Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

The structure of a haemopexin-fold protein from cow pea (*Vigna unguiculata*) suggests functional diversity of haemopexins in plants

Vineet Gaur,[a] Veenu Chanana,[a] Abha Jain[a] and Dinakar M. Salunke[a,b]*

[a]National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India, and [b]Regional Centre for Biotechnology, 180 Udyog Vihar Phase I, Gurgaon 122 016, India

Correspondence e-mail: dmsalunke@rcb.res.in

Received 8 October 2010
Accepted 6 December 2010

PDB Reference: CP4, 3oyo.

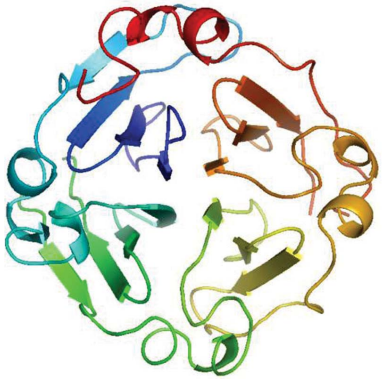
The haemopexin fold is present in almost all life forms and is utilized for carrying out diverse physiological functions. The structure of CP4, a haemopexin-fold protein from cow pea (*Vigna unguiculata*), was determined at 2.1 Å resolution. The protein exists as a monomer both in solution and in the crystal. The structure revealed a typical four-bladed β-propeller topology. The protein exhibits 42% sequence similarity to LS-24 from *Lathyrus sativus*, with substantial differences in the surface-charge distribution and in the oligomeric state. A structure-based sequence analysis of haemopexin-fold proteins of plant and mammalian origin established a sequence signature associated with the haemopexin motif. This signature sequence enabled the identification of other proteins with possible haemopexin-like topology of both plant and animal origin. Although CP4 shares a structural fold with LS-24 and other haemopexins, biochemical studies indicated possible functional differences between CP4 and LS-24. While both of these proteins exhibit spermine-binding potential, CP4 does not bind to haem, unlike LS-24.

1. Introduction

The β-propeller superfamily of proteins is known to be associated with structural rigidity. However, it exhibits immense phylogenetic diversity (Fülöp & Jones, 1999). The basic architecture of the fold is based on a simple building block consisting of a four-stranded anti-parallel β-sheet, which is repeated 4–8 times and radially arranged around a central axis (Fülöp & Jones, 1999). The four-bladed members of the β-propeller family constitute the haemopexin superfamily of proteins. To date, more than 500 proteins from organisms as diverse as viruses, prokaryotes and eukaryotes have been designated as containing the haemopexin-like motif (Piccard *et al.*, 2007). Among prokaryotes, two putative proteins, photopexin A and B from *Photorhabdus luminescens*, have been identified as containing haemopexin-like repeats (Crennell *et al.*, 2000). Among eukaryotes, various proteins, including limunectin from *Limulus* amoebocytes (Liu *et al.*, 1991), albumin proteins from plant seeds (Jenne, 1991), nectinepsin from the neuroretina, liver, brain and intestine of quail (*Coturnix coturnix*; Blancher *et al.*, 1996) and numerous vitronectins and matrix metalloproteinases (MMPs) from mammals, have been identified as containing haemopexin-like domains (Piccard *et al.*, 2007).

Proteins containing the haemopexin domain exhibit tremendous functional diversity. This functional diversity also corroborates with a correspondingly high sequence diversity. Interestingly, based on the limited structural data available it is evident that these proteins adapt similar tertiary structures. This was particularly observed in the mammalian system. The haemopexin fold has been employed for the execution of a variety of functions, as demonstrated by mammalian serum haemopexin, MMPs and vitronectin (Piccard *et al.*, 2007).

In plants, the haemopexin fold was first reported in the protein pea albumin 2 (PA2) from *Pisum sativum* (Jenne, 1991), which is a cytosolic protein (Harris & Croy, 1985) and a major albumin from pea seeds (Croy *et al.*, 1984; Harris & Croy, 1985; Higgins *et al.*, 1987; Gruen *et al.*, 1987). Subsequently, similar proteins have been reported from various other sources, including *Cicer arietinum* (Kolberg *et al.*, 1983; Vioque *et al.*, 1998), *Lathyrus sativus* (Qureshi *et al.*, 2006; Gaur



et al., 2010) and *Vigna unguiculata* (Chanana *et al.*, 2004). The involvement of plant haemopexins in spermine biosynthesis (Vigeolas *et al.*, 2008; Gaur *et al.*, 2010) and regulation of oxidative stress has been proposed (Gaur *et al.*, 2010). The study presented in this manuscript focuses on the structural diversity of the haemopexin fold in plants in analogy to that in the mammalian system. In the present study, we have determined the structure of CP4, a protein with haemopexin-like topology from *V. unguiculata*. While CP4 exhibits the haemopexin fold, similar to the previously reported structure of LS-24, significant differences could be observed in the surface-charge distribution and in the oligomeric structure. These may imply possible differences in their biochemical characteristics, indicating the involvement of the haemopexin fold in carrying out diverse physiological functions. Thus, this study highlights the extent of the diversity of the haemopexin fold in plants.

