



Structural Mechanisms of Inactivation in Scabies Mite Serine Protease Paralogues

Katja Fischer^{1*†}, Christopher G. Langendorf^{2†}, James A. Irving^{2†}, Simone Reynolds^{1,4}, Charlene Willis¹, Simone Beckham¹, Ruby H. P. Law², Sundry Yang², Tanya A. Bashtannyk-Puhlovich², Sheena McGowan², James C. Whisstock^{2,3}, Robert N. Pike², David J. Kemp¹ and Ashley M. Buckle^{2,3*}

¹Scabies Laboratory,
Queensland Institute of Medical
Research, Brisbane, 4029,
Australia

²The Department of
Biochemistry and Molecular
Biology, School of Biomedical
Sciences, Faculty of Medicine,
Nursing and Health Sciences,
Monash University, Clayton,
Victoria 3800, Australia

³ARC Centre for Structural and
Functional Microbial Genomics,
Monash University, Clayton,
Victoria 3800, Australia

⁴The University of Queensland,
School of Veterinary Sciences,
St Lucia, Qld 4072, Australia

Received 22 January 2009;
received in revised form
28 April 2009;
accepted 30 April 2009
Available online
7 May 2009

The scabies mite (*Sarcoptes scabiei*) is a parasite responsible for major morbidity in disadvantaged communities and immuno-compromised patients worldwide. In addition to the physical discomfort caused by the disease, scabies infestations facilitate infection by Streptococcal species *via* skin lesions, resulting in a high prevalence of rheumatic fever/heart disease in affected communities. The scabies mite produces 33 proteins that are closely related to those in the dust mite group 3 allergen and belong to the S1-like protease family (chymotrypsin-like). However, all but one of these molecules contain mutations in the conserved active-site catalytic triad that are predicted to render them catalytically inactive. These molecules are thus termed scabies mite inactivated protease paralogues (SMIPPs). The precise function of SMIPPs is unclear; however, it has been suggested that these proteins might function by binding and protecting target substrates from cleavage by host immune proteases, thus preventing the host from mounting an effective immune challenge. In order to begin to understand the structural basis for SMIPP function, we solved the crystal structures of SMIPP-S-I1 and SMIPP-S-D1 at 1.85 Å and 2.0 Å resolution, respectively. Both structures adopt the characteristic serine protease fold, albeit with large structural variations over much of the molecule. In both structures, mutations in the catalytic triad together with occlusion of the S1 subsite by a conserved Tyr200 residue is predicted to block substrate ingress. Accordingly, we show that both proteases lack catalytic function. Attempts to restore function (*via* site-directed mutagenesis of catalytic residues as well as Tyr200) were unsuccessful. Taken together, these data suggest that SMIPPs have lost the ability to bind substrates in a classical “canonical” fashion, and instead have evolved alternative functions in the lifecycle of the scabies mite.

© 2009 Elsevier Ltd. All rights reserved.

Edited by R. Huber

Keywords: scabies; protease; inactivated catalytic triad

Introduction

Scabies is a parasitic infection of the skin caused by the burrowing of the ectoparasitic itch mite

Sarcoptes scabiei.^{1–3} The disease afflicts about 300 million people worldwide and is particularly problematic within socially disadvantaged populations. Infections are endemic in Aboriginal communities in Northern Australia that are characterized by inadequate medical facilities and overcrowding.^{4,5} Scabies is prevalent also in immuno-deficient patients and in residents of nursing homes, despite the availability of chemotherapy control measures.⁶ Infestation occurs when the adult female mite burrows into the skin. Pruritic scabies lesions are often accompanied by opportunistic streptococcal

*Corresponding authors. E-mail addresses:

Katja.Fischer@qimr.edu.au;

Ashley.Buckle@med.monash.edu.au.

† K.F., C.G.L. and J.A.I. contributed equally to this work.

Abbreviations used: SMIPP, scabies mite inactivated protease paralogue.

infections with significant sequelae (cellulitis, septicemia and glomerulonephritis) and the increased community streptococcal burden contributes to extreme levels of rheumatic fever and rheumatic heart disease.^{4,5} The wide prevalence and decreasing efficacy of treatment for scabies indicate that development of novel control strategies, in the form of vaccination or immunotherapy, are warranted; this is feasible as clinical immunity to scabies develops in humans.^{7,8} A comprehensive review of these aspects was published recently.⁹

The paucity of molecular data on scabies is due to the difficulty of obtaining mites in sufficient numbers, because the parasite burden is generally very low and no *in vitro* culture system for propagating mites is available. In order to address this problem, using scabies mites from skin shed into the bedding of severe crusted scabies patients, we recently created a sequence database and identified gene homologues to major allergens of house dust mites.^{9–11} In particular, 33 genes homologous to the serine protease group 3 major allergen were identified. Remarkably, with one exception, all proteins contain mutations in the catalytic triad. This suggests strongly that members of this family cannot act as proteases by a known mechanism, and they have accordingly been named scabies mite inactivated protease paralogues (SMIPPs).¹² Here, we present the result of a structural, biochemical and phylogenetic investigation of the SMIPPs. The crystal structures of two members reveal considerable structural rearrangements within and around the active site. This provides an explanation for the absence of protease function, and suggests that, in contrast to active serine proteases, these molecules are unable to bind substrates in a canonical fashion.¹³

Results

SMIPPs adopt the chymotrypsin-like serine protease fold

Both SMIPPs crystallised as a dimer in the asymmetric unit of the crystal, with distinct relative orientations. There is no experimental evidence, however, to support the presence or physiological relevance of the dimeric form, and therefore we conclude that the dimers are crystalline artefacts. Both structures adopt the chymotrypsin-like serine protease fold, with residues 12–229 and 1–235 visible in the electron density for SMIPP-S-I1 and SMIPP-S-D1, respectively. There was no significant difference between molecules related *via* non-crystallographic symmetry in either case, and the A chains were used in this analysis. Structure determination and refinement statistics are given in Table 1. Numbering refers to SMIPP-S-I1 starting at the N-terminus of the mature protein, unless stated otherwise. To aid comparison with other structures, chymotrypsin template numbering is shown in

