The crystal structure of the *Leishmania major* surface proteinase leishmanolysin (gp63)

Edith Schlagenhauf†, Robert Etges‡ and Peter Metcalf*

Background: Despite their medical importance, there is little available structural information for the surface antigens of infectious protozoa. Diseases caused by the protozoan parasite *Leishmania* are common in many developing countries. Human infection occurs during the bite of infected sandflies, when *Leishmania* promastigote cells from the insect gut enter the bloodstream. Promastigotes in the blood parasitize macrophages, often causing serious disease. Leishmanolysin is the predominant protein surface antigen of promastigotes, and is assumed to have a key role during infection. Leishmanolysin is a membrane-bound zinc proteinase, active *in situ*. Similar molecules exist in other trypanomastid protozoa.

Results: Two crystal forms of leishmanolysin were obtained from protein purified from promastigote membranes. A single lead derivative in both crystal forms was used to solve the structure. The structure reveals three domains, two of which have novel folds. The N-terminal domain has a similar structure to the catalytic modules of zinc proteinases. The structure clearly shows that leishmanolysin is a member of the metzincin class of zinc proteinases.

Conclusions: The unexpected metzincin features of the leishmanolysin structure suggest that the metzincin fold may be more widespread than indicated by sequence homologies amongst existing metzincin zinc proteinases. The similarity of the active-site structure to previously well characterized metzincin class zinc proteinases should aid the development of specific inhibitors. These inhibitors might be used to determine the function of leishmanolysin in the insect and during mammalian infection, and may aid the development of drugs for human leishmaniasis.

Address: EMBL Heidelberg, Biological Structures and Biocomputing Programme, Meyerhofstrasse1, Postfach 10.2209, D-69012 Heidelberg, Germany.

Present addresses: †Institute of Plant Biology, University of Zurich, CH-8008 Zurich, Switzerland and ‡Institut fuer Biomedizinsche Forschung und Beratung (IBFB) GmbH, Arthur-Hoffman Strasse 7, D-04107 Leipzig, Germany.

*Corresponding author.
E-mail: Peter.Metcalf@auckland.ac.nz

Key words: crystal structure, gp63, *Leishmania,* leishmanolysin

Received: **19 February 1998**
Revisions requested: **27 March 1998**
Revisions received: **30 June 1998**
Accepted: **2 July 1998**

Structure 15 August 1998, **6**:1035–1046
http://biomednet.com/elecref/0969212600601035

Introduction

The kinetoplastid protozoan parasites of the genus *Leishmania* cause diseases for which treatment is difficult and for which no effective vaccine exists [1,2]. Promastigotes injected during the feeding of infected sandflies parasitize macrophages, causing symptoms ranging in severity from self-healing sores, through severely disfiguring mucosal lesions (*espundia*), to a fatal systemic infection involving spleen and bone marrow (*kala azar*). Promastigotes of all *Leishmania* species express an abundant surface glycoprotein, leishmanolysin [3], that is thought to be a ligand involved in the interaction of the parasite with defensive systems of the host, including components of the complement system and the macrophage surface [4–6]. Leishmanolysin is therefore an attractive vaccine candidate [7]. This glycophosphatidylinositol (GPI)-anchored glycoprotein [8], was identified as a neutral metalloproteinase with activity in a broad pH range extending from around pH 6 to 9 (EC 3.4.24.36; leishmanolysin was previously termed gp63 or PSP promastigote surface proteinase) [9]. Protein sequence databases currently contain 13 distinct leishmanolysin sequences which, with the exception of a short zinc-binding motif (HEXXH), have no apparent homology to other proteins. Similar molecules occur in *Trypanosoma brucei* [10], the parasite causing African sleeping sickness, and in the insect parasites without mammalian hosts (*Crithidia* and *Herpetomonas*) [3]. At least 60% of the leishmanolysin residues are strictly conserved, including the zinc proteinase motif HEXXH and all 19 cysteine residues.

Leishmanolysin is the major protein component of the promastigote surface (~5×10^5 molecules/cell), where it is enzymatically active against polypeptide substrates [9]. Together with the other major cell surface components, the lipophosphoglycans (LPGs) and glycoinositolphospholipid (GIPL) [11], it is likely to have important roles in enabling the small number (< 1000) of promastigotes injected by the feeding sandfly [12] to reach and infect sufficient macrophages to establish an infection. Membrane-bound leishmanolysin expression by amastigotes (the form of *Leishmania* within macrophages) is much reduced, although a soluble form of leishmanolysin is expressed at lower levels within the lysosome of the parasite [13,14]. The protein is synthesized in the endoplasmic reticulum where the signal

sequence is cleaved post-translationally, N-linked carbohydrate is added, and the C-terminal ~25 amino acid residues are replaced with a GPI membrane anchor structure. Subsequent proteolytic processing involves the cleavage of a propeptide, to produce the mature polypeptide chain. In the case of *Leishmania major* the mature protein starts at residue Val100 [15], has carbohydrate attachment sites at Asn300, Asn407 and Asn534 and terminates at residue Asn577, which is covalently bound via ethanolamine to the carbohydrate part of the GPI anchor.

We present here the crystal structure of leishmanolysin. This is the first structure of a *Leishmania* surface molecule to be reported and may also prove to be the first for a membrane-bound proteinase. We describe the molecule and show that the active-site region is similar to that of previously well characterized metzincin class zinc proteinases for which inhibitors already exist, suggesting future drug design possibilities.

Results and discussion

Overall structure

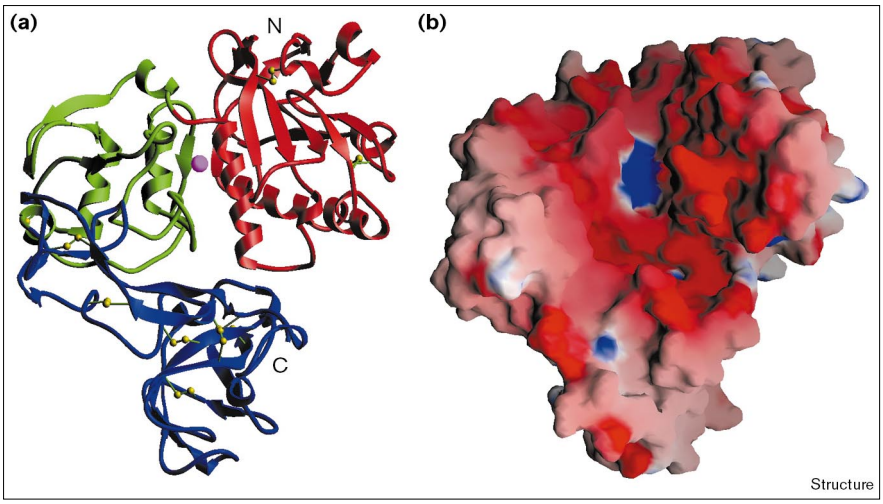
Leishmanolysin was purified from membrane preparations of cultured stationary phase *L. major* strain LRC-L119 promastigotes [16]. Treatment with phosphatidylinositol-specific phospholipase C was used to remove the lipid part of the GPI anchor, and subsequent purification produced yields sufficient for crystallization trials [15,17]. Tetragonal crystals diffracting to 2.5 Å obtained with the precipitant 2-methyl-2,4-pentanediol (MPD) were used for much of the derivative search. Monoclinic crystals diffracting to at least 1.8 Å were obtained using macroseeding methods with similar precipitant conditions. The tetragonal and monoclinic crystal forms each contain a single leishmanolysin molecule in the asymmetric unit. The crystallized molecule is intact solubilized leishmanolysin, with 478 amino acid residues and three carbohydrate groups including the ~1100 Dalton GPI anchor. The structure was solved using crystals of both forms derivatized with trimethyllead acetate. Phases derived from five lead-atom sites found in the monoclinic cell were sufficient to determine an approximate molecular boundary. Electron density within the boundary could be correlated with electron density derived from the single lead site found in the tetragonal cell and averaging methods were used to produce an interpretable map.