2. Materials and methods

2.1. Protein purification

Mature *V. unguiculata* seeds were purchased from the Indian Agricultural Research Institute (IARI), India. The seeds were ground to a fine powder, defatted with petroleum ether and subsequently homogenized with 50 m*M* Tris–HCl pH 7.5 containing 140 m*M* NaCl by continuous stirring for 4 h at 277 K. The crude extract was prepared by centrifugation at 48 384*g* for 30 min and was then subjected to ammonium sulfate fractionation. CP4 was purified from the 90% ammonium sulfate fraction using a PI/M weak anion-exchange column (Applied Biosystems) pre-equilibrated with 50 m*M* Tris–HCl pH 7.5. Elution was carried out using a gradient of 0–1 *M* NaCl in 50 m*M* Tris–HCl buffer pH 7.5 for 35 min.

2.2. Protein sequencing

Protein sequencing was carried out by subjecting the purified protein to limited proteolyses using endoproteinase Asn-N, V8 protease, trypsin and pepsin in order to obtain internal protein fragments. The internal proteolytic fragments thus obtained were transferred onto PVDF membrane and subjected to N-terminal sequencing using Edman chemistry on a Procise protein sequencer (Applied Biosystems). All sequence searches were performed with *BLAST* (Altschul *et al.*, 1997) and sequence alignments were carried out using *ClustalW* (Thompson *et al.*, 1994). The sequence of mung been seed albumin from *V. radiata* was used as a guide to align the sequences of the proteolytic fragments. The portions of the protein that could not be sequenced biochemically were identified crystallographically. In addition to electron density, features such as chemical environment and comparison of the electron densities of the two protomers in the asymmetric unit were taken into account when identifying the residues during structure refinement.

2.3. Structure solution and refinement

Data collection was carried out at 120 K on a Rigaku RU-H3R rotating-anode X-ray generator equipped with Osmic focusing mirrors and a MAR345dtb detector (MAR Research, Germany) using 33%(*v*/*v*) glycerol in the mother liquor as a cryoprotectant. The *automar* program was used for data processing. The crystal structure was determined by molecular replacement using *AMoRe* (Navaza, 2001) in the resolution range 50–4.0 Å. LS-24 (PDB entry 3lp9; Gaur *et al.*, 2010) was used as the phasing model. *CNS* (Brünger *et al.*, 1998) was used for the calculation of NCS (noncrystallographic symmetry) rotation and translation matrices and structure refinement. A randomly assigned 10% of the reflections were used for the calculation of $R_{\rm free}$. TLS refinement was carried out using *REFMAC*5 (Murshudov *et al.*, 1999; Winn *et al.*, 2001). The quality of the structure was assessed using *PROCHECK* (Laskowski *et al.*, 1993). Structure analysis was carried out using *PyMOL*. Most of the superpositions were carried out using secondary-structure superposition and the least-squares method of superposition in *Coot* (Emsley & Cowtan, 2004) and *UCSF Chimera* (Pettersen *et al.*, 2004).

2.4. Biochemical analyses

The molecular weight of CP4 was estimated using gel-exclusion chromatography. An S-300 TSK-GEL SW(XL) gel-filtration column

Figure 1
The complete sequence of CP4 obtained as a result of N-terminal sequencing of fragments obtained by enzymatic proteolysis and by interpretation of the electron-density map (shown in grey).

(Tosoh Corporation) pre-equilibrated with 50 m*M* Tris–HCl pH 7.5 at a flow rate of 0.75 ml min^{-1} was used. A low-molecular-weight gel-filtration calibration kit (Amersham Biosciences) was used to plot the standard curve: molecular weight (M_r) against K_{av} ($V_e - V_o/V_c - V_o$). V_e is the elution volume, V_o is the void volume of the column (10.8 ml) and V_c is the total bed volume of the column (26 ml). V_o was determined using blue dextran.

Haem-binding studies were carried out using a mobility-shift assay and Biacore studies as described previously (Gaur *et al.*, 2010). In the mobility-shift assay, LS-24 pre-incubated with increasing haem concentrations was used as a positive control. A dot-blot assay was carried out to demonstrate the presence of bound spermine in CP4 purified directly from *V. unguiculata* seeds as described previously (Gaur *et al.*, 2010). Polyclonal anti-spermine antibody (Abcam) and horseradish peroxidase (HRP) conjugated goat anti-rabbit IgG antibodies (Abcam) were used as the primary (1:1200 dilution) and secondary (1:3000 dilution) antibodies, respectively.

3. Results and discussion

3.1. Protein sequencing and primary-structure analysis

Sequence analysis is one of the most useful bioinformatics tools that is employed to establish relationships between proteins and their folds. However, it has failed to identify many relationships that have subsequently emerged after the three-dimensional structures of the proteins have been determined. Since structure is better conserved than sequence, structural data facilitate the recognition of relationships that are otherwise hidden at the sequence level. Combined analyses of structure and sequence could enable the identification of a signature motif associated with a structural fold that could identify protein relationships even at the sequence level (Landschulz *et al.*, 1988). Haemopexin is one such family of proteins, classified under the β-propeller fold, in which the members are found in diverse phylogenic classes and share very low sequence identity. We have analysed the structures and sequences of haemopexin-fold proteins from plant seeds.

