment and host cell invasion. © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

**Key words:** N-terminal domain of plasminogen module; Apple domain; Apicomplexa; Microneme; Mass spectrometry

## 1. Introduction

Apicomplexan protozoa, such as *Cryptosporidium*, *Eimeria*, *Plasmodium* and *Toxoplasma*, cause life-threatening diseases of man and livestock. They are obligate intracellular parasites that invade target host cells in order to replicate. Invasion is an active process, powered by parasite actinomyosin [1–7], and requiring apical attachment of parasites to the host cell membrane. Micronemes are secretory organelles, located in-

side the parasite apical tip, which secrete their contents when parasites make contact with host cells [8–10]. It has been proposed that micronemes store proteins that are essential for parasite attachment to host cells and that they release these in response to environmental stimuli [11–13]. In support of this hypothesis, genetic knock-out of the microneme protein PbTRAP of *Plasmodium berghei* results in parasites that cannot infect mosquito salivary glands and rat liver cells [14].

Recently, microneme proteins have been described with sequence identity to the Apple domains (A-domains) of the heavy chain of blood coagulation factor XI (FXI) and plasma pre-kallikrein (PK) [15–17]. FXI and PK have four A-domains, each of which is stabilised by three disulphide bonds [18,19] to form the ‘Apple’ structure. The A-domains of PK and FXI mediate binding interactions, some of which have been mapped using conformationally constrained peptides, in vitro mutagenesis, expression and characterisation of recombinant variants and monoclonal antibodies. Thus, in FXI, A1 has binding sites for high molecular weight kininogen [20] and for thrombin, a proteolytic activator of FXI [21]. A2 and A3 have binding sites for FIX, [22,23] and A4 has one for the activating protease, FXIIa [24]. Platelet-binding maps to A3 [25] as does heparin-binding, to a site that is overlapping, but distinct from, the platelet-binding region [26]. FXI circulates in plasma as a homodimer and dimerisation is mediated through the A4-domain and stabilised by an inter-chain disulphide bond. Recently it was suggested on the basis of homology search and structure prediction methods that the A-domains of FXI and PK belong to the PAN module superfamily [27] which also includes a number of *Caenorhabditis elegans* proteins and the N-terminal domains of the plasminogen/hepatocyte growth factor (HGF) family. Interestingly, the PAN module of HGF also binds heparin [28].

Although the percent sequence similarity between PAN family members is low, the homology is supported by a characteristic pattern of four cysteines and clusters of hydrophobic residues. The A-domains of FXI and PK, and of the majority of the nematode proteins share this pattern but also contain an extra pair of conserved cysteines at the amino- and carboxy-terminal boundaries of the domain. The six cysteines that form the disulphide bonds of the A-domains in PK and FXI are also conserved in the A-domains of the apicomplexan microneme proteins, although the overall level of sequence identity between the apicomplexan sequences and those of FXI and PK is low. For example, the mammalian structures

\*Corresponding author. Fax: (44)-1635-577 263.  
E-mail: fiona.tomley@bbsrc.ac.uk

<sup>1</sup> Present address: Wellcome Trust Laboratories for Molecular Parasitology, Department of Biochemistry, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK.

**Abbreviations:** A-domain, Apple domain; PAN module, N-terminal domain of plasminogen module; FXI, coagulation factor XI; PK, plasma pre-kallikrein

have an additional string of seven or eight residues between the fourth and fifth cysteines (Fig. 2, [17]) which, from the determined structure of HGF, would be predicted to form an extended surface loop connecting two anti-parallel  $\beta$ -sheets.

Since there is mounting evidence for the role of microneme proteins in binding to host cells during parasite invasion, sequence similarity and hence potential homology with PAN modules of proteins involved in protein–protein and/or protein–carbohydrate interactions is of great interest. In this paper we show that the six conserved cysteines of an A-domain from the microneme protein EtMIC5 of the parasite *Eimeria tenella*, form disulphide bonds in a pattern similar to that in PK and FXI, strongly suggesting that this domain is a true A-domain and thus a member of the PAN module superfamily.

## 2. Materials and methods

### 2.1. Sub-cloning of EtMIC5 A-domains into plasmid expression vector pFLAG-ATS

The coding sequences for six putative A-domains of EtMIC5 (A1, A3, A5, A7, A9 and A11) and for a fusion of three domains (A9–A11) were amplified from plasmid pPB7 [17] by polymerase chain reaction (PCR) with *Pfu* polymerase using specific primers to incorporate *Hind*III sites at the 5'-end and *Eco*RI sites (in A1, A3 and A9) or *Bgl*II sites (in A5, A7, A9 and A9–A11). Digested PCR products were cloned into *Hind*III and *Eco*RI/*Bgl*II digested pFLAG-ATS (IBI/Sigma), and the constructs checked by nucleotide sequencing. Recombinant plasmids were transformed into *Escherichia coli* strain BL21, and protein expression induced in log-phase cultures by exposure to 1 mmol/l isopropyl  $\beta$ -D-thiogalactopyranoside for 1 h at 37°C followed by 3 h at 30°C. Whole cell lysates were analysed by SDS-PAGE and Western blotting with a polyclonal hyperimmune rabbit serum against EtMIC5 [17] and anti-FLAG peptide monoclonal antibodies M1 and M2 (IBI/Sigma).

### 2.2. Purification of FLAGA9

Recombinant FLAGA9 was purified from a periplasmic shock fraction. Briefly, the cell pellet was washed in 10 mmol/l Tris–HCl, pH 8.0, and suspended at 30 ml/g in 0.5 mol/l sucrose, 30 mmol/l Tris–HCl, pH 8.0, 1 mmol/l EDTA at room temperature. The cells were centrifuged at 5000  $\times g$  for 10 min then immediately suspended in ice-cold distilled water and re-centrifuged at 8000  $\times g$  for 15 min. The supernatant (periplasmic fraction) was decanted, made 0.1% (v/v) in polyethyleneimine and DNA removed by centrifugation at 10000  $\times g$  for 10 min and filtration of the supernatant through a 0.2  $\mu$ m membrane. The filtrate was adjusted to 50 mmol/l Tris–HCl, pH 8.5 and subjected to two rounds of anion exchange chromatography according to manufacturer's instructions (SQ and RQ columns, Amersham Pharmacia Biotech). Fractions containing FLAGA9 were pooled, adjusted to 0.1% (v/v) trifluoroacetic acid (TFA) and applied to a reversed-phase column (PepRPCMTM HR 5/5 column, Amersham Pharmacia Biotech). Proteins were eluted with a linear gradient of 0–100% acetonitrile in 0.1% TFA and fractions containing FLAGA9 were pooled, buffer exchanged (PD-10 column, Amersham Pharmacia Biotech) into 10 mmol/l phosphate, pH 7.0 and concentrated (CP3, CC3, Millipore).