Leishmanolysin is a compact molecule containing predominantly β sheet secondary structure. This is in contrast to the only other protozoan surface antigen structures determined to date: two *Trypanosoma brucei* variant surface glycoprotein (VSG) ~45 kDa N-terminal fragments that are dimers with elongated coiled α-helical folds [18,19]. The overall dimensions of the leishmanolysin molecule are $45 \times 50 \times 70$ Å. The protein consists of three domains: the N-terminal, central and C-terminal domains (Figure 1). The N-terminal domain was recognized at an early stage of the structure determination as having a fold similar to that of the catalytic modules of zinc proteinases [20]. This domain contains a catalytic zinc atom and an active-site helix containing the two histidines of the zinc proteinase sequence motif HEXXH. There are deep surface indentations at domain interfaces that join the active-site cleft, which is located between the N-terminal domain and the central domain of the molecule (Figure 1). The calculated surface charge distribution shows a large region of negative charge surrounding the active-site cleft, but charge distribution features suggesting the orientation of the molecule on the promastigote membrane are not evident.

N-terminal domain

Three-dimensional structures of zinc proteinases containing the HEXXH sequence motif have been grouped into

Figure 1



Leishmanolysin domain structure and surface representation. **(a)** Ribbon representation of the structure with the N-terminal domain shown in red, the central domain in green, and the C-terminal domain in blue. Disulfide bonds are shown in yellow and the active site zinc atom is represented as a magenta sphere. The molecule is viewed looking into the active-site cleft. **(b)** Surface representation of the molecule with regions of negative potential shown in red and positive regions in blue.

two families: the bacterial thermolysins and the metzincins (including serralysins, reprolysins, matrixins and astacin [21]). Molecules of both classes have a major catalytic domain with a topology consisting of two α helices (A and B) packed against one side of a five-stranded twisted β sheet. The HEXXH motif is located on helix B and the two histidine sidechains are ligands of the active site zinc atom. The N-terminal domain of leishmanolysin has these basic features (Figures 2 and 3) and includes 40- and 35-residue inserted 'flaps' which cover the side of the β sheet opposite from the helices (Figure 4). The 40-residue flap replaces the second strand of the central sheet in the conserved topology. There are two disulfide bonds within this domain, one within the 35-residue 'flap' and the other linking the 40-residue flap to strand S11 of the central sheet, which borders the active-site cleft.

Central domain

Residues 274–391 of leishmanolysin (Figures 2b and 3b) fold into a compact domain with antiparallel helices H11 and H12 forming the core of the domain together with the antiparallel sheet formed by strands S13, S14 and S15. Residues in the long loop joining H11 and H12 form part of the active site. Strand S16 forms an antiparallel sheet with strand S6, and in doing so forms the main connection between the N-terminal and central domains. A single disulfide bond connects the C terminus of the domain to the beginning of the long loop between H11 and H12.

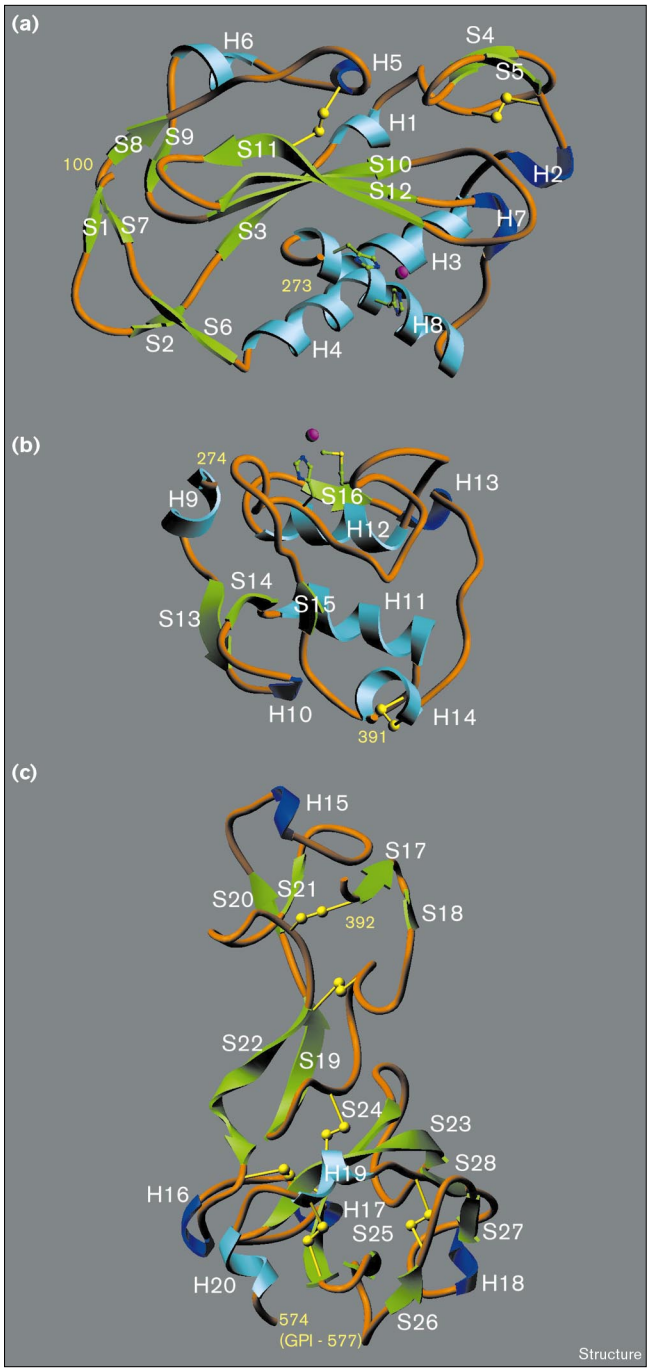
Metzincin class zinc proteinases have an extended zinc proteinase motif HExxHxxGxxH, where the glycine residue is part of a tight turn at the end of active-site helix B, allowing the additional histidine of the motif to form the third ligand of the active site zinc atom. Metzincins share two additional structural features, an unusual tight 1,4 'met-turn' with a conserved methionine at the base of the active site, and a following helix (helix C), which is roughly parallel to the active-site helix B [21].

The central domain of leishmanolysin contains a corresponding met-turn (Asp342–Ala348) and a helix C (H12), but there is an unexpected 62-amino acid insertion between the glycine and third histidine residue (Gly271–His334) of the metzincin motif that represents half of the central domain (Figure 4). Leishmanolysin is clearly a metzincin class zinc proteinase, but does not satisfy the defining sequence motif HExxHxxGxxH because of the insert between the glycine (Gly271) and the last histidine (His334).