While screening the proteome of *V. unguiculata* for potential allergenic proteins, CP4 was identified from the 90% ammonium sulfate fraction of the whole seed protein extract and was found to show significant sequence identity to mung bean seed albumin (MBSA) from *V. radiata* and the major albumin from pea (PA2; Chanana *et al.*, 2004). The protein sequence was determined by proteolytic fragmentation and interpretation of the electron-density map during iterative structural refinement. 70% of the sequence was elucidated by proteolytic fragmentation followed by N-terminal sequencing. The remainder of the sequence was determined by interpretation of the electron-density map (Fig. 1). CP4 showed 42% sequence identity to LS-24. Most of the residues that are conserved between the two proteins are restricted to the β-sheets of the fold. The loops interconnecting the β-strands and the interblade linker regions are highly variable in sequence. The extent of sequence

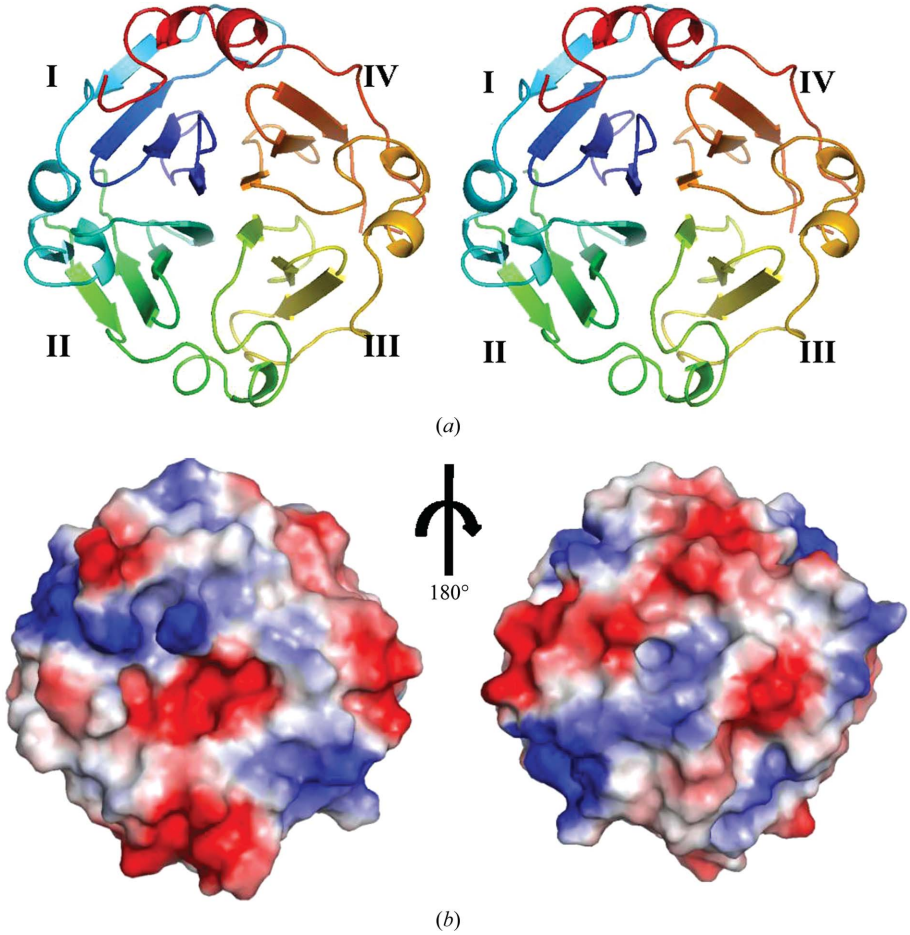


Figure 2
Crystal structure of the CP4 protein from *V. unguiculata*. (*a*) Stereoview of a ribbon representation of one CP4 molecule in the asymmetric unit, presenting a four-bladed β-propeller structure. (*b*) Surface-charge distribution of CP4 shown in two different views.

homology with LS-24 and the presence of an internal fourfold repeat qualifies CP4 as a member of the haemopexin family of proteins.

3.2. Overall structure

The CP4 protein crystallized in space group *C*2, with unit-cell parameters $a = 124.9$, $b = 60.1$, $c = 67.5$ Å, $\beta = 111.1^\circ$. There are two protomers in the asymmetric unit. The crystallization has been reported previously and the same data set was used in structure determination and refinement (Chanana *et al.*, 2004). The structure was determined at 2.1 Å resolution by molecular replacement using LS-24 as the phasing model. The refinement statistics are given in Table 1.

The asymmetric unit contains two polypeptide chains, each with three bound ions: calcium, sodium and chloride. The overall structure of CP4 is shown in Fig. 2. The molecular structure shows a β-propeller domain characteristic of proteins belonging to the haemopexin superfamily. Each molecule exhibits a pseudo-fourfold axis of symmetry passing through the centre of each of the monomers, which is characterized by a channel (Fig. 2*a*). The central channel contains the three bound ions: calcium, sodium and chloride. The presence of ions within the central channel appears to be a characteristic feature of the haemopexin domain; similar ions have also been reported in other structures with haemopexin-like topology.