### 2.3. PAGE staining of proteins, Western blotting

Protein samples were solubilised by boiling in SDS–PAGE loading buffer with 100 mmol/l dithiothreitol (DTT) and resolved by electrophoresis through 15% PAGE gels with 0.1% SDS. Proteins were stained with Coomassie brilliant blue or silver (SilverXpress, Novex) or transferred to nitrocellulose membranes by semi-dry blotting. Membranes were blocked for 1 h in phosphate buffered saline, 5% (w/v) non-fat milk powder, 0.05% (w/v) Tween 20 and then proteins were detected by probing with specific antibodies [17].

### 2.4. On-line capillary high performance liquid chromatography (HPLC)-mass spectrometry (MS)

Home-made capillary columns of 180  $\mu$ m internal diameter were packed with Supelcosil C8 reversed-phase media (Sigma). An ABI 140C Microgradient System HPLC Pump (Applied Biosystems, Fos-

ter City, CA, USA) provided a constant flow rate of 50  $\mu$ l/min, which was split by a simple T-connector splitter to direct approximately 1.5% of the flow (750 nl/min) to the capillary column via a Valco low dispersion injector. For analysis of intact FLAGA9 protein and of tryptic peptides, samples were loaded with a Valco injector equipped with an internal loop of 1  $\mu$ l. For analysis of peptides produced by endoproteinase Glu-C digestion, samples were pre-concentrated and desalted with a Valco injector equipped with an external loop containing a home-made protein trap. In all experiments, samples were loaded in 0.1% TFA in water (solvent A) and eluted with a 15–60% gradient of 80:20 methanol:acetonitrile in 0.1% TFA (solvent B) over 40 min. The eluate was passed directly to the mass spectrometer (Quattro II tandem quadrupole, Micromass, Altrincham, UK) via a 25  $\mu$ m internal diameter fused silica transfer capillary. The mass spectrometer was equipped with a continuous flow nanospray source consisting of a Valco stainless steel union attached to the end of a commercial electrospray probe (Micromass, Altrincham, UK) with 20  $\mu$ m internal diameter fused silica as the spraying capillary. A potential of 2.25 kV was applied to the union and the instrument was scanned between  $m/z$  300 and  $m/z$  2000 with the cone voltage ramping between 40 and 100 V to allow improved detection of ions of high  $m/z$ .

### 2.5. Enzymatic digestion of FLAGA9

10  $\mu$ l of trypsin (Sigma) (50  $\mu$ g/ml in water) was added to 10  $\mu$ l of FLAGA9 protein in 10 mmol/l sodium phosphate, pH 7.0 and the solution incubated at 35°C for 24 h. 1  $\mu$ l was removed for HPLC-MS analysis and the remaining sample diluted to 200  $\mu$ l by the addition of 100 mmol/l ammonium acetate (pH 4). 5  $\mu$ l of endoproteinase Glu-C (Sigma) (100  $\mu$ g/ml in 100 mmol/l ammonium acetate) were added and the solution incubated at 35°C for a further 24 h. 20  $\mu$ l of this sample was subjected to HPLC-MS analysis.

### 2.6. Confirmation of the presence of disulphide bonds and of peptide identity

5  $\mu$ l (2  $\mu$ g) of recombinant FLAGA9 protein were diluted 10-fold in 100 mmol/l Tris, pH 9.0 and alkylated by incubation in 10 mmol/l iodoacetic acid at 37°C for 1 h, either with or without prior reduction by incubation in 10 mmol/l DTT at 50°C for 2 h. Accurate masses for these derivatised proteins were obtained by on-line HPLC-MS (see above). Tryptic and endoproteinase Glu-C peptides were isolated by reversed-phase HPLC (PepRPCMTM HR 5/5 column, Amersham Pharmacia Biotech), fractions were concentrated by drying, 1  $\mu$ l aliquots were loaded into pulled, platinum-coated borosilicate glass capillaries and peptides were analysed by single or tandem MS. For MS/MS experiments, argon at  $4 \times 10^{-3}$  mbar was used as the collision gas, collision energies of 30 V were used and at least 100 scans of 20 s were summed to produce each spectrum. Fractions containing putative disulphide-linked peptides were lyophilised, reconstituted in 100 mmol/l Tris, pH 9.0 and analysed by HPLC-MS both before and after reduction (as above).

## 3. Results and discussion

A-domains from the microneme protein EtMIC5 were sub-cloned, expressed as recombinant fusion proteins in *E. coli*, and analysed by Western blotting (Fig. 1a). All the recombinant A-domains were recognised by anti-FLAG monoclonal antibody M2, which binds the FLAG peptide in any context, and all but FLAGA5 and FLAGA7 were recognised by anti-FLAG monoclonal antibody M1, which binds specifically when the FLAG peptide is exposed at the N-terminus, i.e. after cleavage of the OmpA signal peptide during translocation to the periplasmic space. All but FLAGA5 and FLAGA7 were also recognised by a hyperimmune rabbit serum raised against parasite-derived EtMIC5 [17], indicating that they were probably, at least partially, correctly folded. FLAGA9 recombinant protein was selected for structural studies because it was highly expressed and successfully translocated to the periplasmic space. The FLAGA9 fusion protein contains the 21-residue OmpA signal, eight-residue FLAG tag, three residues from the plasmid vector, the 75 residues of