Table 1. Data collection and refinement statistics

Data collection	SMIPP-S-I1	SMIPP-S-D1
Space group	$P2_12_12_1$	$P6_3$
Cell dimensions (Å)	$a=47.415$, $b=80.578$, $c=114.693$	$a=133.159$, $b=133.159$, $c=83.706$
Resolution (Å)	57.35–1.85	57.64–2.0
No. of unique observations	34,997	67,071
Multiplicity	4.0 (2.5)	7.0 (6.4)
Data completeness (%)	91.8 (72.2)	99.4 (96.7)
$\langle I/\sigma I \rangle$	7.1 (2.3)	16.4 (3.5)
R_{pim} (%) ^a	5.0 (26.2)	4.1 (22.6)
<i>Structure refinement</i>		
Non hydrogen atoms		
Protein	3486	3540
Solvent	347	310
R_{free} (%)	22.0	17.5
R_{cryst} (%)	18.7	15.7
r.m.s.d.'s from ideality		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	1.0	1.1
Ramachandran plot		
Favoured (%)	98.0	98.1
Outliers (%)	0	0
B -factors (Å ²)		
Mean main chain	32.2	33.6
Mean side chain	34.5	35.9
Mean water molecule	42.1	46.1
r.m.s.d. bonded B s		
Main chain	0.828	0.830
Side chain	3.1	2.8

Values in parentheses refer to the highest-resolution shell.

^a Agreement between intensities of repeated measurements of the same reflections, and can be defined as: $\sum (I_{h,i} - \langle I_h \rangle) / \sum I_{h,i}$, where $I_{h,i}$ are individual values and $\langle I_h \rangle$ is the mean value of the intensity of reflection h .

parentheses where appropriate. A detailed description of SMIPP nomenclature is provided as [Supplementary Data](#).

Structural diversity within the active site and throughout the molecules

We compared the structures of each of the SMIPPs with human trypsin (PDB identifier [5PTP](#)) (Fig. 1a and b). RMSD (% sequence identities) between SMIPP-S-I1 and SMIPP-S-D1, SMIPP-S-I1 and trypsin, and SMIPP-S-D1 and trypsin are 2.4 Å (40 %), 2.9 Å (24 %) and 1.8 Å (25 %), respectively. The net RMSD between all three structures is 2.4 Å. Both in comparison with each other and with trypsin, the SMIPPs display relatively large (~4 Å) backbone shifts extending throughout the structure, particularly in loop regions (residues 20–26, 131–140, and 201–206, corresponding to residues 35–40, 142–153 and 218–223 in chymotrypsin numbering). However, the core interiors of the molecules are relatively well conserved, and the positions of the catalytic triad (or equivalent residues) are maintained (Fig. 1a and b). The relatively large structural divergence between the two SMIPPs, and in comparison to trypsin, reflects the low level of sequence identity between the SMIPPs and with the mammalian enzyme (Fig. 1a). Indeed, the sequence divergence