The active site

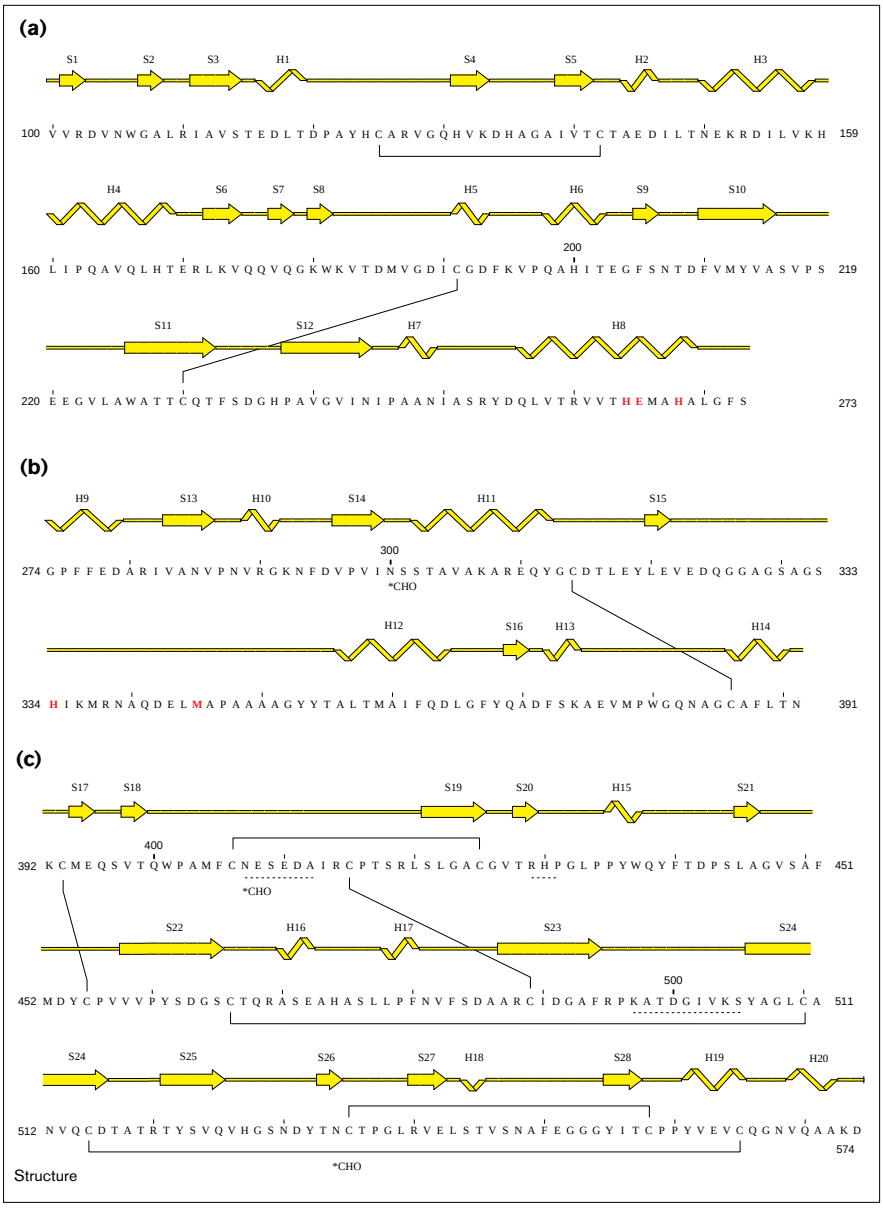
The geometry of the leishmanolysin active-site residues in the immediate vicinity of the catalytic zinc atom is similar to that of other zinc proteinases. The zinc atom is coordinated to the ε-nitrogen atoms of His264, His268 and His334 at distances of 2.18 Å, 2.18 Å and 2.12 Å respectively, similar to values observed in other zinc proteinase

Figure 2



Leishmanolysin domains and secondary structure. **(a)** The N-terminal domain containing the active-site helix H8 with the conserved zinc-binding motif HEXXH. **(b)** The central domain. Residues in the loop preceding H12 contribute to the active site: the zinc-binding ligand His334 and Met345 are structurally conserved in metzincin family metalloproteinases. **(c)** The C-terminal domain. Active-site histidine, glutamate and methionine residues, and disulfide bonds (in yellow) are shown in ball-and-stick representation; the zinc atom is shown as a magenta sphere. The elements of secondary structure are color-coded: random-coil structure is shown in brown, β sheet in green and α and 3_{10} helices are in light and dark blue, respectively.

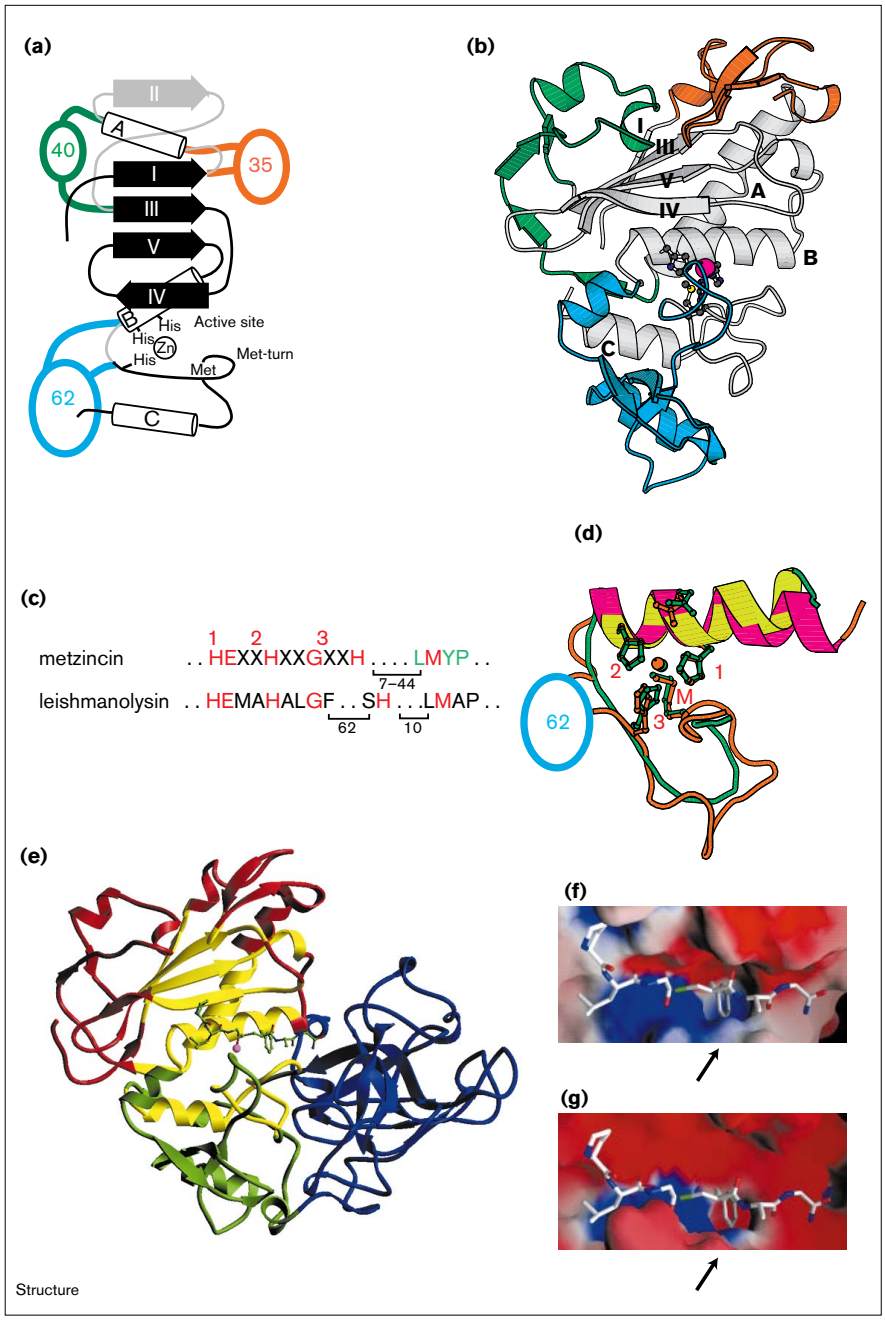
Figure 3



The amino acid sequence and secondary structure elements of leishmanolysin. The sequences of the three domains are shown: **(a)** the N-terminal domain; **(b)** the central domain; and **(c)** the C-terminal domain. In the sequence, active-site residues are shown in red, disulfide bonds are indicated by connecting thin lines, and carbohydrate attachment residues are marked by *CHO. Dotted lines below the sequence show residues which are disordered in the structure. The GPI anchor is attached to Asn577.

structures (2.0–2.2 Å) [20]. The arrangement of the three zinc ligand histidine sidechains (His264, His268 and His334) and the underlying met-turn methionine (Met345) are very similar to those for metzincin class zinc proteinases, such as astacin [22] and collagenase [23]. The structures of zinc proteinases determined in the absence of inhibitors have a fourth zinc ligand — a water molecule — in addition to the three histidine ligands. This water molecule is held between the zinc and the oxygens of the conserved glutamic acid in the zinc proteinase motif HEXXH. Astacin is unusual in having a fifth zinc ligand — a tyrosine phenolic oxygen. The water molecule zinc ligand is thought to have a key role in proteolysis, being the nucleophile which attacks the substrate peptide bond. Electron density at the expected position of the catalytic water molecule zinc ligand is absent in the leishmanolysin electron-density map and the glutamate sidechain is twisted away from the usual position it adopts in other zinc proteinases. The leishmanolysin active-site zinc appears to be pentacoordinated, with two ligands to a region of connected density modeled as glycine (Figure 5a). (The origin of this electron density, modeled as glycine, is unclear. Similar density is seen in both the monoclinic and tetragonal maps, and although 100 mM pH 9.5 ethanolamine-HCl and 100 mM pH 9 glycine-NaOH buffers were used to produce monoclinic crystals,