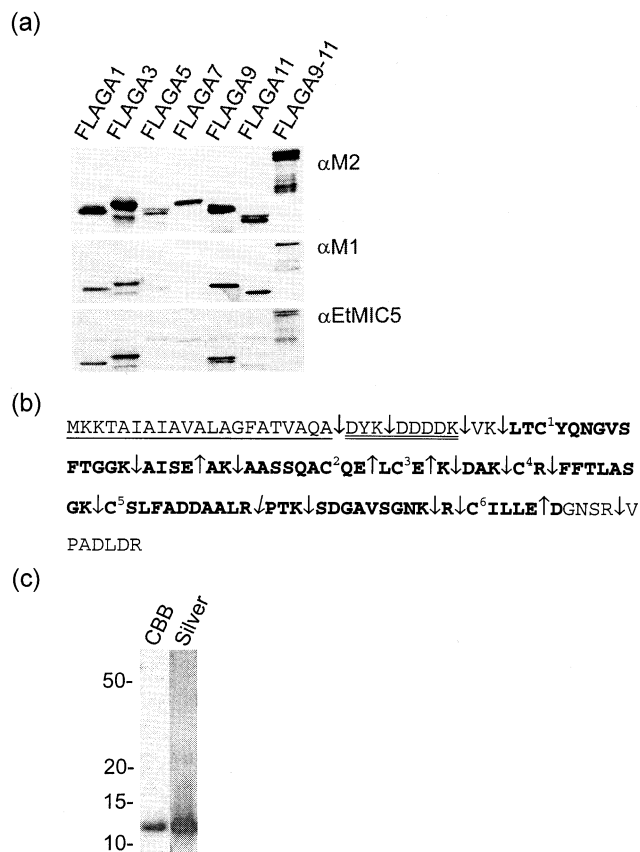
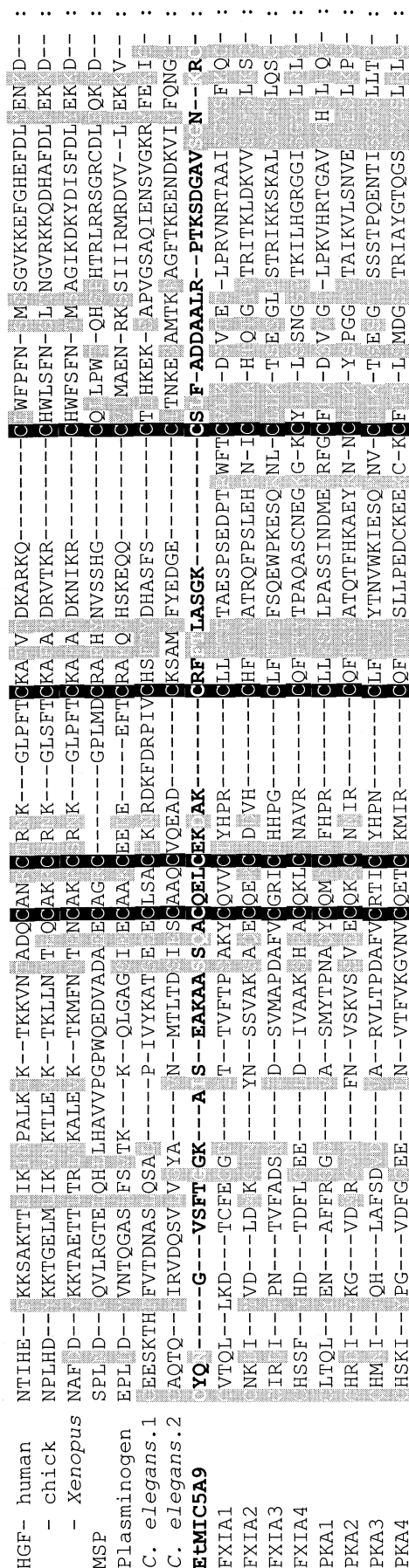


Fig. 1. Expression of EtMIC5 A-domains in *E. coli*. a: Western blot analysis of pFLAG-ATS recombinant A-domains. Whole cell samples were resolved by SDS-15% PAGE, blotted onto nitrocellulose and probed with antibodies M2, which recognises the FLAG peptide in any context, M1, which recognises the FLAG peptide only when at the N-terminus of the protein and rabbit hyperimmune anti-EtMIC5 serum. b: Primary sequence of the FLAGA9 peptide. Expression of A9 in the pFLAG-ATS vector leads to the production of a fusion protein which contains: an N-terminal OmpA signal peptide (single underlined) followed by the FLAG peptide (double underlined), residues VKL derived from the vector, the A9 peptide (bold) and finally residues GNSRVPADLDR derived from the vector. Arrows indicate OmpA signal cleavage (↓); trypsin (↓) and Glu-C cleavage sites (↑). Trypsin digestion is not efficient where arginine is followed by proline (✓). c: SDS-PAGE analysis. Coomassie brilliant blue and silver staining of pooled fractions corresponding to the first major peak eluted from the C18 column.

A9 and a further 11 residues from the vector (Fig. 1b). During translocation into the periplasmic space, the OmpA signal sequence is removed leaving a fusion of 97 residues. This was purified from the periplasmic shock fraction by two rounds of anion exchange chromatography followed by re-

Fig. 2. ClustalW alignment. Selected sequences are: HGF of human (JH0579, residues 37–123), chick (Q90978, residues 33–119) and *Xenopus* (I51283, residues 24–110); human MSP (P26927, residues 21–105); human plasminogen (P00747, residues 20–98); a theoretical protein of *C. elegans* (Q18777, residues 26–114 and 123–203); A-domain 9 of EtMIC5 (CAB52368, residues 713–781); human FXI (P03951, residues 20–103, 110–193, 200–283 and 291–374); and human PK (P03952, residues 21–104, 111–194, 201–284 and 292–375). The ClustalW.aln file was presented using Genedoc and identical/similar residues are indicated by shading; 100% black, 50% grey.