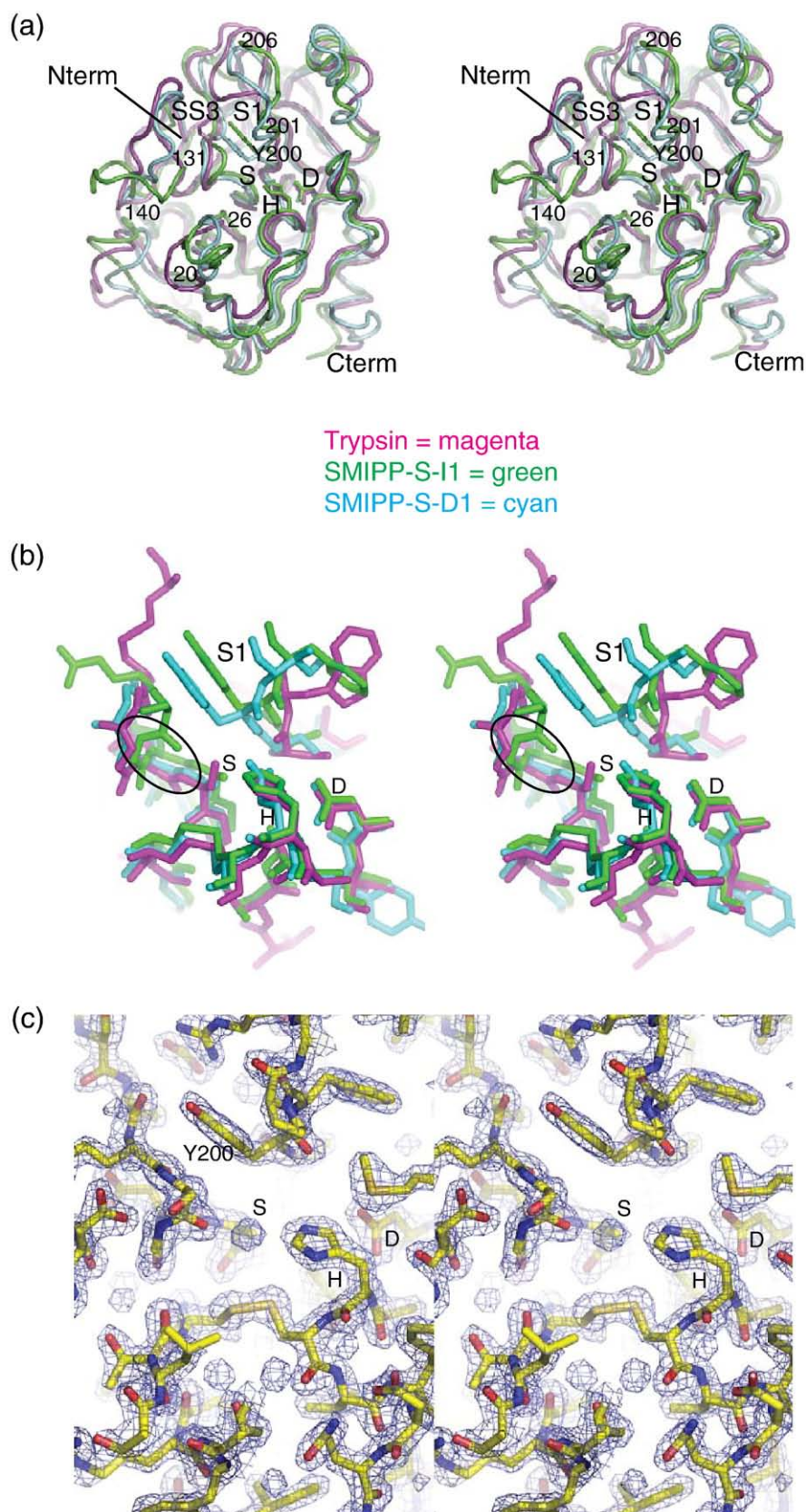


Fig. 1. Overall structure of SMIPPs and comparison with human trypsin. (a) Stereo cartoon representation showing the structural alignment of SMIPP-S-I1 (green), SMIPP-S-D1 (cyan) and trypsin (1PTP; magenta). Regions of structural differences are indicated by residue labels. The location of the disulphide bond (SS3) that is present in trypsin but absent from SMIPPs is indicated. The side chain of Y200 that blocks the S1 subsite is shown as stick bonds and labelled. (b) Structural differences in the active site, oxyanion (ellipsoid) hole and S1 subsite. (c) Electron density in active site, contoured at 1 σ , where σ is the rmsd of electron density in the unit cell.

within the SMIPP family is sufficiently large to imply that the majority of other SMIPPS display large, unpredictable structural variations outside the core of the molecule.

In both structures, the S1 subsite is blocked by a conserved tyrosine (Y200, 215 in chymotrypsin numbering), which prevents substrate access to this vital part of the active site (Fig. 1a). Structural shifts in this region are relatively large, particularly in the loop that contains Cys220 in trypsin (201–206, 218–223 in chymotrypsin numbering). Since all SMIPPs lack Cys191 and Cys220, which form a conserved disulphide bond in the serine protease family, it is likely that all display similarly large loop shifts resulting in a pronounced structural rearrangement at the S1 subsite.

Interestingly, although SMIPPs do not contain zymogen sequences,¹² predictions show signal peptidase cleavage at the site usually expected after cleavage of a zymogen, for every SMIPP known. In addition, they have several structural characteristics in common with the zymogen form. First, some of the regions showing structural differences correspond to the “zymogen activation domain” of chymotrypsinogen (i.e. the 147-loop, the 180-loop and the 220-loop, in chymotrypsin numbering). Second, whereas the oxyanion hole is properly formed in SMIPP-S-D1, it is slightly distorted in SMIPP-S-I1 (though this is relatively small compared to the disruption seen in chymotrypsinogen). Third, in SMIPP-S-I1 the N-terminus incorporation has not occurred, resulting in a zymogen-like conformation of the N-terminus, but not the “zymogen-triad” seen in chymotrypsinogen (this requires a greater rearrangement of the oxyanion hole).

Superimposing SMIPP structures with a trypsin-protein inhibitor complex (2PTC) reveals that the structural changes are so large that formation of a complex with a protein substrate would be impeded without extensive structural rearrangement (Fig. 2). It has been suggested that SMIPPs may function by binding and protecting target substrates from cleavage by host immune proteases, thus preventing the host from mounting an effective immune challenge.¹³ This hypothesis is clearly incompatible with our structural data, but nevertheless we were interested in testing whether SMIPPs are able to bind peptide substrates.

Peptide phage display screening did not identify binding partners

Phage-displayed peptide libraries have been used to identify peptide substrates of several chymotrypsin-like serine proteases.^{14,15} Since it was likely that the SMIPPs were inactive due to mutations of their catalytic residues, we investigated whether 20mer peptides fused to the minor coat protein pIII in a phage display library^{16,17} could bind to the SMIPPs. Screening of the phage display peptide library with three different SMIPPs did not demonstrate a preference for any of the amino acid sequences displayed. During successive rounds of panning,