Figure 4

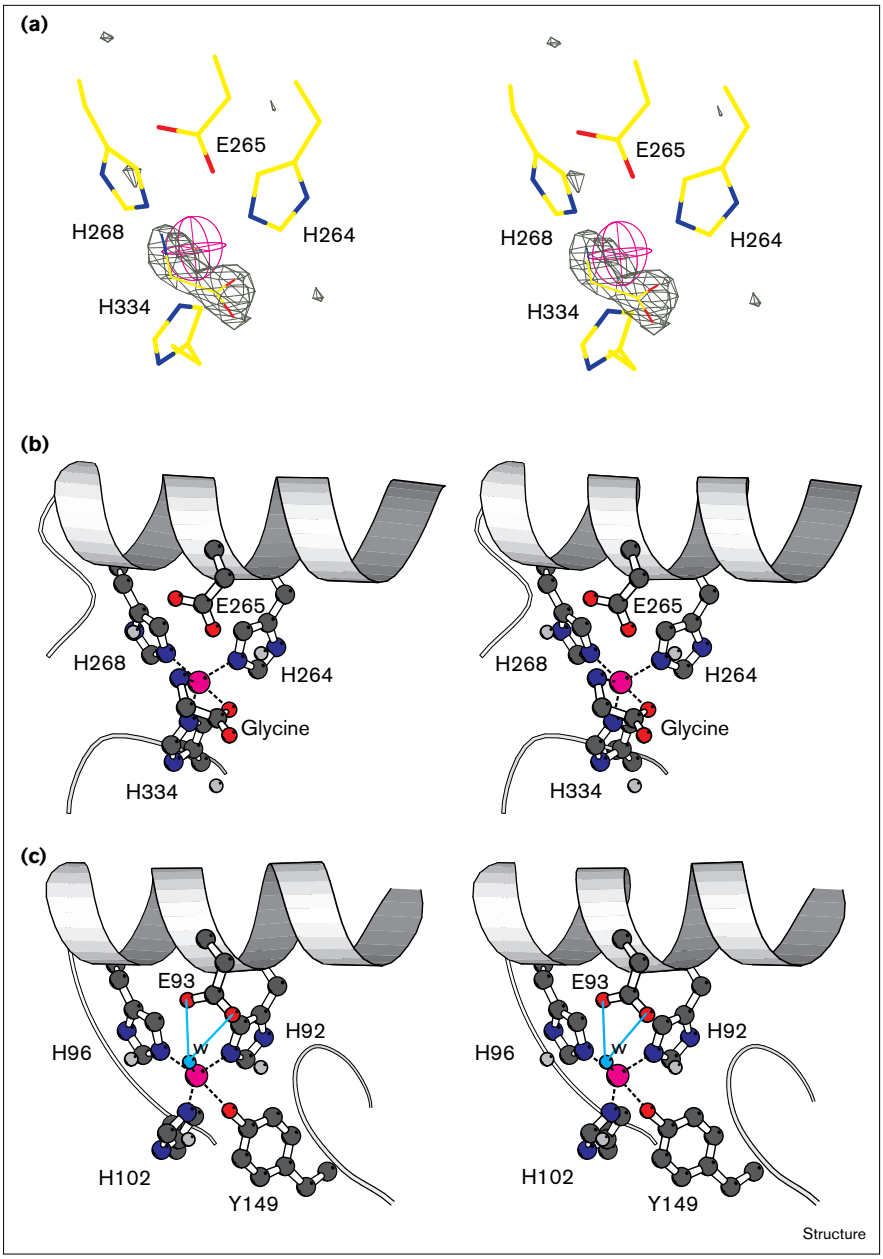


The classification of leishmanolysin as a metzincin class zinc proteinase and comparisons with collagenase. **(a)** The conserved secondary structure elements of the metzincin family [21]. Leishmanolysin lacks strand II of the central β sheet and includes three insertions (colored circles with number of inserted residues). **(b)** Ribbon diagram showing the part of leishmanolysin (residues 110–366) with the metzincin structure (grey) and the three insertions (colored as in (a)) in the standard zinc proteinase orientation [21]. The conserved histidine ligands and the met-turn methionine sidechains are shown in ball-and-stick representation; the active site zinc atom is shown as a magenta sphere. The conserved metzincin secondary structure elements and the leishmanolysin equivalents are as follows: metzincin strand I = leishmanolysin strand S3; III = S10, IV = S11; V = S12; helix A = H3, H4; B = H8; and C = H12. **(c)** The metzincin and leishmanolysin sequence motifs. The conserved histidine, glycine and methionine residues are shown in red (additional residues in the collagenase met-turn are shown in green). The spacer between the third histidine and the methionine residues varies between 7–44 residues depending on the metzincin member. **(d–g)** Comparison of leishmanolysin and the metzincin collagenase. The structures of leishmanolysin and two similar collagenase catalytic module–inhibitor complexes (PDB codes 1jap [23] and 1jao [44]) were superimposed using the program O [36]. (d) Overlay of the leishmanolysin (orange) and collagenase (green) active-site regions. (e) Ribbon diagram of leishmanolysin with domains coloured as in Figure 1 showing metzincin core regions (yellow) where corresponding collagenase Cα atoms are within 3.6 Å of aligned leishmanolysin equivalents. The inhibitors of 1jap and 1jao indicate the position of the active-site cleft. The inhibitors on the surface of (f) collagenase and (g) leishmanolysin. The S1′ hole is marked with an arrow in each figure.

only the ethanolamine buffer was used to prepare tetragonal crystals [17]. Both ethanolamine and glycine can be modeled into the density with similar N–O zinc coordination and glycine was chosen as it fits the observed electron density better.)

In the model, one zinc ligand is the glycine amino nitrogen which is near to the position normally occupied by the catalytic water in zinc proteinase structures (Figures 5b,c). A glycine carboxyl oxygen forms the fifth zinc ligand and is located near to the position of the phenolic oxygen of the tyrosine which forms the fifth zinc ligand in astacin (Figures 5b,c). Similar N–O zinc coordination has been observed recently in a complex of the metzincin human matrix metalloproteinase stromelysin-1 with the protein inhibitor TIMP-1 (tissue inhibitor of metalloproteinase-1),

Figure 5



The active site of leishmanolysin and comparison with the metzincin zinc proteinase astacin. **(a)** Stereoview figure illustrating the electron density attributed to a glycine molecule adjacent to the leishmanolysin active site zinc atom (magenta sphere). Zinc ligands are shown in stick representation. The glycine shown within the density was omitted from the model during calculation of the simulated annealing F_o–F_c omit map, shown contoured at 3.5σ. **(b)** Stereoview of the leishmanolysin active site. The zinc atom (magenta sphere) is shown with five ligands: εN atoms from His264, His265 and His334 and the amino nitrogen and a carboxyl oxygen from the glycine molecule fitted to the electron density shown in (a). Water molecules are shown as light grey spheres; atoms are in standard colors. **(c)** Stereoview of the active-site of astacin. The catalytic water (W) is held in place by the sidechain of Glu93 (cyan lines). The corresponding water molecule is absent in leishmanolysin. The five astacin zinc ligands include the catalytic water and the phenolic oxygen of Tyr149, which are in similar positions to bonds to the glycine molecule in the leishmanolysin model.

involving the amino nitrogen and carbonyl oxygen of the first residue of the inhibitor [24].