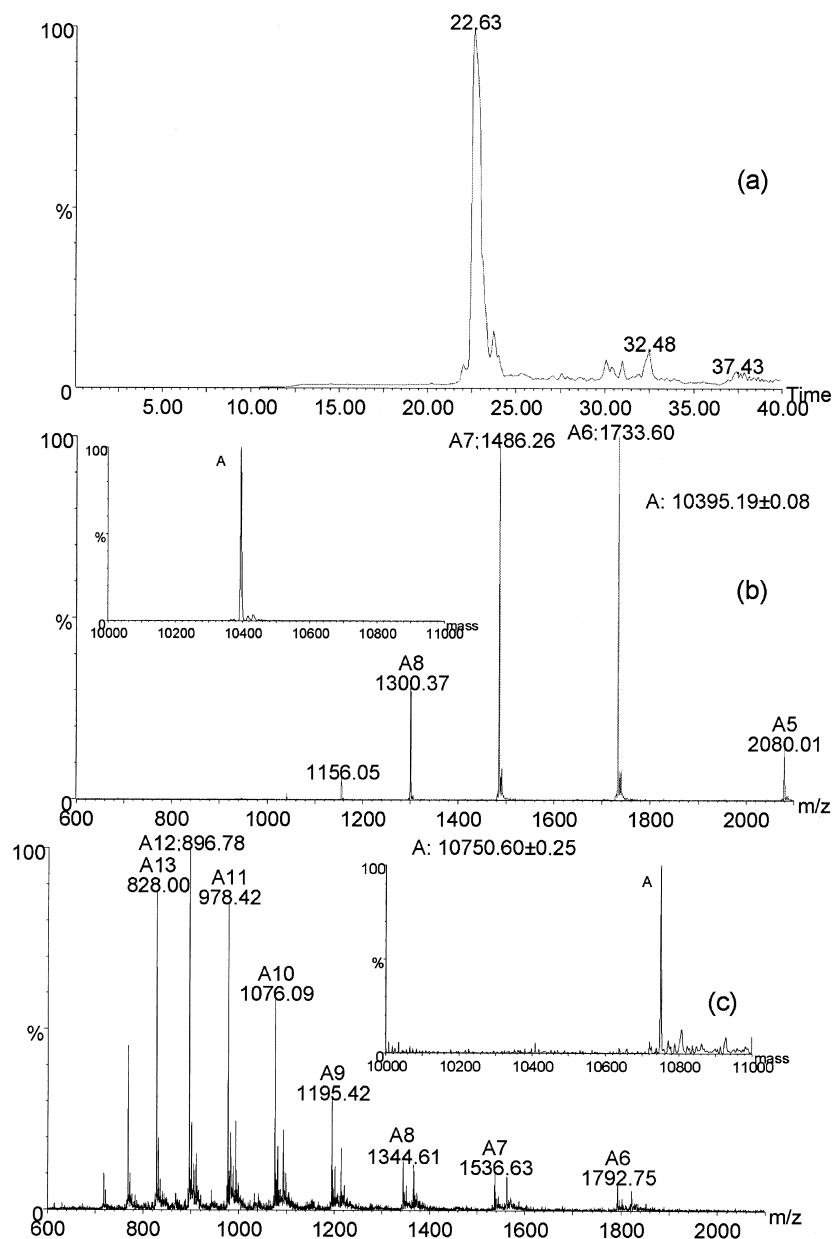


Fig. 3. On-line capillary RP-HPLC-MS analysis of purified FLAGA9. a: Mass spectrometric TIC chromatogram of purified protein showing a major peak at 22.63 min corresponding to FLAGA9. Minor peaks at higher retention times are interpreted as higher molecular weight contaminants. b: Mass spectrum and (inset) deconvoluted mass spectrum resulting from summing scans under the chromatographic peak at 22.63 min. The calculated, isotopically averaged mass of FLAGA9, assuming formation of all three disulphide bonds, is 10395.60 Da. Reduced cysteine residues cause an increase in the observed mass of 2 Da per disulphide bond. c: Mass spectrum and (inset) deconvoluted mass spectrum of FLAGA9 after reduction with DTT and alkylation with iodoacetic acid. Reduction and alkylation results in the formation of carboxymethyl-cysteine residues and increases the mass of the protein by ~59 Da per free thiol. The measured mass of the protein is 10750.60 indicating that six free thiols were generated by DTT reduction.

versed-phase (C18) chromatography. SDS-PAGE separation and staining with Coomassie brilliant blue or silver indicated that the protein was purified to homogeneity (Fig. 1c).

EtMIC5 has 11 domains which, from sequence alignment, appear to be homologous to the A-domains of FXI and PK [17] and to numerous *C. elegans* proteins [27]. The PAN module sequence of HGF, plasminogen and macrophage stimulating proteins (MSPs) contain only four conserved cysteines and lack C1 and C6 which are conserved at the amino- and carboxy-termini of all the PK, FXI and apicomplexan sequences (Fig. 2, [17,27]). In addition, FXI, PK and the apicomplexan

A-domains have a consistent number (five) of residues separating C3 and C4, an FT motif after C4 and a highly conserved SG motif towards the C-terminus, which further distinguishes them from the HGF, plasminogen and MSP sequences. However, structural predictions suggest that A-domains belong to the PAN module superfamily [27]. Predictions from the PHD server [29] also indicate that EtMIC5 A9 has a secondary structure that is broadly similar to the PAN module, with a single helix though with fewer  $\beta$ -sheets and more loop regions from C4 onwards. Thus, we would predict that the EtMIC5 A-domains, and those of microneme