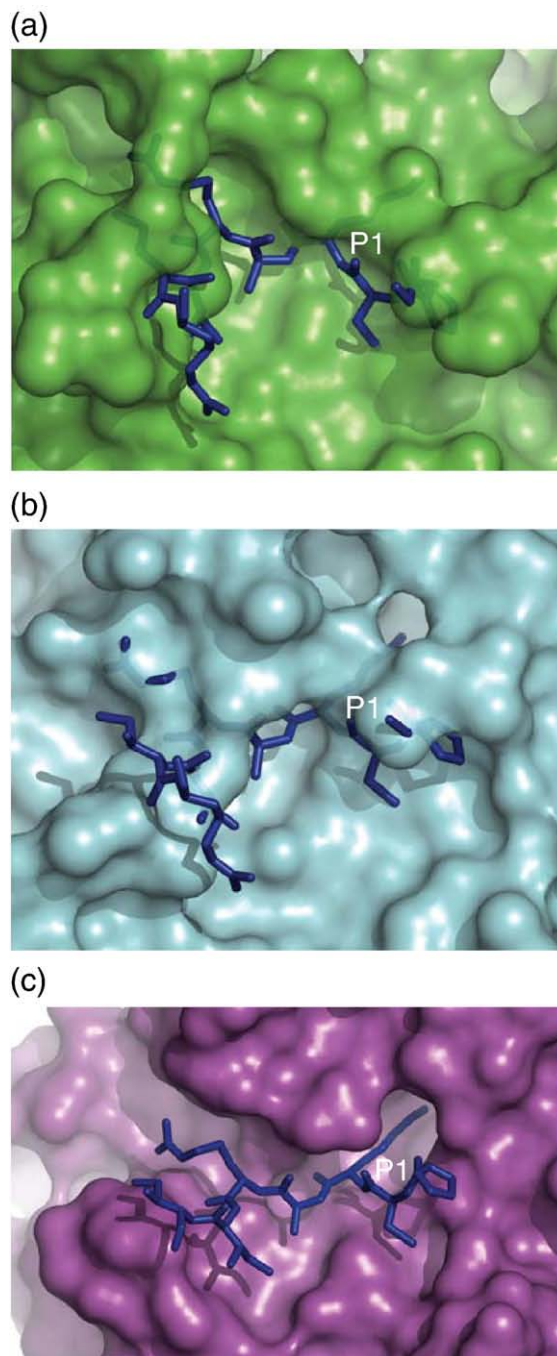


Fig. 2. SMIPP-protease inhibitor binding is blocked. (a and b) Surface representations of SMIPP-S-I1 and SMIPP-S-D1, respectively, showing modelled peptide (sticks) and protein (cartoon) trypsin inhibitors. (c) Interactions between trypsin and inhibitor (2PTC) shown for contrast.

titrations showed that the plaque-forming units/ml counts for a positive control known to select phage-displayed peptides from the library (5G8¹⁶) increased very quickly and remained high, while the plaque-forming units/ml of phage displaying peptides that bound to the SMIPP-S proteins did not increase (Supplementary Data Table 1A). While phage displayed peptides that have an affinity for binding to 5G8 increased in number during each

round of panning, the same was not found for the SMIPPs. The titration data were confirmed using an ELISA to detect phage that displayed peptides that bound to the protein of interest (Supplementary Data Table 1B). These results are consistent with the structural data described above, implying strongly that SMIPPs do not bind peptides as might be expected for a catalytically inactive protease with an otherwise fully formed active site.

Restoration of the catalytic triad by mutagenesis does not rescue protease activity

Initial enzyme assays performed with SMIPP-Ss I1, G2 and B2 on methylcoumarin peptide substrates (AAPF, AAF, LLVY, LY, VKM, R, GGR, GPR and APA) revealed no proteolytic activity for wild type proteins or for mutants with their glycosylation sites removed (typically by mutation of Asn to Gln).

In order to restore the catalytic triad as well as unblock the S1 subsite, we created the mutant Q41H/N85D/A186S/Y202A of the SMIPP-S-D1 mature sequence and the mutant A184S/Y200A of the SMIPP-S-I1 mature sequence. Consistent with structural data, neither mutant displayed protease activity on the range of substrates described above, suggesting that the absence of protease activity is not due solely to these mutations and is probably a result of structural rearrangements in the oxyanion hole and S1 subsites and, at least in the case of SMIPP-S-I1, these “zymogen-like” characteristics appear to have a role in the failure to reactivate SMIPPs by mutagenesis.

Taken together with the structural data, our findings suggest that SMIPPs have diverged significantly from the ancestral serine protease, resulting in an irreversible culmination of mutations in and around the active site that result in extensive structural rearrangements, effectively preventing the binding of substrates (Figs. 1 and 2).

Phylogenetic analysis of the SMIPP family

A codon-based alignment of the SMIPP sequences was divided, with reference to the crystal structure of SMIPP-S-I1, into three functionally distinct regions: surface-exposed active site residues; surface-exposed residues outside the active site; and buried residues. The rates of synonymous mutation (a polymorphism yielding an identical amino acid) *versus* non-synonymous mutation (causing an amino acid substitution) were inferred, and are shown in Supplementary Data Table 2. The values reveal that surface residues outside the active site are subject to significant ($p < 0.001$) purifying (negative) selection, suggesting that evolutionary pressure is operating to maintain the functional and/or structural integrity of the SMIPPs as a whole. In contrast, residues within the active site appear to be undergoing neutral drift and gradual diversification. This indicates that, on average, selective pressure has been lost within the active site, and therefore this region is likely to be of a markedly different importance to the function of the

inactive SMIPPs than for active trypsin-like serine proteases.

The phylogeny of the SMIPP family (Supplementary Data Fig. 1a) is consistent with the early duplication of a gene encoding an active enzyme, yielding the modern-day SARS3-A1 enzyme and the common ancestor of all non-enzymatic members of the family. Following this duplication, a significant lag period is observed, followed by two duplications within a short period of time, and subsequently a further series of expansions, yielding the modern-day repertoire of the parasite. Maximum likelihood ancestral sequence reconstruction, based on codon data, was used to infer sequences at ancestral nodes with the highest posterior probability in PAML4.1. This analysis suggests that the common ancestor of the inactive SMIPPs lost its enzyme activity before the expansion of the non-enzymatic subfamily. Consistently, none of the SMIPPs appears to possess a sequence that is intermediate between the enzymatically active and inactive forms.

Other tests performed on the data indicated that there is significant heterogeneity in the rate of evolution between lineages ($p < 0.01$), reflecting differences in selective pressure operating on the different genes. Consequently, the data are not clock-like and it is not possible to infer actual dates for the divergence events. Nevertheless, it was found that, over the course of their evolution, the inactive SMIPPs have accumulated mutations in their amino acid sequence 1.7-fold, on average, more rapidly than their enzymatic counterpart. Collectively, these observations support the notion that the SMIPPs have acquired distinct roles that have, in turn, altered the selective pressures to which they are subject.