Each protomer contains four β-sheets, each of which consists of four β-strands arranged around a pseudo-fourfold axis. The discoid subunit, which has an average diameter of 40 Å and a thickness of 30 Å, is characterized by a uniform charge distribution on the surface (Fig. 2*b*), in contrast to that of LS-24, which shows a distribution of positive and negative potential between the two faces of the discoid protein (Gaur *et al.*, 2010). The two monomers in the asymmetric unit superimpose with an r.m.s.d. of 0.17 Å. The molecules of LS-24 and CP4 superimpose with an r.m.s.d. of 0.75 Å. Striking differences are observed in the conformations of the loops connecting the outermost strands of the blades in CP4 and LS-24 (Fig. 3).

Table 1
Refinement statistics.

Resolution range (Å)	25.0–2.1
No. of molecules in asymmetric unit	2
R_{work}† (%)	20.9
R_{free}‡ (%)	24.6
R.m.s.d. bond lengths (Å)	0.02
R.m.s.d. bond angles (°)	1.6
No. of reflections	22725
Ramachandran parameters	
Residues in most favoured region (%)	84.2
Residues in allowed region (%)	14.2
Residues in generously allowed region (%)	1.6
Residues in disallowed region (%)	0.0
No. of non-H atoms used in refinement	
Protein atoms	3504
Heterogen atoms	6
Solvent atoms	182
Average *B* factors (Å^2)	
Protein atoms	15.91
Heterogen atoms	19.35
Solvent atoms	21.13

† $R_{work} = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$ for the 89.8% of the reflection data that were used in refinement. ‡ $R_{free} = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$ for the remaining 10.2% of the reflection data that were not used in refinement.

3.3. Signature sequence

The signature sequence motif corresponding to a single blade of the haemopexin fold was identified by superimposition of each of the blades of the haemopexin domains of proteins in the PDB (http://