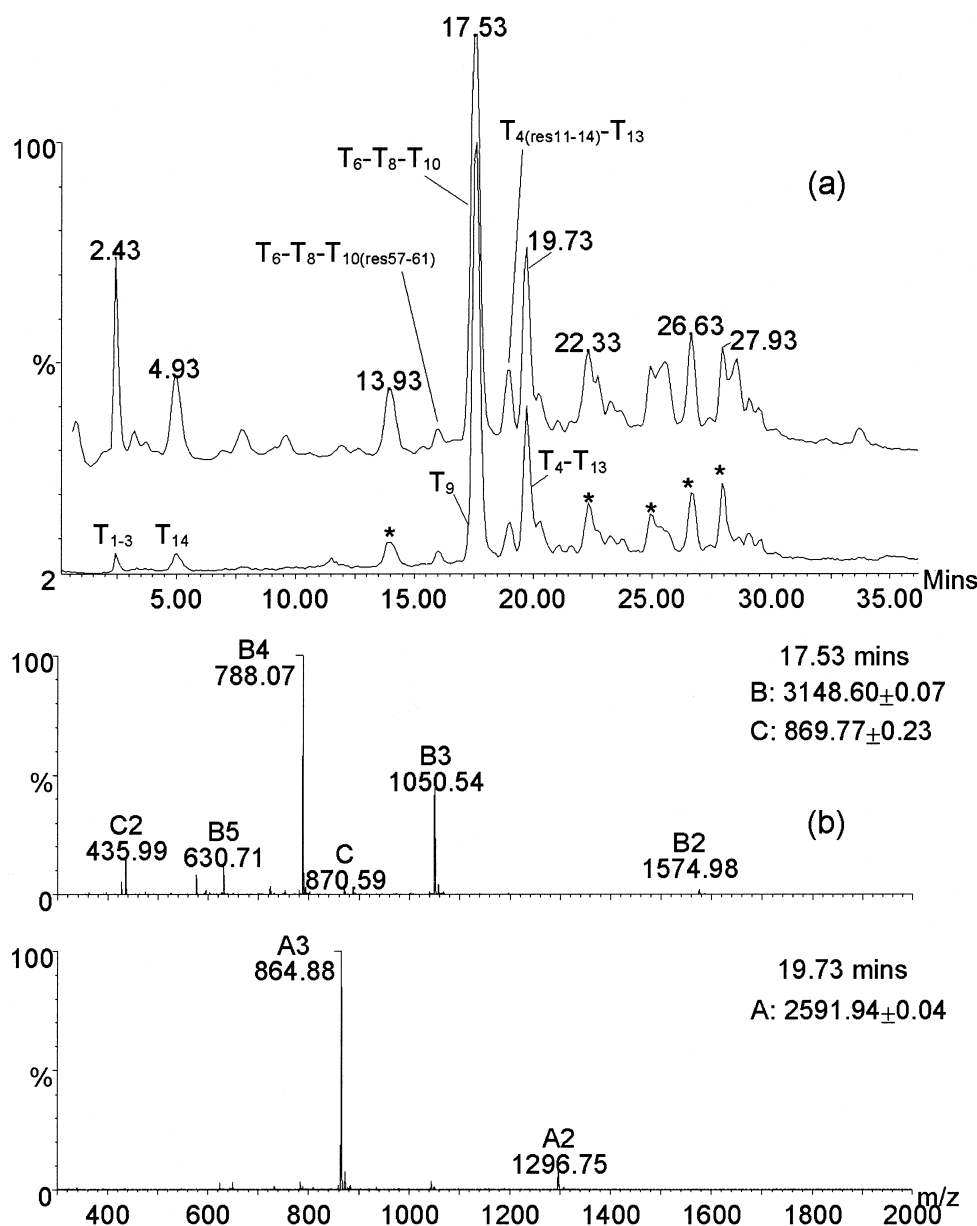


Fig. 4. On-line capillary RP-HPLC-MS analysis of peptides produced by tryptic digestion of FLAGA9. a: Chromatographic separation of tryptic peptides with detection by UV at 214 nm (upper trace) and mass spectrometric TIC chromatogram (lower trace). Peaks are labelled according to retention time and with the corresponding tryptic peptides. Trace levels of chymotrypsin result in additional cleavages indicated in parentheses. Peaks labelled with an asterisk are trypsin autolysis products. b: Mass spectra produced by summing scans under chromatographic peaks at retention times of 17.53 min (upper spectrum) and 19.73 min (lower spectrum). Peaks are labelled with the series label and the number of charges and the  $m/z$  ratio at which they occur. Series A corresponds to peptides T4 and T13 containing disulphide linking C1 and C6, calculated mass 2591.91 Da. Series B corresponds to peptides T6, T8 and T10 containing C2, C3, C4 and C5, calculated mass 3148.57. Series C corresponds to peptide T9 which co-elutes with series B, calculated mass 869.46.

proteins from two other apicomplexan parasites [30], have a fold with similarity to that described for the N-terminal domain of HGF [31,32].

In the A-domains of FXI and PK, disulphide bonds are formed between C1 and C6, C2 and C5, and C3 and C4. The domains are referred to as 'Apples' simply because of the constraints that the disulphide bonds place on the primary sequence when it is shown in two dimensions. To determine the number of intramolecular disulphide bridges in A9, an accurate mass for the FLAGA9 recombinant protein was measured. After purification, FLAGA9 was dialysed into a sodium phosphate buffer which precluded mass spectrometric

measurement of molecular mass by direct injection of the protein since sodium cations adduct to proteins when electrospray ionisation is utilised. The buffer was removed on-line by a reversed-phase HPLC column and analysis of FLAGA9 by MS produced the mass spectrometric total ion current (TIC) chromatogram in Fig. 3a. The peak at 22.63 min corresponds to FLAGA9. A small amount (estimated to be <1%) of higher molecular weight species were also seen, eluting from the column at around 30 min, and these were not identified. Fig. 3b shows the mass spectrum produced by summing the scans under the chromatographic peak at 22.63 min. The measured mass of FLAGA9 is  $10\,395.19 \pm 0.08$  Da (Fig. 3b), which is