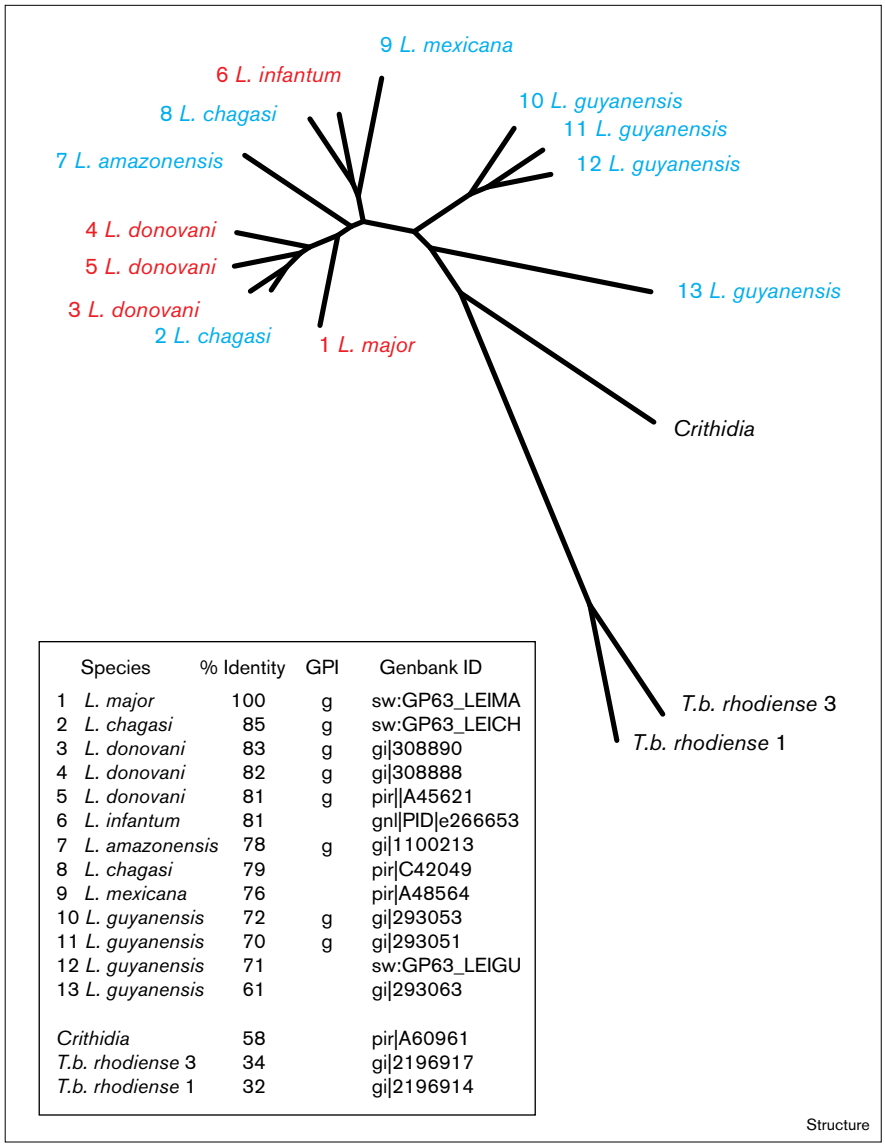
The similarity of leishmanolysin with metzincin class zinc proteinases extends beyond the immediate vicinity of the active site zinc atom. When the metzincin human neutrophil collagenase (PDB code 1jap [23]) is aligned to leishmanolysin, 89 corresponding Cα positions in the active-site regions of the two molecules match with a root mean square (rms) difference of 1.8 Å (53 of these residues are located in the leishmanolysin N-terminal domain and 36 in the central domain; Figure 4e). Leishmanolysin and collagenase share residues near the active site in addition to the conserved metzincin motif residues (Figure 4c). The similarity of leishmanolysin and collagenase in this region is evident in the active site surface representation, where both molecules have large S1′ substrate pockets. Collagenase inhibitors in the structural alignment fit reasonably well to the leishmanolysin surface (Figures 4f,g).

C-terminal domain

The elongated C-terminal domain of leishmanolysin (Figures 2c and 3c) consists of mainly antiparallel β strand and random coil structure with only minor helical contributions.

Figure 6

Tree representation of amino acid sequence distances of leishmanolysin homologs in the GenBank database. The diagram includes leishmanolysin sequences from old world (red) and new world (blue) *Leishmania* species, the insect parasite *Crithidia fasciculata* and the African trypanosome *Trypanosoma brucei*. Distances were calculated using sequences corresponding to mature *L. major* leishmanolysin (residues 100–577). The key gives the GenBank sequence identification codes and homology (% identity) with *L. major* leishmanolysin. Sequences sharing the *L. major* GPI anchor attachment site, Asn577, are marked (g).



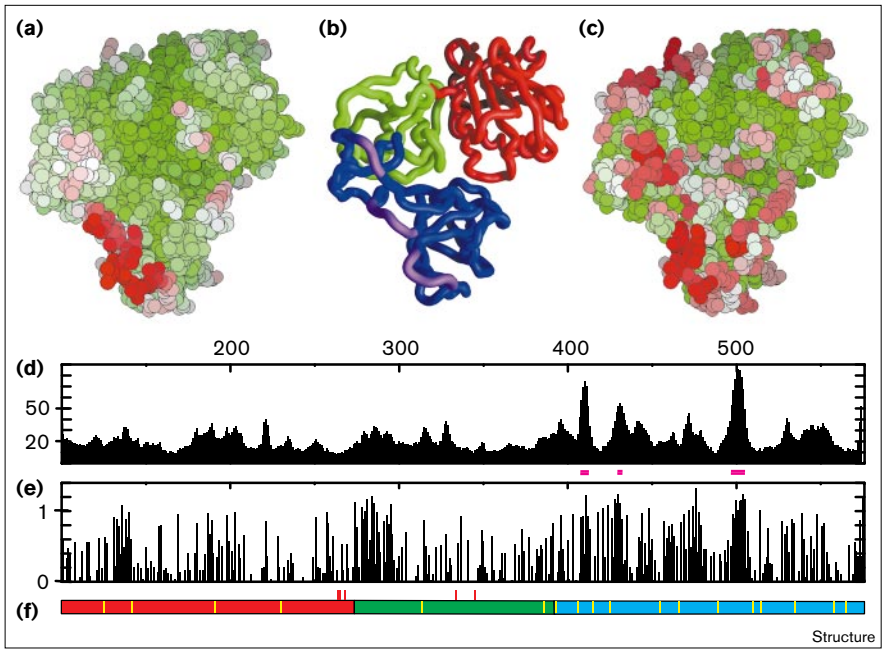
	Species	% Identity	GPI	Genbank ID
1	*L. major*	100	g	sw:GP63_LEIMA
2	*L. chagasi*	85	g	sw:GP63_LEICH
3	*L. donovani*	83	g	gi|308890
4	*L. donovani*	82	g	gi|308888
5	*L. donovani*	81	g	pir||A45621
6	*L. infantum*	81		gnl|PID|e266653
7	*L. amazonensis*	78	g	gi|1100213
8	*L. chagasi*	79		pir|C42049
9	*L. mexicana*	76		pir|A48564
10	*L. guyanensis*	72	g	gi|293053
11	*L. guyanensis*	70	g	gi|293051
12	*L. guyanensis*	71		sw:GP63_LEIGU
13	*L. guyanensis*	61		gi|293063
	Crithidia	58		pir|A60961
	T.b. rhodiense 3	34		gi|2196917
	T.b. rhodiense 1	32		gi|2196914

The domain contains six disulfide bonds. The N-terminal part of the domain contains three two-stranded antiparallel sheets (formed from strands S17 and S18; S20 and S21; S19 and S22). These sheets are connected via two short helices (H16 and H17) to an antiparallel six-stranded β sheet folded into a sandwich structure (comprising strands S23–28) and this is in turn attached via two short helices (H19 and H20) to the GPI anchor. The two C-terminal residues, the GPI anchor, and three other segments of this domain (indicated in the sequence underlined with dots; Figures 2c and 3c) have weak electron density, and are assumed to be either mobile or disordered in the crystal. There is no evident similarity with known structures or with the structure of acetylcholinesterase [25], the only other known structure of a GPI-anchored protein.