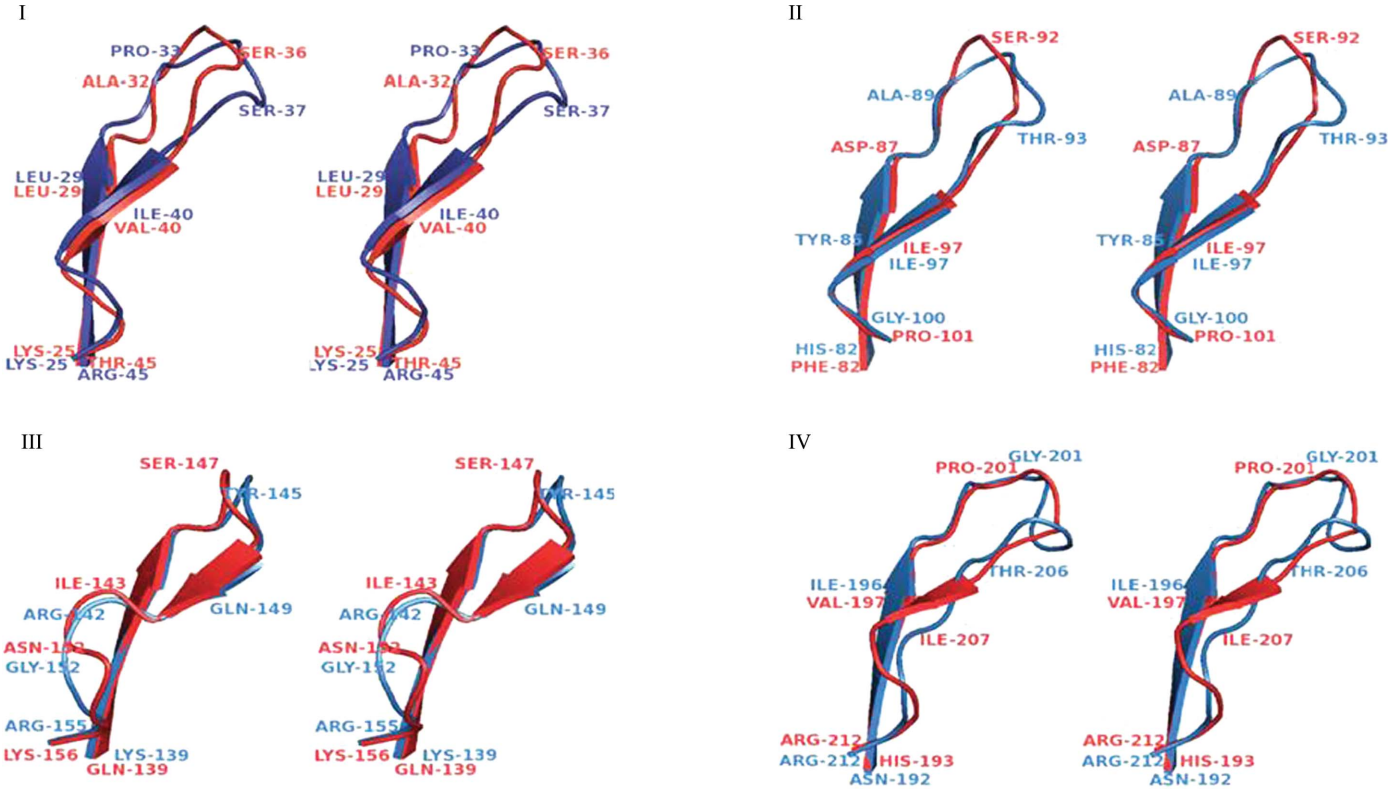


Figure 3
Structural differences between LS-24 from *L. sativus* and CP4 from *V. unguiculata*. Differences can be seen in the conformation of the loop interconnecting the outermost β-strands of the four blades (I–IV).

www.rcsb.org/pdb) with those of LS-24 (Gaur *et al.*, 2010) and CP4 (Fig. 4*a*) and alignment of the corresponding sequences. The haemopexin domains used to arrive at the signature sequence motif exist as subdomains of multimeric proteins, as in the cases of gelatinase A (PDB entry 1gen; Libson *et al.*, 1995), rabbit serum haemopexin (PDB entry 1hxn; Faber *et al.*, 1995), mammalian serum haemopexin (PDB entry 1qhu; Paoli *et al.*, 1999), collagenase-3 (PDB entry 1pex; Gomis-Ruth *et al.*, 1996), human proMMP-1 (PDB entry 1su3; Jozic *et al.*, 2005) and human MMP-12 (PDB entry 3ba0; Bertini *et al.*, 2008). The signature sequence consists of 21 amino acids and is comprised of three strands of a β-sheet and a short α-helix in the linker interconnecting the two blades. The fourth strand is ill-defined in terms of sequence consensus. The first and second β-strands are connected by a loop of 2–4 amino acids and the second and third β-strands are connected by a β-turn (Fig. 4*b*).

The conserved residues on the β-strands of each blade exhibit a definite amino-acid sequence pattern, being more hydrophobic in the middle and more hydrophilic towards the ends, with the length of the side chain gradually increasing from the N-terminus to the C-terminus. The striking feature of the signature sequence is the distribution of residues in the structural context. Of the eight conserved hydrophobic residues in the motif, three are oriented towards the concave surface, while the remaining five bulkier aromatic residues are oriented towards the convex surface of the β-sheet, resulting in differential hydrophobicity of the two surfaces (Fig. 4*b*). The predominance of bulky aromatic residues on one surface could indeed have resulted in a significantly curved topology of the β-sheet. Residue 9 of the signature sequence on the second β-strand, which is oriented towards the less hydrophobic surface, tethers the blade to the other surface by interacting with the more hydrophobic surface of the next blade. Likewise, residue 18 of the signature sequence cements the conserved helix to the core structure. An aspartate residue at the beginning of the first β-strand of each of the four blades is involved in chelating a Ca^{2+} ion, thereby tethering the blades at one central hub (Fig. 4*b*). Therefore, calcium seems to provide structural stability to the entire haemopexin fold.

It is interesting that the outermost strand of each blade, together with the inter-blade linker, in the haemopexin fold are highly variable and do not contribute to the signature sequence. The functional significance of the outermost strand in either ligand binding or