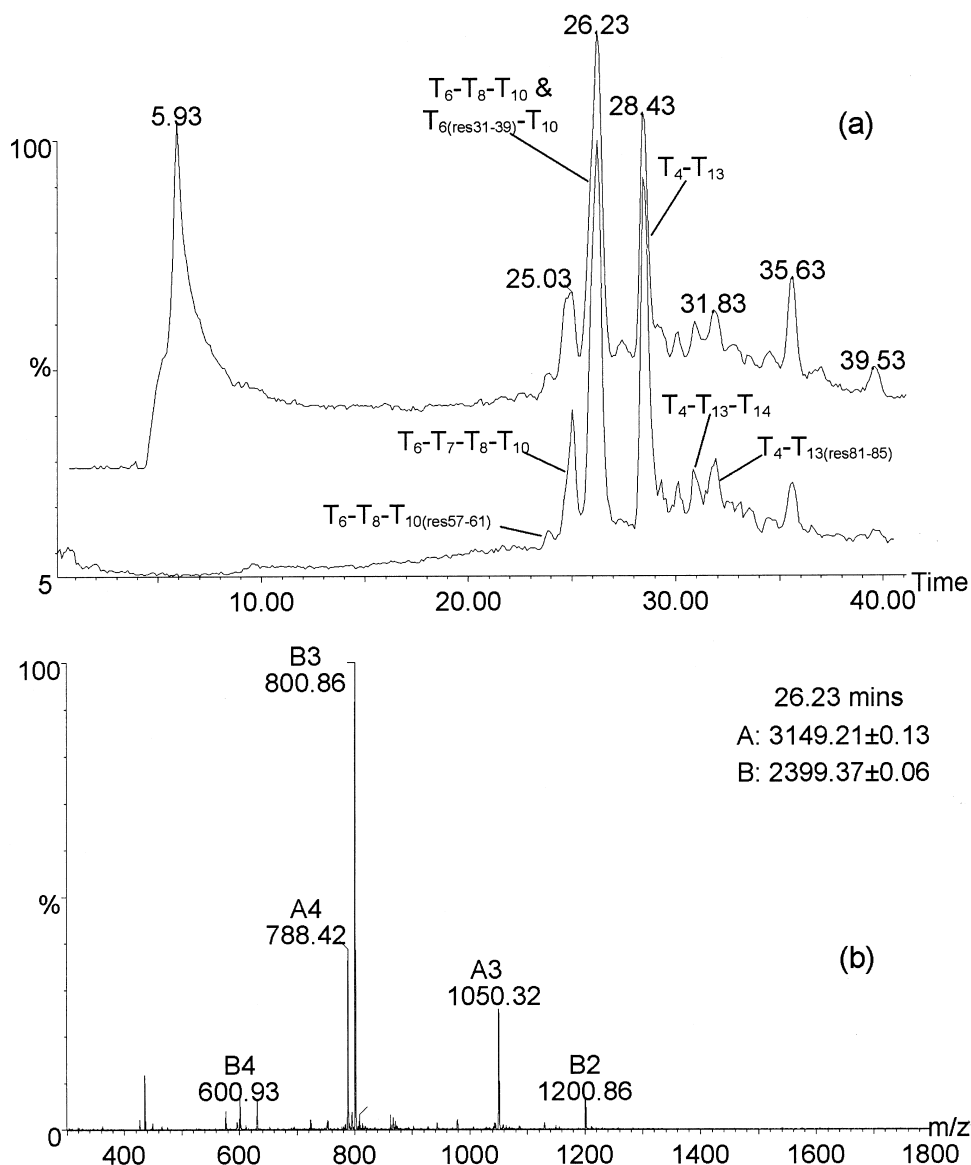


Fig. 5. On-line capillary RP-HPLC-MS analysis of peptides produced by endoproteinase Glu-C digestion of tryptic peptides of FLAGA9. a: Chromatographic separation of peptides with detection by UV at 214 nm (upper trace) and mass spectrometric TIC (lower trace). Peaks are labelled according to time and with tryptic peptide assignments. Where residues have been lost by chymotrypsin or endoproteinase Glu-C digestion this is indicated in parentheses. b: Mass spectra produced by summing scans under the chromatographic peak at retention time 26.23 min. Peaks are labelled with the series identifier and the number of charges along with the  $m/z$  ratio at which they occur. Series A corresponds to Glu-C undigested tryptic peptides T6, T8 and T10 containing C2, C3, C4 and C5, calculated mass 3148.57 Da. Series B corresponds to the same peptides with Glu-C enzymatic cleavage between residues 39 and 40 liberating residues 40–43, containing C3, and residues 47–48 containing C4. Peptide B therefore contains C2 and C5 linked by a disulphide bridge, calculated mass 2399.65.

within 0.4 Da of the predicted, isotopically averaged molecular mass of the fully oxidised peptide (10395.5 Da). Therefore, the measured mass confirms not only the correct cleavage of the OmpA signal sequence but also the presence of three disulphide bridges in the molecule. For each disulphide bond not formed the mass of the protein would increase by 2 Da. The inset in Fig. 3b shows the mass spectrum produced by maximum entropy deconvolution of the mass to charge spectrum of FLAGA9, which confirms the measured mass of the protein.

To provide additional evidence for the formation of three disulphide bonds, intact FLAGA9 was treated with the alkylating agent iodoacetic acid, either with or without reduction of disulphide bonds with DTT. When FLAGA9 was not re-

duced prior to alkylation, its mass remained unchanged (data not shown). After reduction of all disulphide bonds within the protein, alkylation raised the mass to 10750.60 Da (Fig. 3c). Reduction and derivatisation of free thiols with iodoacetic acids results in an increase in mass of  $\sim 59.2$  Da per newly generated free thiol group. In this case the mass change is  $\sim 355$  Da, indicating the presence of six free thiol groups after reduction of intramolecular disulphide bonds.

To probe the linkage positions of the disulphide bonds in FLAGA9, the sample was digested with trypsin and the resulting peptides were analysed by capillary HPLC-MS with additional UV detection. The chromatograms produced are shown in Fig. 4a. The mass spectra of peptides eluting at 17.63 and 19.73 min are shown in Fig. 4b. The peptide eluting