Sequence variability

Amino acid sequences of surface antigens from organisms that infect mammals often show increased variation, reflecting changes occurring to overcome developing host immunity. Well characterized examples of antigenic variation in human pathogens include the evolution of viral coat protein sequences associated with influenza epidemics and the extreme variability of VSG sequences expressed during *T. brucei* infection. Although there is no evidence for similar antigenic variation in *Leishmania*, leishmanolysin sequences derived from different species do differ, and developmental changes in the expression of different genes in *Leishmania chagasi* has been reported [26]. Leishmanolysin genes occur in multiple copies with genomic complexity ranging from estimates of six genes

Figure 7



Structural flexibility and sequence variation. **(a)** Surface representation of crystallographic B factors. Red regions indicate ill-defined parts of the structure where there is weak electron density and high B factors; green areas indicate well-defined regions with low B factors. **(b)** Skeleton representation showing the three domains of leishmanolysin (N-terminal, red; central, green; and C-terminal, blue). The three segments with ill-defined structure in the C-terminal domain are shown in purple. These segments are those shown underlined with dots in the sequence (Figure 3) and with magenta bars below in part (d) of this figure. **(c)** Surface representation of sequence variability. Regions in red differ most in an alignment of 13 leishmanolysin sequences from different species (Figure 6; regions in green differ least). **(d)** Averaged mainchain B factors for each residue; magenta bars mark the three ill-defined regions shown in (b) where B > 50. **(e)** Sequence variability. The plot shows the inverse of the per-residue similarity measure generated by the UWGCG program PLOTSIMILARITY [45] from the 13 sequence leishmanolysin alignment. **(f)** Bar diagram of the sequence showing the three domains. Upper red bars show conserved active-site residues (Figure 5c) and yellow lines indicate cysteine residues, which are conserved in leishmanolysin sequences.

with five in a single tandem array in *L. major* [27] to two or more tandem arrays containing a total of more than 70 gene copies in *Leishmania braziliensis* [28]. Although restriction fragment length polymorphism (RFLP) analysis of natural isolates indicates extensive polymorphism [28,29], amino acid sequence variability within the coding region of the mature protein (residues 100–577 for *L. major*) is limited [30–32]. *Leishmania* sequences are at least 60% identical (Figure 6) and the mature proteins share all 18 cysteines, the active site zinc ligand and methionine-turn residues with the less conserved leishmanolysin homologs from *Crithidia* and *T. brucei*. The *L. major* carbohydrate-linked residues (Asn300, Asn407 and Asn534) and the GPI anchor attachment site at Asn577 are not conserved in *Leishmania* species. As expected, surface residues are less conserved than interior residues, and a surface representation of variability suggests sequence variability is correlated with structural flexibility, as indicated by crystallographic B factors (Figure 7). The C-terminal domain is the least conserved with the three disordered loops showing the greatest sequence variation (Figure 7).

Discussion

The discovery of a metzincin fold in leishmanolysin raises interesting questions about the evolutionary relationships of zinc proteinases and suggests that this fold may be more widespread than indicated by the metzincin sequence motif HExxHxxGxxH alone. The 62-residue segment seen in leishmanolysin, which replaces the two residues between the glycine and last histidine of the motif, indicates that significant changes can occur in metzincin proteinases close to the active site that are consistent with proteolytic activity. Because trypanomastids apparently predate other species where metzincins are known to occur [33], leishmanolysin may well have existed before mammalian metzincins. Despite the apparent evolutionary gap, the structural similarity seen in the vicinity of the active site is still sufficient to suggest that existing collagenase inhibitors could be adapted to produce specific inhibitors for leishmanolysin.

Leishmanolysin is a major component of the promastigote surface, and is probably involved in host–parasite interactions both in infected insects, and in the early stages of mammalian infection. Little is known of the role of leishmanolysin in insects, but similar proteinases occur in insect parasites without alternative hosts (*Crithidia* and *Herpetomonas*) or in parasites with very different mammalian host environments (*T. brucei*), suggesting a primary role for the proteinase in the insect stage of the *Leishmania* life cycle. Mammalian infection is an obligatory stage of the *Leishmania* life cycle, so there is likely to be strong selective pressure on the surface molecules which first contact the mammalian immune system in order to maximize infection probability. Immediately after infection promastigotes fix and activate complement [14] and leishmanolysin has been shown to reduce subsequent complement-mediated lysis [34]. C3-opsonized leishmanolysin has been reported to be

involved in promastigote–macrophage binding [5]. Leishmanolysin antigen-specific immune responses to promastigotes and macrophages carrying amastigotes may be significant and trials of several leishmanolysin-based vaccines have been carried out. Although there is no evidence for antigenic variation in *Leishmania*, the polymorphism seen in leishmanolysin sequences from different strains does allow invariant surface regions to be identified on the structure which may prove to be useful for future peptide vaccine studies.

Treatment for leishmaniasis, although usually effective, is not easy and improved drugs are clearly needed. Furthermore, parasite resistance to the currently used antimonials has been reported. Leishmanolysin expression in amastigotes is low [13], but if a role for the proteinase in chronic leishmaniasis is established, leishmanolysin inhibitors may prove medically useful. We expect that the structure reported here will initially be used to design specific inhibitors that could be used to investigate the role of leishmanolysin in both insect and mammalian hosts.

Biological implications

We report here the structure of the leishmanolysin membrane-bound proteinase. The structure provides detailed information about the main protein surface antigen of the infective insect form of *Leishmania*, a parasitic protozoan which causes widespread disease in many parts of the world. Leishmanolysin is one of the best characterized of a number of key surface antigens from medically important pathogenic protozoa, a class of molecules for which there is as yet very little structural information. Although most of the leishmanolysin structure is novel, the region surrounding the active site is similar to the active-site region of metzincin class zinc proteinases. The best characterized examples of these proteinases are secreted extracellular matrix modifying enzymes, such as collagenase. To date, all metzincin proteinase sequences include a sequence motif containing three active-site histidine residues within a short stretch of 11 amino acids, HExxHxxGxxH. The leishmanolysin sequence contains only the first two histidine residues of the motif, and there was no prior reason to believe that leishmanolysin was a metzincin. The structure revealed that the third histidine was present, but separated in the sequence from the first two histidines by a 62-residue insert. These results demonstrate that the current 11-residue metzincin sequence motif can fail to identify metzincins and that extensive changes can occur in the immediate proximity of the metzincin active site, presumably while maintaining the geometry of the active site and proteolytic activity.

Leishmanolysin molecules on the surface of promastigotes probably form a significant part of the interface between the invading parasite cell and the mammalian host at the time of infection. Mammalian infection is a necessary step of the *Leishmania* life-cycle, and because comparatively few parasite cells are involved at this critical stage, there is likely to be strong selective pressure on leishmanolysin to maximize the probability of infection. Leishmanolysin might act before infection, ensuring maximum numbers of parasites are produced and released from the insect, or just after infection by facilitating binding and entry into target cells. The existence of a leishmanolysin homolog in *Crithidia fasciculata*, an insect parasite with no mammalian host, suggests a primary role in the insect. The structure should facilitate the development of non-toxic specific inhibitors required for experimental work to determine the function of leishmanolysin.

The medical emphasis of *Leishmania* research has meant that work has focused on human reactions to infection, and on vaccination and drug treatment. Leishmanolysin antibodies do occur in infected individuals and trials of vaccines containing leishmanolysin have been carried out, some with promising results. Reducing infection by stimulating immune responses to promastigotes is complicated by the fact that promastigotes target macrophages, cells which phagocytose microorganisms and process antigens in the development of the immune response. The structure may be used in future to design peptide vaccines that comprise just the appropriate antigenic parts of leishmanolysin, producing a specific protective response. Although the role of leishmanolysin in the amastigote macrophage form of *Leishmania* present in established infection is unknown, inhibitors may prove to be medically useful and the structure might be used together with existing zinc-proteinase inhibitors to develop therapeutic drugs.