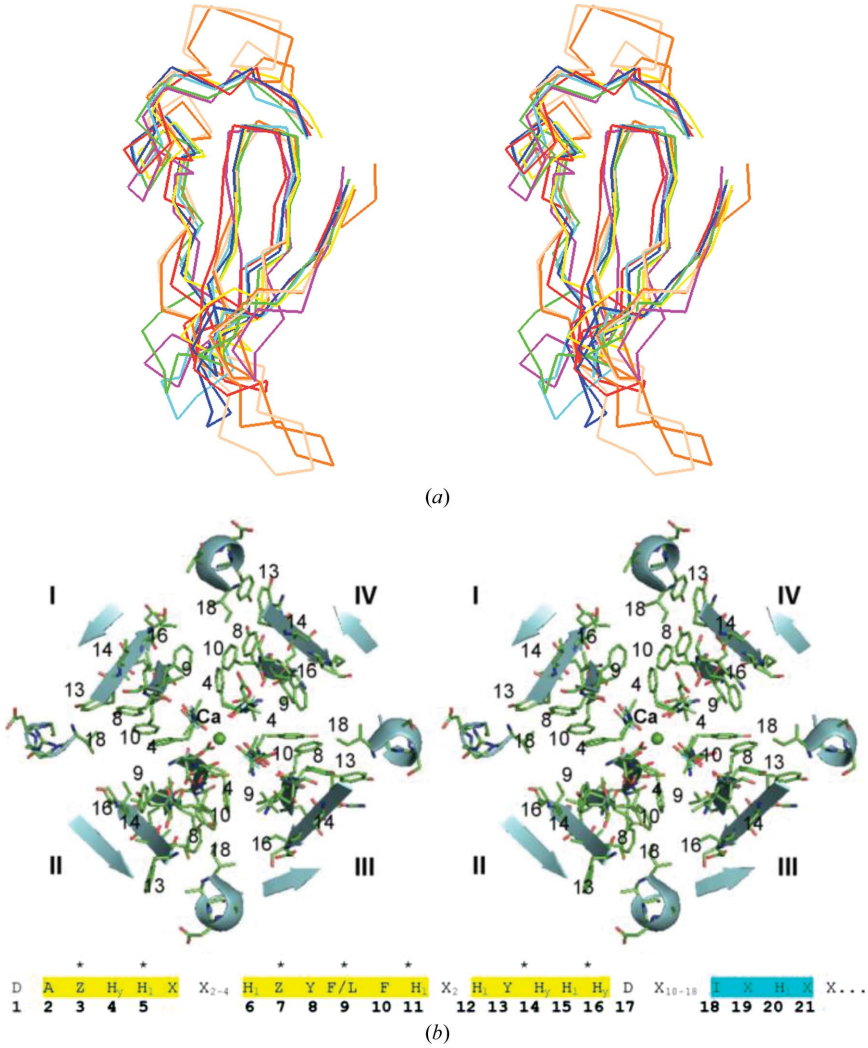


Figure 4
(*a*) Superimpositions of the N-terminal blades of proteins with a haemopexin-like motif (red, 1gen, C-terminal domain of gelatinase A; green, 1hxn, C-terminal domain of rabbit serum haemopexin; blue, 1pex, C-terminal domain of collagenase 3; yellow, 1qhu, mammalian serum haemopexin; magenta, 1su3, human proMMP-1; cyan, 3ba0, human MMP-12; wheat, 3lp9, LS-24; orange, CP4). (*b*) The signature sequence identified from structure-based sequence alignment (Z, residues with small side chains; H_y, hydrophobic residues; H_l, hydrophilic residues; X, any residue). Sequences highlighted in yellow and turquoise demarcate β-sheets and α-helices, respectively. The residues marked with asterisks are oriented toward the less hydrophobic surface of the β-sheet. The figure also presents a stereoview of CP4 with the signature sequence motif.

interaction with a protein subunit for oligomerization is evident in proteins adopting the haemopexin fold (Hrkal *et al.*, 1974; Paoli *et al.*, 1999; Jozic *et al.*, 2005; Iyer *et al.*, 2006).

In order to identify other possible haemopexin-like proteins from the plant world, a sequence search was carried out against a non-redundant database of protein sequences using the haemopexin signature sequence. Only three sequences belonging to plants could be picked up, including those of PA2 from *P. sativum*, MBSA from *V. radiata* and a protein from *Medicago truncatula*. Interestingly, in *M. truncatula* the fold exists as one of the domains of a larger protein. From the available sequences of plant haemopexins, a diversification of the haemopexin fold is therefore evident (Fig. 5*a*). A comparison

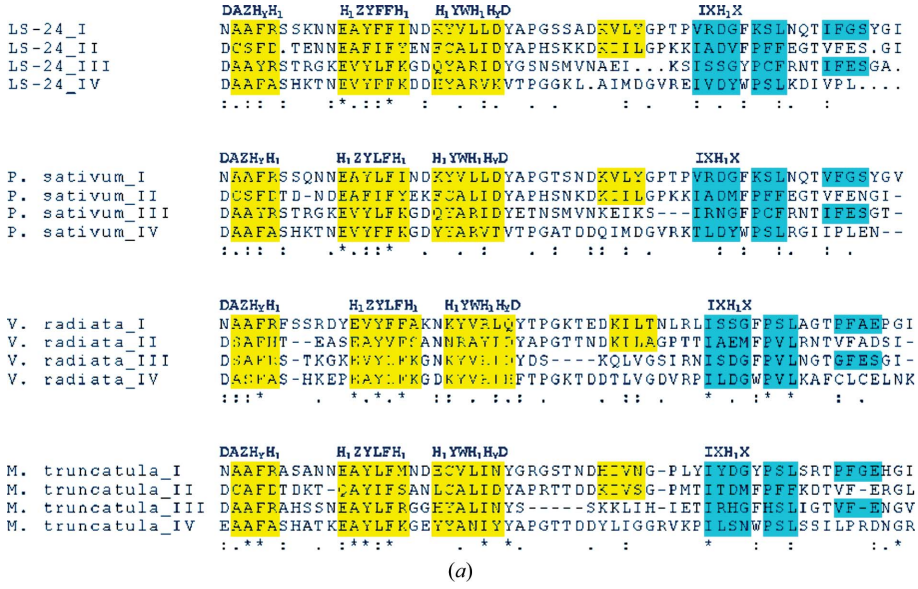


Figure 5
(*a*) Different blades together with secondary-structural assignments of haemopexin fold-containing proteins of plant origin as identified using the signature sequence motif. β-Strands are shown in yellow and α-helices are depicted in turquoise. (*b*) Dendrogram of haemopexin-like proteins from plants.