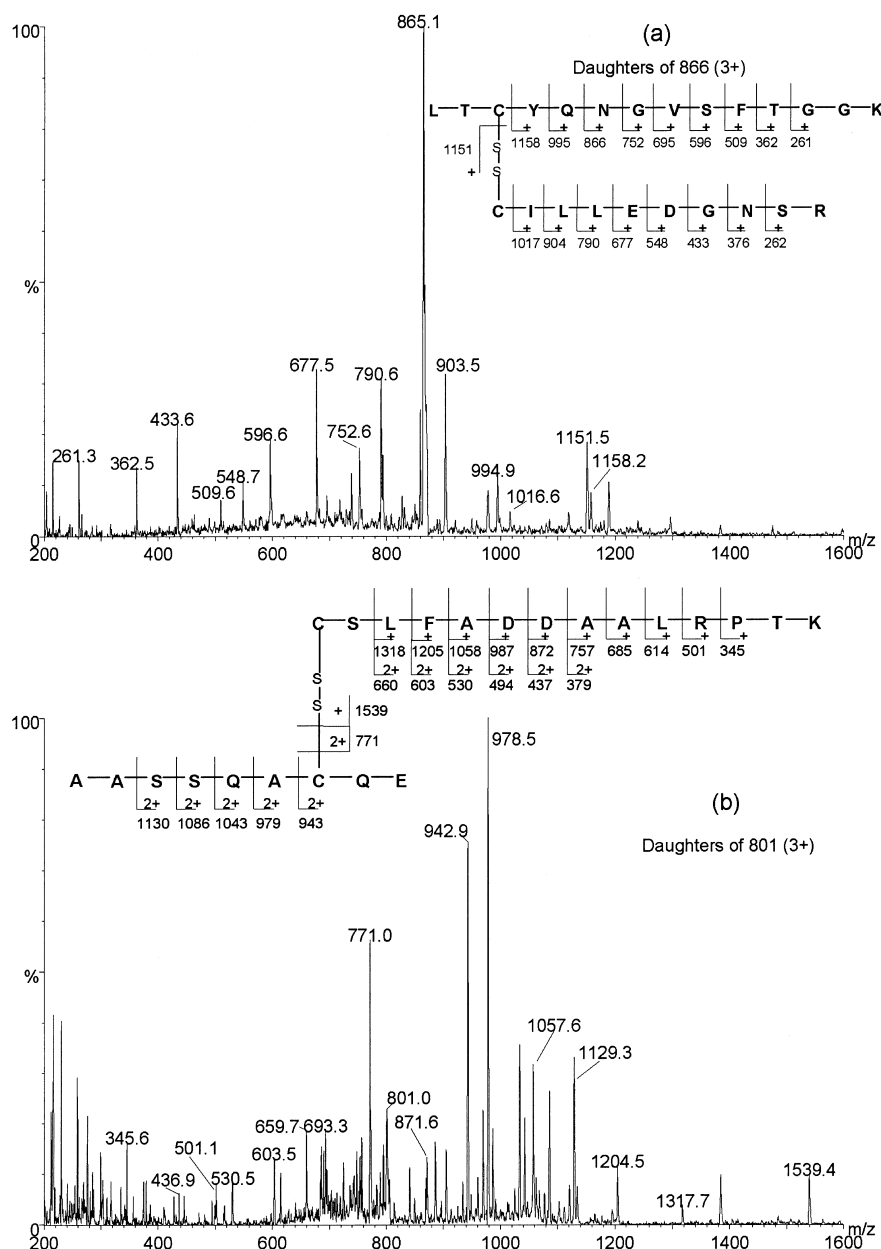


Fig. 6. MS/MS spectra of putative disulphide-linked peptides. Relevant ions are labelled with the mass to charge ratio of the peak and these signals are rationalised in diagrammatic form in the insets. a: Tandem mass spectrum of the tryptic peptide of mass 2592 Da. The triply charged ion ( $m/z$  865.7) was selected as the parent ion. The peptide covers residues 11–24, including C1 and residues 81–90 including C6. Sequence-specific daughter ions are evident from both peptide chains. b: Tandem mass spectrum of the 2399 Da peptide resulting from trypsin and endoproteinase Glu-C digestion. The triply charged ion ( $m/z$  800.6) was selected as the parent ion. The peptide covers residues 57–70 including C5 and residues 31–39 including C2. Sequence-specific daughter ions, some of which are doubly charged, are evident from both peptide chains.

at 19.73 min has a measured mass of  $2591.94 \pm 0.04$  Da. Based on the mass of this peptide, it was tentatively assigned as corresponding to residues 11–24, containing C1, linked to residues 81–90 containing C6. The predicted average mass for this linked peptide is 2591.94 Da and identification of this peptide therefore confirms disulphide linkage C1–C6. The peptide eluting at 17.63 min has a measured mass of  $3148.60 \pm 0.07$  Da and was tentatively assigned as corresponding to residues 31–43, containing C2 and C3, linked to 47–48, containing C4 and to 57–70 containing C5. The predicted average mass for this combination is 3148.57 Da. Also in the spectrum at 17.53 min are peaks resulting from a co-elut-

ing peptide of measured mass  $869.77 \pm 0.23$  Da. This peptide was tentatively assigned as corresponding to residues 49–56, which have a predicted mass of 869.46 Da. From these tentative assignments based on peptide masses it is possible to discount the possible linkages C2–C3 and C4–C5 since formation of these bridges would lead to two separate disulphide-containing peptides and not a single peptide of mass 3149 Da. It is not possible, however, to distinguish between the two other possible linkages, i.e. C2–C4 and C3–C5 versus C2–C5 and C3–C4.

The major species in the chromatographic analysis of the tryptic peptides of FLA9 have all been identified and are

labelled in Fig. 4a. Those peaks labelled with an asterisk correspond to trypsin autolysis products and low levels of peptides cleaved by trace amounts of chymotrypsin in the enzyme were also found and are labelled. Over 80% of the protein sequence has been detected and, crucially, no peptides were identified that correspond to other disulphide linkage combinations.

To characterise further the remaining disulphide bridges, the pH of the sample of tryptic peptides was lowered to pH 4.0 by the addition of approximately 10 volumes of 100 mmol/l ammonium acetate solution. The peptides were redigested with endoproteinase Glu-C which, under these conditions, cleaves specifically at glutamic acid residues. Peptides were separated by capillary HPLC and the UV and TIC chromatograms produced are shown in Fig. 5a. The mass spectrum of the peptides eluting at 26.33 min is shown in Fig. 5b. In addition to some undigested peptide of mass 3149 Da, there is a co-eluting peptide of mass  $2399.37 \pm 0.06$  Da. A peptide predicted from Glu-C digestion, corresponding to residues 31–39 containing C2 and residues 57–70 containing C5 has a predicted average mass of 2399.65 Da. These data confirm the presence of the C2–C5 disulphide linkage and since the measured mass of the intact protein, and our initial derivatisation experiments, showed that three disulphides were formed, the remaining linkage is inferred as C3–C4, although the peptide containing this linkage could not be detected. Again, the most abundant peptides were identified and no peptides were observed containing linkages other than those described above.