Discussion

Although the proteolytically active member of the family, SARS3-A1, exhibits the hallmarks of an active serine protease, the unique sequence and structural features of the SMIPP family discussed above suggest strongly that this family functions in a way that is distinctly different from that of a catalytically active serine protease. Given the clear sequence homology with house dust mite serine protease allergens, and the knowledge that scabies infestation elicits a delayed-type hypersensitivity reaction, it is likely that SARS3-A1 and SMIPPs have a role in the defence against the host immunological response.¹² Although it is plausible that the sequence diversity within the SMIPP family represents an adaptation that enables scabies mites to evade the host defence mechanisms, our structural data do not support the hypothesis that this is in any way mediated by their capacity to competitively bind peptide substrates, at least by a canonical protease-like mechanism. In particular, Tyr200 is anticipated to occlude the S1 subsite completely. In active serine proteases, proper interaction of the P1 residue of substrates¹⁸ with the S1 site of the

protease is a primary determinant of specificity. We therefore postulate that SMIPPs do not function by competing for substrates with host proteases that have a role in host defence mechanisms. It is likely that all members of the SMIPP family have lost the ability to bind substrates by any known mechanism but, instead, have evolved alternative binding

activities. To investigate alternative binding sites, we analysed the molecular surface properties of each molecule. Consistent with their large structural difference, the shape and electrostatic properties of their active sites are strikingly dissimilar (Fig. 3a and b). Although we have demonstrated that the structural changes observed in SMIPP-S-I1 preclude

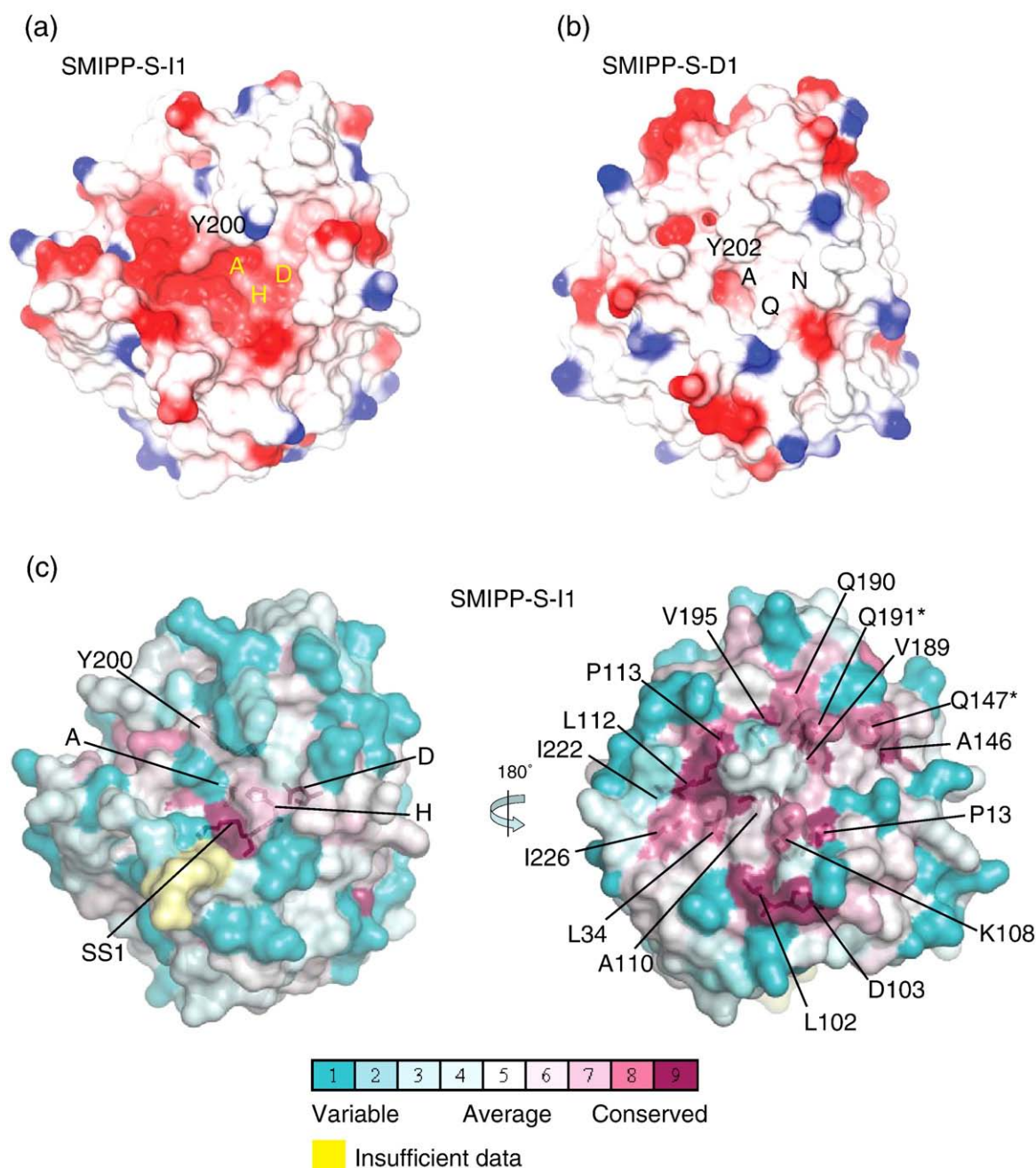


Fig. 3. Molecular surfaces of: (a) SMIPP-S-I1 and (b) SMIPP-S-D1 colour-coded according to the electrostatic potential (red, negative; blue, positive). (c) Sequence conservation of all 33 SMIPP family members mapped onto the molecular surface of SMIPP-S-I1. Left and right views are orthogonal. Highly conserved surface-exposed residues, as well as catalytic triad residues are shown as sticks and labelled. Predicted N-glycosylation sites are labelled with an asterisk (*). (d) Sequence conservation mapped onto SMIPP-S-D1, as for c. (e) Structural superposition of SMIPP-S-I1 and SMIPP-S-D1, both colour-coded according to sequence conservation. Colour coding is shown in the inset. Catalytic triad residues, highly conserved surface-exposed residues, disulphide bonds (and location of absent SS3 bond), and Y200 residue are shown as sticks. Left and right views are orthogonal. Views looking at the active site show molecules in similar orientations.