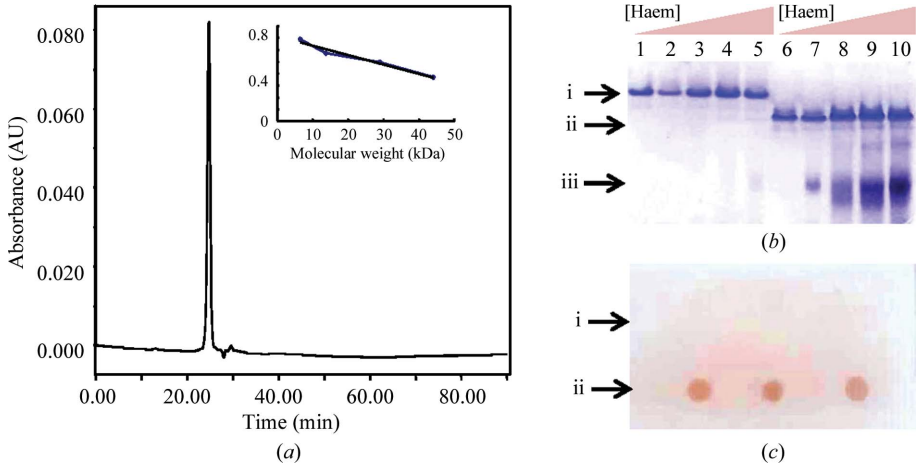


Figure 6
(*a*) CP4 molecular-weight estimation using gel-exclusion chromatography. The standard curve used to estimate the molecular weight of CP4 is shown in the inset. (*b*) Haem-binding analysis using an electrophoretic mobility-shift assay. Lanes 1–5 and lanes 6–10 in the assay contained CP4 and LS-24, respectively, pre-incubated with increasing concentrations of haem. i and ii depict the unbound forms of CP4 and LS-24, respectively, whereas iii depicts haem-bound LS-24. (*c*) Dot-blot assay in which native CP4 was probed with anti-spermine antibody (lane ii). Lactalbumin was used as a negative control (lane i).

of their primary structures revealed that all known proteins from plants with the haemopexin fold belong to one of at least two offshoots: one comprised of LS-24 and PA2 and the other consisting of CP4 and MBSA. Thus, it is apparent that the haemopexin fold has undergone divergent evolution in plants (Fig. 5*b*).

Screening of the protein-sequence database led to the detection of the motif in diverse taxonomic groups with *E* values ranging from 3.8 to 97.0. Essentially, 20 different proteins belonging to four subgroups were identified from diverse organisms: vitronectin, matrix metalloproteinases, hypothetical proteins and several assorted proteins including APC1 from *Dictyostelium discoideum*, CG2 from *Plasmodium falciparum* and cysteinyl-tRNA synthetases from *Taeniopygia guttata*, *Xenopus tropicalis*, *X. laevis*, *Gallus gallus* and *Mus musculus*.

3.4. Physiological implications

In light of the differences observed between CP4 and LS-24 both at the sequence and the structural levels, we compared the biochemical properties of the two proteins. CP4 was observed to be a monomer in solution using gel-exclusion chromatography (Fig. 6*a*). On the other hand, LS-24 was shown to be a dimer (Gaur *et al.*, 2010). Comparison of the dimerization sites in LS-24 with the corresponding regions in CP4 highlighted major differences in surface-charge properties that directly corroborated with the differences in the quaternary structure. Blade II of the LS-24 monomer is involved in dimerization and a direct structural comparison of the blade II region between LS-24 and CP4 revealed that the region is hydrophobic in the case of LS-24, whereas blade II is more polar in CP4. Hydrophobic interactions have previously been reported to be responsible for dimerization of LS-24 (Gaur *et al.*, 2010; Fig. 7).

It was evident from previous studies that LS-24 interacts with two different ligands, haem and spermine, utilizing two different oligomeric states (Gaur *et al.*, 2010). The haem-binding ability of CP4 was analyzed using a mobility-shift assay (Fig. 6*b*) as well as a surface plasmon resonance-based bioaffinity sensor (data not shown). No mobility shift of CP4 was observed with increasing concentrations of haem, suggesting that CP4 does not show any haem-binding ability, unlike LS-24. On the other hand, when probed with antispermine antibody in a dot-blot assay CP4 showed binding to spermine similar to that of LS-24 (Fig. 6*c*). While biochemical studies indicate the presence of spermine in the protein purified from the source, spermine was not located bound to CP4 in the crystal structure. This could be a consequence of a low occupancy of bound spermine or packing constraints in the crystal lattice.

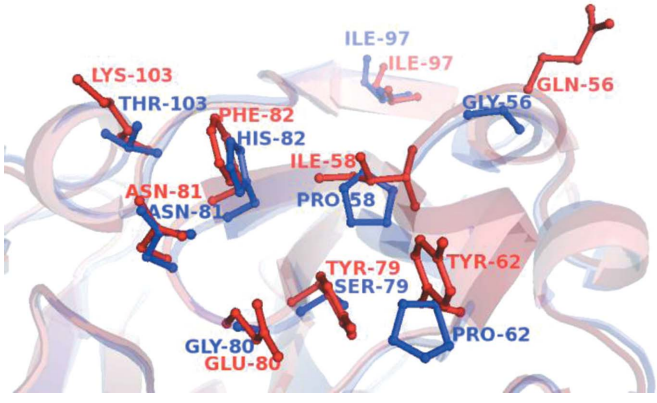


Figure 7
Superimposition of LS-24 residues involved in dimerization (red) with the corresponding residues of CP4 (blue).