In order to confirm our tentative assignments, which were based solely on peptide mass, we isolated the putative single disulphide-containing peptides of masses 2592 Da (containing linkage C1–C6) and 2399 Da (containing linkage C2–C5) from the tryptic and endoproteinase Glu-C digests by reversed-phase HPLC. These peptides were analysed by collision-induced dissociation experiments and the MS/MS spectra produced are shown in Fig. 6. Although the fragmentation behaviour of disulphide-containing peptides can be difficult to interpret, in both cases the sequences of specific daughter ions were generated which confirmed the identity of both halves of the disulphide-linked peptides. Since sequence-specific ions from both peptides chains were produced by these experiments, it is complicated to use traditional nomenclature to label the peaks. For simplicity, peaks are labelled only with their mass-to-charge ratio and the rationalisation of these for the daughter ions is given in diagrammatic form.

To further confirm that peptides were correctly identified, the two single disulphide-containing peptides (2592 Da containing linkage C1–C6 and 2399 Da containing linkage C2–C5) and the tryptic peptide that contained two disulphide linkages (3149 Da containing linkage C2–C5 and the inferred linkage C3–C4) were reduced by addition of DTT and analysed by on-line HPLC-MS. In each case, reduction produced new peptides with masses corresponding to those expected from our previous assignments (data not shown).

In conclusion, we have expressed isolated A-domains of the microneme protein EtMIC5 from the apicomplexan parasite *E. tenella*, purified the A9-domain to homogeneity and analysed the disulphide bonds formed by MS. We have shown that in addition to the limited sequence similarity, this A-

domain shares an identical disulphide bond pattern (C1–C6; C2–C5; C3–C4) to the A-domains of FXI and PK, consistent with the hypothesis that the EtMIC5 repeat regions are A-domains and therefore members of the PAN module superfamily.

**Acknowledgements:** We would like to thank Laurence Hunt for advice on protein purification. P.J.B. was funded by a BBSRC studentship.

## References

- [1] Russell, D.G. and Sinden, R.E. (1981) *J. Cell Sci.* 50, 345–359.
- [2] Russell, D.G. (1983) *Parasitology* 87, 199–209.
- [3] Dobrowolski, J.M. and Sibley, L.D. (1996) *Cell* 84, 933–939.
- [4] Dobrowolski, J.M., Carruthers, V.B. and Sibley, L.D. (1997) *Mol. Microbiol.* 26, 163–173.
- [5] Forney, J.R., Vaughan, D.K., Yang, S. and Healey, M.C. (1998) *J. Parasitol.* 84, 908–913.
- [6] Sibley, L.D., Hakansson, S. and Carruthers, V.B. (1998) *Curr. Biol.* 8, R12–14.
- [7] Poupel, O. and Tardieux, I. (1999) *Microbes Infect.* 1, 653–662.
- [8] Entzeroth, R., Kerckhoff, H. and Konig, A. (1992) *Eur. J. Cell Biol.* 59, 405–413.
- [9] Tomley, F.M., Bumstead, J.M., Billington, K.J. and Dunn, P.P. (1996) *Mol. Biochem. Parasitol.* 79, 195–206.
- [10] Carruthers, V.B. and Sibley, L.D. (1997) *Eur. J. Cell Biol.* 73, 114–123.
- [11] Dubremetz, J.F., Garcia-Reguet, N., Conseil, V. and Fourmaux, M.N. (1998) *Int. J. Parasitol.* 28, 1007–1013.
- [12] Carruthers, V.B., Giddings, O.K. and Sibley, L.D. (1999) *Cell Microbiol.* 1, 225–235.
- [13] Bumstead, J.M. and Tomley, F.M. (2000) *Mol. Biochem. Parasitol.* 110, 311–321.
- [14] Sultan, A.A., Thathy, V., Frevert, U., Robson, K.J., Crisanti, A., Nussenzweig, V., Nussenzweig, R.S. and Menard, R. (1997) *Cell* 90, 511–522.
- [15] Eschenbacher, K.H., Klein, H., Sommer, I., Meyer, H.E., Entzeroth, R., Mehlhorn, H. and Ruger, W. (1993) *Mol. Biochem. Parasitol.* 62, 27–36.
- [16] Klein, H., Mehlhorn, H. and Ruger, W. (1996) *Parasitol. Res.* 82, 230–237.
- [17] Brown, P.J., Billington, K.J., Bumstead, J.M., Clark, J.C. and Tomley, F.M. (2000) *Mol. Biochem. Parasitol.* 107, 91–102.
- [18] McMullen, B.A., Fujikawa, K. and Davie, E.W. (1991) *Biochemistry* 30, 2050–2056.
- [19] McMullen, B.A., Fujikawa, K. and Davie, E.W. (1991) *Biochemistry* 30, 2056–2060.
- [20] Baglia, F.A., Jameson, B.A. and Walsh, P.N. (1990) *J. Biol. Chem.* 265, 4149–4154.
- [21] Baglia, F.A. and Walsh, P.N. (1996) *J. Biol. Chem.* 271, 3652–3658.
- [22] Baglia, F.A., Jameson, B.A. and Walsh, P.N. (1991) *J. Biol. Chem.* 266, 24190–24197.
- [23] Sun, Y. and Gailani, D. (1996) *J. Biol. Chem.* 271, 29023–29028.
- [24] Baglia, F.A., Jameson, B.A. and Walsh, P.N. (1993) *J. Biol. Chem.* 268, 3838–3844.
- [25] Baglia, F.A., Jameson, B.A. and Walsh, P.N. (1995) *J. Biol. Chem.* 270, 6734–6740.
- [26] Ho, D.H., Badellino, K., Baglia, F.A. and Walsh, P.N. (1998) *J. Biol. Chem.* 273, 16382–16390.
- [27] Tordai, H., Banyai, L. and Patthy, L. (1999) *FEBS Lett.* 461, 63–67.
- [28] Zhou, H. et al. (1999) *Biochemistry* 38, 14793–14802.
- [29] Rost, B. (1996) *Methods Enzymol.* 266, 525–539.
- [30] Tomley, F.M. and Soldati, D.S. (2001) *Trends in Parasitology* 17, 81–88.
- [31] Ultsch, M., Lokker, N.A., Godowski, P.J. and de Vos, A.M. (1998) *Structure* 6, 1383–1393.
- [32] Zhou, H., Mazzulla, M.J., Kaufman, J.D., Stahl, S.J., Wingfield, P.T., Rubin, J.S., Bottaro, D.P. and Byrd, R.A. (1998) *Structure* 6, 109–116.