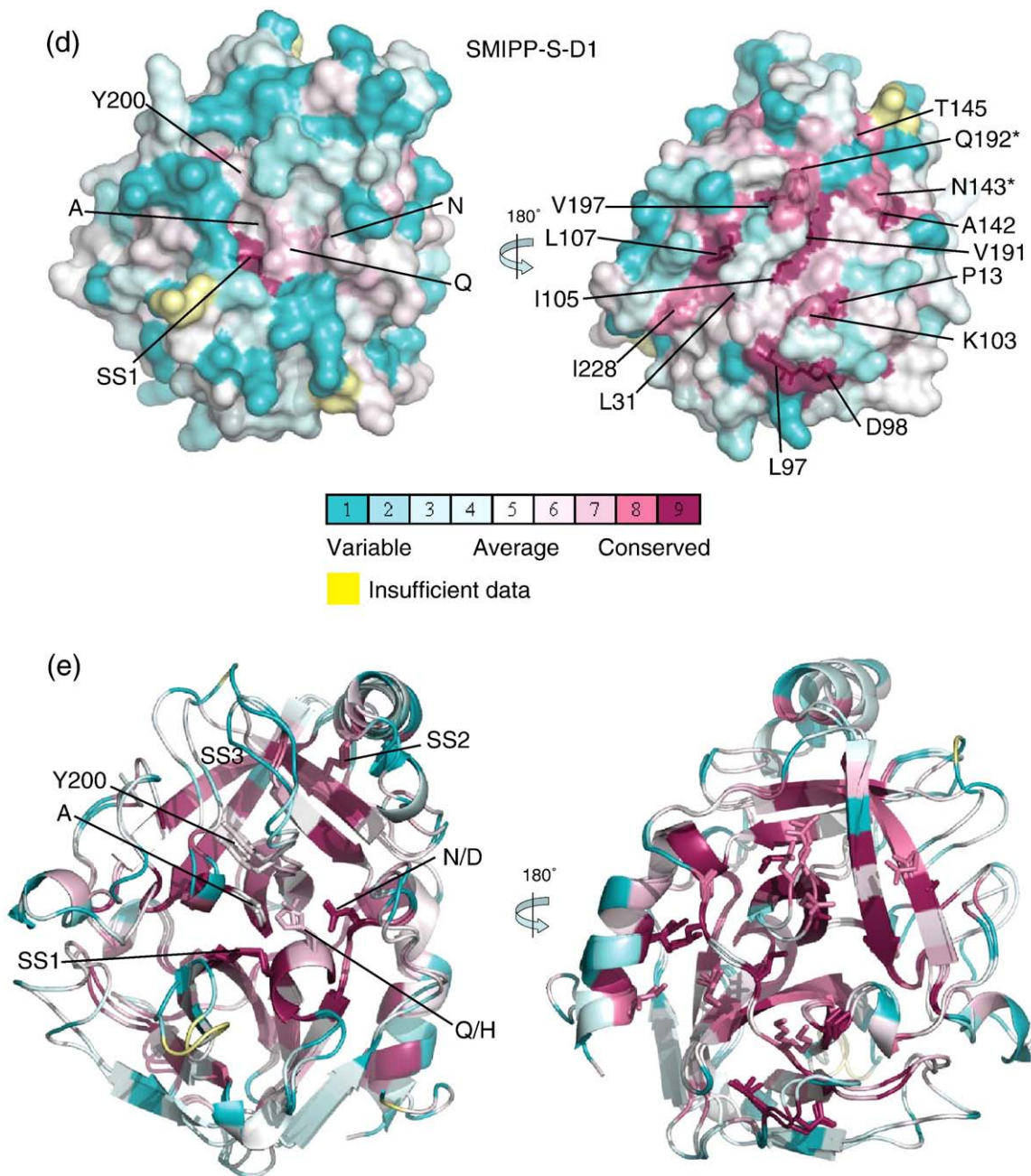


Fig. 3 (legend on previous page)

peptide or protein substrate binding in a serine protease-like fashion (Fig. 2), its "active site" still resembles a deep groove (Fig. 3a). Further, the surface of the groove is highly negatively charged (with the charge extending beyond the groove), suggestive of a binding function. In contrast, the equivalent region in SMIPP-S-D1 is relatively shallow and, with the exception of two negatively charged small pockets, is relatively neutral.

In a search for further clues to function, we mapped the sequence conservation within the SMIPP family onto both structures (Fig. 3c–e). It has been shown that sequence conservation of surface residues is a robust indicator of functional sites.^{19,20} The low level of sequence conservation

within the SMIPP family is revealed clearly: the highest level of conservation is found within the core of the molecule (with the exception of the catalytic triad; Fig. 3c and d), whereas the surrounding, more surface-exposed, regions show little, if any, conservation. This is consistent with the structural differences between the SMIPPs (Fig. 3e); regions with lower levels of sequence conservation show the largest structural shifts. Interestingly, there are small patches of conserved regions on the face of the molecule opposite to that of the active site (Fig. 3c–e). These are in regions of good agreement between the SMIPP structures, suggesting that they may represent potential binding sites or exosites. All but one of the predicted N-glycosylation sites for

both SMIPPs lie within conserved regions on this face, and this is a plausible explanation for the conservation at these sites (Fig. 3c and d).