Materials and methods

Crystallization, data collection and processing

Membrane-bound GPI-anchored leishmanolysin, isolated from *L. major* strain LRC-L119 stationary phase promastigotes, was solubilized by phospholipase C cleavage of the lipid part of the GPI anchor and purified as described [16]. Crystallization trials were carried out with the mature, glycosylated form of the protein, containing 478 amino acid residues (Val100–Asn577) and the carbohydrate part of the GPI anchor. Tetragonal (space group $P4_12_12$, $a = b = 63.6$ Å, $c = 251.4$ Å) and monoclinic (space group C2, $a = 107.2$ Å, $b = 90.6$ Å, $c = 70.6$ Å, $\beta = 110.6°$) crystals were obtained, grown under very similar crystallization conditions [17]. Diffraction data (Table 1) from tetragonal crystals (native; $(CH_3)_3PbAc$ 2) were recorded on a Siemens X100 multiwire area detector, and data from native monoclinic crystals (native 1) on a MAR Research image plate detector, both mounted on a rotating-anode X-ray source. Data were processed using XDS. A high-resolution native monoclinic data set (native 2), collected on the synchrotron beamline X11 at European Molecular Biology Laboratory–Deutsches Elektronen Synchrotron (EMBL–DESY), Hamburg, was measured on a MAR detector and processed using the software packages DENZO and SCALEPACK.

Phasing

The structure was solved using data from the two crystal forms (Table 2). Unless otherwise specified, programs used for the structure determination were from the CCP4 suite [35]) and the software package

Table 1

Crystallographic data.

Crystal	Resolution (Å)	Multiplicity	Completeness (%)	R_{merge} (%)*
Tetragonal crystals				
Native	2.7	4.7	91.8	8.0
$(CH_3)_3PbAc$ 1	2.85	4.9	99.2	7.5
$(CH_3)_3PbAc$ 2	3.8	3.1	77.1	14.9
Monoclinic crystals				
Native 1	2.8	4.2	87.6	5.3
Native 2	1.86	5.9	99.5	5.8
$(CH_3)_3PbAc$	2.5	4.0	99.0	8.6

*$R_{merge} = \Sigma_{hkl} \Sigma^{i} | I_{hkl,i} - \langle I \rangle_{hkl} | / \Sigma |\langle I \rangle_{hkl}| \times 100\%$. $(CH_3)_3PbAc$, trimethyllead acetate.

Table 2

Phasing statistics.

Derivative	No. of sites	R_{iso}*	d_{phas} (Å)†	R_{Cullis}‡	Phasing power§
Tetragonal crystals					
$(CH_3)_3PbAc$ 1	1	18.2	3.5	0.73, 0.74	0.7
$1.0(CH_3)_3PbAc$ 2	1	20.9	3.8	0.73, 0.67	0.7, 1.3
Monoclinic crystals					
$(CH_3)_3PbAc$	5	10.9	3.0	0.68, 0.78	0.8, 1.0

*$R_{iso} = \Sigma ||F_{PH}| - |F_P|| / \Sigma |F_P| \times 100\%$. †$d_{phas}$, maximum resolution of phased reflections. ‡$R_{Cullis} = \Sigma_{hkl} ||F_{PH} \pm F_P| - F_{H(calc)}| / \Sigma_{hkl} |F_{PH} - F_P|$. §Phasing power $= \Sigma_{hkl} |F_H| / \Sigma_{hkl} |$ lack of closure $|$. The pairs of values given for R_{Cullis} and phasing power are for the centric and acentric reflections, respectively. For the tetragonal crystals the figure of merit (FOM) within the resolution range 10–3.5 Å was calculated as 0.51 (for centric reflections) and 0.38 (for acentric reflections); the FOM values for the monoclinic crystals (resolution range 10–3.0 Å) were 0.56 (centric) and 0.30 (acentric).

developed in Uppsala [36]. Soaking experiments with the native crystals (stabilized in 60% (v/v) 2-methyl-2,4-pentanediol, 100 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$) provided a single trimethyllead acetate (25–30 mM, 3–4 weeks) derivative for both crystals forms. Tetragonal ($(CH_3)_3PbAc$ 1) and monoclinic ($(CH_3)_3PbAc$) derivative data were collected on the synchrotron beamline X31 at EMBL/DESY, Hamburg, at $\lambda = 0.92$ Å to optimize the anomalous scattering signal from lead. Data were recorded on a MAR detector and processed using MOSFLM and CCP4 programs. Scaling of heavy-atom derivative data and determination of heavy-atom positions was carried out using CCP4 programs. The correct enantiomorph $P4_{1(3)}2_12$ was assigned using anomalous scattering data from the trimethyllead acetate derivative and a poor independent mercury acetate derivative. In C2, the anomalous signal contribution was too weak to distinguish the hand and the less noisier electron-density map proved to be correct. Heavy-atom parameters were refined and SIRAS phases calculated using MLPHARE. Solvent-flattened maps in the monoclinic and tetragonal cells at 3.0 Å and 3.5 Å, respectively, showed clear molecular boundaries, but were not interpretable. The relationship between the maps from each crystal form was determined using AMORE with a 'bones atom' search model made from the 3.0 Å C2 map using the programs O and MAMA. The initial correlation coefficient for the electron density within the molecular boundary in the two maps was 0.23. A 30 step iterative averaging procedure (using the program RAVE), density modification (PRISM [37]), and phase combination (SIGMAA), was applied using data progressively extending from 3.5 to 2.8 Å resolution. Maps produced in this way were solvent flattened using DM. The procedure resulted in an interpretable map where the active-site helix and the zinc proteinase domain β sheet structure could be recognized.

Refinement

Models consisting at first largely of polyalanine were refined (using X-PLOR [38]) against the amplitudes and experimental phases of both the monoclinic and tetragonal data sets and $2F_o - F_c$ maps from each were averaged for further rebuilding. Refinement with the high-resolution synchrotron monoclinic data set to 1.86 Å (native 2) and rebuilding with density-modified maps (using DM) allowed completion of the model (Table 3). Waters were added using ARP software. All model building was performed using the program O. Overall, the structure

Table 3

Refinement statistics.

Resolution (Å)	3.36	2.81	2.51	2.30	2.15	2.04	1.94	1.86	5–1.86
R_{merge} (%)	4.5	6.9	8.2	9.5	10.3	12.3	15.3	21.4	5.8
R_{work} (%)*	15.4	18.5	20.2	21.0	21.5	22.5	23.7	25.8	19.5
R_{free} (%)†	19.1	18.3	20.3	21.8	22.9	23.1	24.6	27.2	21.0

*R_{work}, the crystallographic R factor calculated using 90% of all data in the resolution range used in refinement of models. †R_{free}, R factor calculated for a randomly selected 10% of data omitted in the model building and refinement process.

Table 4

Model parameters.

Number of nonhydrogen atoms	3600
Number of metal atoms	1
Number of amino acid residues	475
Number of water molecules	212
Rms deviation in nonhydrogen bond lengths (Å)	0.004
Rms deviation in nonhydrogen bond angles (°)	1.12

maintains very good stereochemical quality with the deviations from ideality of geometric parameters lying within the ranges expected for protein structures at this resolution (Table 4; as determined using PROCHECK). Three amino acid residues are in disallowed phi/psi regions in the Ramachandran plot: two (Thr499 and Ile502) belong to a poorly defined surface loop with high B factors and the third (Ile190), is part of a classic type I γ turn.

Figures

Skeleton representations were made with RIBBONS [39] or MOLSCRIPT [40] using secondary structure assignments from PROCHECK. The surface images were made with GRASP [41]. Figure 6 was prepared using DRAWTREE [42] from a multiple sequence alignment made using CLUSTALW [43].

Accession numbers

The coordinates of the leishmanolysin structure and high-resolution data set native 2 have been submitted to the Brookhaven Protein Data Bank (accession code 1LML).

Acknowledgements

We thank B Anderson and EMBL colleagues at Heidelberg and Hamburg for their help. We are also indebted to Jacques Mauel of the Institute of Biochemistry at the University of Lausanne, Switzerland for his help with the purification of leishmanolysin in his laboratory.