The ability of CP4 to interact with spermine strengthens the notion that haemopexin-fold proteins are involved in polyamine metabolism in plants. However, the inability of CP4 to interact with haem indicates that it may not be involved in the direct regulation of oxidative stress as is the case for LS-24. With regard to other haemopexin domains, no spermine-binding ability has been reported in the case of mammalian haemopexins. A high haem-binding affinity has been observed and demonstrated crystallographically for mammalian serum haemopexin (Paoli *et al.*, 1999). It has often been observed that a single protein fold has diversified enormously in order to account for multiple functionalities. In fact, the haemopexin fold itself is employed to carry out numerous unrelated functions in mammals. Indeed, while the repertoire of structural folds is limited, the number of functions that are carried out by utilizing this limited number of structural folds is very large. The observed differences between CP4 and LS-24 reflect this functional diversification.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bertini, I., Calderone, V., Fragai, M., Jaiswal, R., Luchinat, C., Melikian, M., Mylonas, E. & Svergun, D. I. (2008). *J. Am. Chem. Soc.* **130**, 7011–7021.
Blancher, C., Omri, B., Bidou, L., Pessac, B. & Crisanti, P. (1996). *J. Biol. Chem.* **271**, 26220–26226.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chanana, V., Kaur, K. J. & Salunke, D. M. (2004). *Acta Cryst.* D**60**, 2100–2103.
Crennell, S. J., Tickler, P. M., Bowen, D. J. & ffrench-Constant, R. H. (2000). *FEMS Microbiol. Lett.* **191**, 139–144.
Croy, R. R., Hoque, M. S., Gatehouse, J. A. & Boulter, D. (1984). *Biochem. J.* **218**, 795–803.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Faber, H. R., Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. (1995). *Structure*, **3**, 551–559.
Fülöp, V. & Jones, D. T. (1999). *Curr. Opin. Struct. Biol.* **9**, 715–721.
Gaur, V., Qureshi, I. A., Singh, A., Chanana, V. & Salunke, D. M. (2010). *Plant Physiol.* **152**, 1842–1850.
Gomis-Ruth, F. X., Gohlke, U., Betz, M., Knauper, V., Murphy, G., Lopez-Otin, C. & Bode, W. (1996). *J. Mol. Biol.* **264**, 556–566.
Gruen, L. C., Guthrie, R. E. & Blagrove, R. J. (1987). *J. Sci. Food Agric.* **41**, 167–178.
Harris, N. & Croy, R. R. D. (1985). *Planta*, **165**, 522–526.
Higgins, T. J. V., Beach, L. R., Spencer, D., Chandler, P. M., Randall, P. J., Blagrove, R. J., Kortt, A. A. & Guthrie, R. E. (1987). *Plant Mol. Biol.* **8**, 37–45.
Hrkal, Z., Vodrazka, Z. & Kalousek, I. (1974). *Eur. J. Biochem.* **43**, 73–78.
Iyer, S., Visse, R., Nagase, H. & Acharya, K. R. (2006). *J. Mol. Biol.* **362**, 78–88.
Jenne, D. (1991). *Biochem. Biophys. Res. Commun.* **176**, 1000–1006.
Jozic, D., Bourenkov, G., Lim, N.-H., Visse, R., Nagase, H., Bode, W. & Maskos, K. (2005). *J. Biol. Chem.* **280**, 9578–9585.
Kolberg, J., Michaelsen, T. E. & Sletten, K. (1983). *Hoppe Seylers Z. Physiol. Chem.* **364**, 655–664.
Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988). *Science*, **240**, 1759–1764.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Libson, A. M., Gittis, A. G., Collier, I. E., Marmer, B. L., Goldberg, G. I. & Lattman, E. E. (1995). *Nature Struct. Biol.* **2**, 938–942.
Liu, T., Lin, Y., Cislo, T., Minetti, C. A., Baba, J. M. & Liu, T.-Y. (1991). *J. Biol. Chem.* **266**, 14813–14821.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Paoli, M., Anderson, B. F., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. (1999). *Nature Struct. Biol.* **6**, 926–931.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Piccard, H., Van den Steen, P. E. & Opdenakker, G. (2007). *J. Leukoc. Biol.* **81**, 870–892.
Qureshi, I. A., Sethi, D. K. & Salunke, D. M. (2006). *Acta Cryst.* F**62**, 869–872.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vigeolas, H., Chinoy, C., Zuther, E., Blessington, B., Geigenberger, P. & Domoney, C. (2008). *Plant Physiol.* **146**, 74–82.
Vioque, J., Clemente, A., Sánchez-Vioque, R., Pedroche, J., Bautista, J. & Millán, F. (1998). *J. Agric. Food Chem.* **46**, 3609–3613.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.