Analysis of the electrostatic and physical surface characteristics in the context of the sequence diversity within the SMIPP family, suggests that it is highly unlikely that the observed features are general to the SMIPP family as a whole, and indeed may represent specific adaptations important to the individual and specialized roles of each SMIPP in the scabies mite lifecycle. Our data suggest strongly that scabies mites have adapted the protease fold to create a structurally and functionally diverse family of molecules necessary for evading the multifaceted immune system of the host. Accordingly, a more complete understanding of the role of SMIPPs in scabies pathogenicity may emerge only from further structural and functional studies of other members of the SMIPP family. Indeed, the interaction of SMIPPs with specific components of the complement system demonstrated recently, results in inhibition of the activation of this system.²¹

Materials and Methods

Expression of SMIPP protein in *Pichia pastoris* and purification

SMIPP-S-I1 and SMIPP-S-D1 were produced in *Pichia pastoris*. To direct secretion of the expressed protein into the medium, all constructs placed the predicted N-terminus flush with the Kex2 signal peptide cleavage site in the transfer vector pPICZalphaA (Invitrogen Australia Pty Limited, VIC, Australia). A stop codon at the 3' end of the SMIPP sequence was introduced to ensure that the recombinant protein did not include any of the C-terminal tags of the vector. The plasmid was propagated in *Escherichia coli* strain TOP10F', linearised with SacI and electroporated into *P. pastoris* KM71H (Invitrogen Australia Pty Limited, VIC, Australia). Media for colony growth and biomass or induction cultures were made according to the manufacturer's instructions. Cells were initially plated onto supported nitrocellulose on agar plates containing 100 µg/mL zeocin (Invitrogen). After incubation at 29 °C for 24 h, filters were moved onto YPDS medium containing 2000 µg/mL zeocin for selection of transformants with increased numbers of SMIPP gene copies. To ensure pure clonal isolates, single colonies were plated twice on two subsequent YPD/zeocin agar plates. Transformants were assessed to select for the highest expressing clone following the manufacturer's instructions (Invitrogen Australia Pty Limited, VIC, Australia). Biomass cultures were incubated at 29 °C for 24 h and induction cultures were incubated for 60–70 h at 29 °C with shaking. SMIPP expression was detected on a Western blot using mouse antiserum generated against purified SMIPP, produced in *E. coli*. Correct processing of the recombinant protein at the Kex2 cleavage site was confirmed after purification by N-terminal sequencing in the Peptide Biology Laboratory, Monash University, Clayton, Victoria (data not shown).

Mature SMIPP protein secreted from *P. pastoris* was purified as follows: the expression culture supernatant was collected following centrifugation at 12,000 g for 45 min at 4 °C and stored at –80 °C. The supernatant was

thawed and prepared for hydrophobic interaction chromatography by the gradual addition of (NH₄)₂SO₄ to a final concentration of 1.5 M and adjustment of the pH to 5.0. The preparation was centrifuged for 30 min at 20,000 g. The supernatant was filtered through a 0.45 µm pore size filter and the filtrate was loaded onto a 5 mL HiTrap phenyl Sepharose column (GE Healthcare Bio-Sciences Pty. Ltd., NSW, Australia), pre-equilibrated with 1.5 M (NH₄)₂SO₄, 10 mM Mes, 10% (v/v) glycerol, pH 5.0. Unbound material was removed by washing with equilibration buffer. Protein was eluted with a stepwise gradient of 1.0 M – 0 M (NH₄)₂SO₄ (50 mL each of 1 M, 0.5 M and 0 M (NH₄)₂SO₄). Fractions were subjected to dialysis against 50 mM Mes, 10% (v/v) glycerol, pH 5.0 and subjected to SDS-PAGE (10% (w/v) polyacrylamide gel). Fractions containing SMIPP protein were pooled, loaded onto a 5 mL HiTrap SP Sepharose FF column (GE Healthcare Bio-Sciences Pty. Ltd., NSW, Australia) equilibrated in 50 mM Mes, pH 5.0, following which the column was washed with 50 mL of the same buffer to remove unbound material. A linear gradient of 0 M – 0.4 M NaCl over 30 mL was applied to the column. To monitor the purification, eluted fractions were subjected to SDS-PAGE (10% (w/v) polyacrylamide gel) and monitored by visualization with Coomassie brilliant blue R-250 and Western blot analysis. On average, 0.9 – 1.3 mg of pure protein was obtained per 200 mL of culture.

Panning a random peptide library on SMIPPs

A linear peptide library of 20 random amino acids displayed as N-terminal fusions to the minor coat protein pIII of filamentous phage M13 was a kind gift from Professor Mick Foley, La Trobe University, Melbourne. The library of >5 × 10⁸ random peptides was screened for SMIPP-S-binding peptides as described.¹⁶ As a positive control for the experiments reported here, the library was screened with a monoclonal antibody, 5G8, generated against a linear epitope of the *Plasmodium falciparum* apical membrane antigen-1 (AMA1), previously shown to bind phage from this library.¹⁶ Briefly, 1 µg of each protein (four SMIPP-Ss and 5G8, respectively) was used to coat the wells of 96-well ELISA plates overnight at 4 °C. Each well was washed twice with 200 µL of PBS and blocked with 200 µL of 5% (v/v) skimmed milk/PBS for 2–3 h at room temperature. It was ascertained by ELISA using polyclonal antibodies raised against each SMIPP-S that the SMIPP-S protein bound to the microtitre plate. Phage were pre-incubated in 1% skimmed milk/PBS for 15 min. Wells were washed twice with 200 µL of PBS and then 10¹² pre-incubated phage in 100 µL of pre-incubation buffer was added to the wells and the plate was incubated for 2–3 h at room temperature. Non-adherent phages were removed by two washes with PBS, 0.05% (v/v) Tween 20. Adherent phage were eluted by the addition of 100 µL of 0.1 M glycine (pH 2.2) to each well and incubated for 10 min. The eluted phages were neutralised with 50 µL of 1.5 M Tris. Phages were titrated and amplified in *E. coli* K91 using procedures similar to those described previously.²² Phage ELISAs were performed by coating a microtiter plate (Nunc Maxisorp) with 1 µg of protein overnight at 4 °C and subsequently blocked with 5% skimmed milk for 2 h. Phage dilutions (100 µL) were prepared in PBS, transferred in duplicate to the coated blocked wells and incubated for 1 h with shaking. The wells were washed five times with PBS, 0.05% (v/v) Tween 20, and 100 µL of anti-M13 antibody conjugated to horseradish peroxidase (Amersham Pharmacia Biotech, Sweden) at 1:5000 (v/v)