References

1. Desjeux, P. (1992). Human leishmaniases: epidemiology and public health aspects. *World Health Stat Q.* **45**, 267-275.
2. WHO. (1990). Control of the leishmaniases: report of a WHO Expert Committee. *WHO Technical Report Series* **793**.
3. Etges, R. (1992). Identification of a surface metalloproteinase on 13 species of *Leishmania* isolated from humans, *Crithidia fasciculata*, and *Herpetomonas samuelpessoai. Acta Tropica* **50**, 205-217.
4. Russell, D.G. & Wilhelm, H. (1986). The involvement of the major surface glycoprotein (gp63) of *Leishmania* promastigotes in attachment to macrophages. *J. Immunol.* **136**, 2613-2621.
5. Russell, D.G. (1987). The macrophage-attachment glycoprotein gp63 is the predominant C3-acceptor site on *Leishmania mexicana* promastigotes. *Eur. J. Biochem.* **164**, 213-221.
6. Puentes, S.M., Dwyer, D.M., Bates, P.A. & Joiner, K.A. (1989). Binding and release of C3 from *Leishmania donovani* promastigotes during incubation in normal human serum. *J. Immunol.* **143**, 3743-3749.
7. Connell, N.D., Medina-Acosta, E., McMaster, W.R., Bloom, B.R. & Russell, D.G. (1993). Effective immunization against cutaneous leishmaniasis with recombinant bacille Calmette-Guerin expressing the *Leishmania* surface proteinase gp63. *Proc. Natl Acad. Sci. USA* **90**, 11473-11477.
8. Schneider, P., Ferguson, M.A.J., McConville, M.J., Mehlert, A., Homans, S.W. & Bordier, C. (1990). Structure of the glycosylphosphatidylinositol membrane anchor of the *Leishmania major* promastigote surface protease. *J. Biol. Chem.* **265**, 16955-16964.
9. Etges, R., Bouvier, J. & Bordier, C. (1986). The major surface protein of *Leishmania* promastigotes is a protease. *J. Biol. Chem.* **261**, 9098-9101.
10. El-Sayed, N.M. & Donelson, J.E. (1997). African trypanosomes have differentially expressed genes encoding homologues of the *Leishmania* GP63 surface protease. *J. Biol. Chem.* **272**, 26742-26748.
11. McConville, M.J. & Ferguson, M.A. (1993). The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochem. J.* **294**, 305-324.
12. Warburg, A. & Schlein, Y. (1986). The effect of post-bloodmeal nutrition of *Phlebotomus papatasi* on the transmission of *Leishmania major. Am. J. Trop. Med. Hyg.* **35**, 926-930.
13. Schneider, P., Rosat, J.P., Bouvier, J., Louis, J. & Bordier, C. (1992). *Leishmania major*: differential regulation of the surface metalloprotease in amastigote and promastigote stages. *Exp. Parasitol.* **75**, 196-206.
14. Ilg, T., Harbecke, D. & Overath, P. (1993). The lysosomal gp63-related protein in *Leishmania mexicana* amastigotes is a soluble metalloproteinase with an acidic pH optimum. *FEBS Lett.* **327**, 103-107.
15. Schlagenhauf, E. (1995). *X-Ray Crystallographic Studies of Leishmanolysin, the Major Surface Metalloproteinase from* Leishmania major. PhD thesis, University of Heidelberg, Germany.
16. Bouvier, J., Schneider, P. & Etges, R. (1995). Leishmanolysin: surface metalloproteinase of *Leishmania. Methods Enzymol.* **248**, 614-633.
17. Schlagenhauf, E., Etges, R. & Metcalf, P. (1995). Crystallization and preliminary X-ray diffraction studies of leishmanolysin, the major surface metalloproteinase from *Leishmania major. Proteins* **22**, 55-66.
18. Freymann, D., Down, J., Carrington, I., Roditi, I., Turner, M. & Wiley, D. (1990). 2.9 Å Resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei. J. Mol. Biol.* **216**, 141-160.
19. Blum, M.J., Down, J.A., Gurnett, A.M., Carrington, M., Turner, M.J. & Wiley, D.C. (1993). A structural motif in the variant surface glycoproteins of *Trypanosoma brucei. Nature* **362**, 603-609.
20. Stöcker, W. & Bode, W. (1995). Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Curr. Opin. Struct. Biol.* **5**, 383-390.
21. Stöcker, W., *et al*., & Bode, W. (1995). The metzincins – topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* **4**, 823-840.
22. Bode, W., Gomis-Rüth, F.X., Huber, R., Zwilling, R. & Stöcker, W. (1992). Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358**, 164-167.
23. Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S. & Tschesche, H. (1994). The X-ray crystal structure of the catalytic domain of human neutrophil collagenase inhibited by a substrate analogue reveals the essentials for catalysis and specificity. *EMBO J.* **13**, 1263-1269.
24. Gomis-Ruth, F., *et al*., & Bode, W. (1997). Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **389**, 77-81.
25. Sussman, J.L., *et al*., & Silman, I. (1991). Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* **253**, 872-879.
26. Streit, J.A., Donelson, J.E., Agey, M.W. & Wilson, M.E. (1996). Developmental changes in the expression of *Leishmania chagasi* gp63 and heat shock protein in a human macrophage cell line. *Infect. Immun.* **64**, 1810-1818.
27. Button, L.L., Russell, D.G., Klein, H.L., Medina-Acosta, E., Karess, R.E. & McMaster, W.R. (1989). Genes encoding the major surface glycoprotein in *Leishmania* are tandemly linked at a single chromosomal locus and are constitutively transcribed. *Mol. Biochem. Parasitol.* **32**, 271-284.
28. Victoir, K., Dujardin, J.C., de Doncker, S., Barker, D.C., Arevalo, J., Hamers, R. & Le Ray, D. (1995). Plasticity of gp63 gene organization in *Leishmania* (Viannia) *braziliensis* and *Leishmania* (Viannia) *peruviana. Parasitology* **111**, 265-273.
29. Espinoza, J.R., *et al*., & Blackwell, J.M. (1995). Extensive polymorphism at the Gp63 locus in field isolates of *Leishmania peruviana. Mol. Biochem. Parasitol.* **72**, 203-213.
30. Medina-Acosta, E., Karess, R.E. & Russell, D.G. (1993). Structurally distinct genes for the surface protease of *Leishmania mexicana* are developmentally regulated. *Mol. Biochem. Parasitol.* **57**, 31-45.
31. Webb, J.R., Button, L.L. & McMaster, W.R. (1991). Heterogeneity of the genes encoding the major surface glycoprotein of *Leishmania donovani. Mol. Biochem. Parasitol.* **48**, 173-184.
32. Steinkraus, H.B., Greer, J.M., Stephenson, D.C. & Langer, P.J. (1993). Sequence heterogeneity and polymorphic gene arrangements of the *Leishmania guyanensis* gp63 genes. *Mol. Biochem. Parasitol.* **62**, 173-186.
33. Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**, 734-740.

34. Brittingham, A., Morrison, C.J., McMaster, W.R., McGwire, B.S., Chang, K.P. & Mosser, D.M. (1995). Role of the Leishmania surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *J. Immunol.* **155**, 3102-3111.
35. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760-763.
36. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110-119.
37. Bystroff, C., Baker, D., Fletterick, R.J. & Agard, D.A. (1993). PRISM: application to the solution of two protein structures. *Acta Cryst. D* **49**, 440-448.
38. Brünger, A.T. (1992). *X-PLOR, Version 3.1, a System for X-Ray Crystallography and NMR.* Yale University Press, New Haven, CT.
39. Carson, M. (1991). Ribbons 2.0. *J. Appl. Cryst.* **24**, 958-961.
40. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
41. Nicholls, A. Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamics properties of hydrocarbons. *Proteins* **11**, 281–296.
42. Felsenstein, J. (1993). PHYLIP (phylogeny inference package). *Cladistics* **5**, 164-166.
43. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
44. Grams, F., *et al.*, & Bode, W. (1995). X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors. Implications for substrate binding and rational drug design. *Eur. J. Biochem.* **228**, 830-841.
45. Devereux, J. (1995). Wisconsin Package Version 9.0. Genetics Computer Group (GCG), Madison, WI.