dilution with PBS, 0.05% (v/v) Tween 20 was added to each well. After incubation for a further 1 h and washing as described above, bound phage were detected with 100 μ L of ABTS solution (Boehringer Mannheim, Germany) following the manufacturer's instructions. Absorption at 405 nm was read on the POLARstar OPTIMA using FLUOstar OPTIMA software (BMG Labtech, Australia).

Enzyme kinetics

Enzyme assays were performed in assay buffer (100 mM Tris-HCl, 10 mM CaCl₂ and 0.05% (v/v) polyethylene glycol (PEG) 8000, pH 8.2) with 10 μ M commercially available methylcoumarin substrates (Sigma Aldrich) and 25 μ L of SMIPP-S protein maximally concentrated to 100–450 μ M. Assays were done with a broad spectrum of serine protease substrates: AAPF (*N*-succinyl-Ala-Ala-Pro-Phe-7-amino-4-methylcoumarin), AAF (*N*-succinyl-Ala-Ala-Phe-7-amino-4-methylcoumarin), LLVY (*N*-succinyl-Leu-Leu-Val-Tyr-7-amino-4-methylcoumarin), LY (*N*-succinyl-Leu-Tyr-7-amino-4-methylcoumarin) and VKM (*N*-succinyl-Val-Lys-Met-7-amino-4-methylcoumarin) were used to detect chymotrypsin-like protease activities, while R (Z-Arg-7-amino-4-methylcoumarin), GGR (Z-Gly-Gly-Arg-7-amino-4-methylcoumarin) and GPR (Z-Gly-Pro-Arg-7-amino-4-methylcoumarin) were used for trypsin-like and APA (Z-Ala-Pro-Ala-7-amino-4-methylcoumarin) for elastase-like activity. Fluorescence was measured continuously at 37 °C in a final volume of 100 μ L in 96-well plates on a POLARstar fluorescent microtitre plate reader, (BMG Labtech, Melbourne, Australia) at 30 s intervals for 30 min at excitation and emission wavelengths of 370 nm and 460 nm, respectively.

Mutagenesis of N-glycosylation sites

When subjected to SDS-PAGE, each SMIPP produced a major band slightly larger than the same molecule produced in *E. coli* plus a smear of larger material. SMIPP-S-I1 and SMIPP-S-D1 have three and two predicted N-glycosylation sites, respectively. Crystallization can be hindered by glycosylation; therefore, we chose to mutate the predicted glycosylation sites. Mutation of these from N to Q at positions 84, 147, and 191 of the mature protein of SMIPP-S-I1 and at position 143 of the mature protein of SMIPP-S-D1 eliminated the smear of larger material (data not shown).

Crystallization, X-ray data collection, structure determination and refinement

Before crystallization, purified enzymes were concentrated to 5–10 mg/mL. The crystals were grown using the hanging-drop, vapour-diffusion method, with 1:1 (v/v) ratio of protein to mother liquor (well volume 0.5 mL). The crystals appeared overnight in 22% (v/v) PEG 8000, 10% (v/v) glycerol, 0.1 M Tris-HCl (pH 8.5) and 0.2 M magnesium chloride and reached full size in three days. Crystals were dehydrated against reservoir buffer with 15% (v/v) glycerol for 16 h. Crystals were equilibrated for 5 min in reservoir buffer in the presence of 20% (v/v) glycerol. The crystals were flash-frozen in liquid N₂ before data collection with 15% glycerol as the cryoprotectant.

Data were collected from cryo-cooled crystals at 100 K at the BIOCARS and IMCA-CAT beamlines at the Advanced Photon Source, Chicago, USA. Diffraction images were processed using MOSFLM²³ and SCALA²⁴ from the CCP4 suite.²⁵ Randomly chosen 5% of each

dataset was flagged for calculation of R_{free} ²⁶ with neither a σ nor a low-resolution cut-off applied to the data.

Structure solution preceded using the molecular replacement method and the program PHASER.²⁷ A search model was constructed from the crystal structure of granzyme K (PDB 1FI8), the closest structural homologue identified using the FFAS server.²⁸ A "mixed" model consisting of conserved sidechains (all other non-alanine/glycine residues truncated at the C γ atom) was then created using the SCRWL server.²⁸ Two clear peaks in both the rotation and translation functions were evident, and these packed well within the unit cell. Together with the unbiased features in the initial electron density maps, the correctness of the molecular replacement solution was confirmed.

Initial structure refinement and model building proceeded using one molecule in the asymmetric unit (with the other non-crystallographic symmetry-related molecule generated using non-crystallographic symmetry operators). Maximum likelihood refinement using REFMAC,²⁹ incorporating translation, liberation and screw-rotation displacement refinement was carried out, using a bulk solvent correction (Babinet model with mask). Non-crystallographic symmetry restraints were used throughout refinement to improve the observation to parameter ratio. Imposed restraints were guided by manual inspection of the model and R_{free} . Simulated annealing composite omit maps were generated using CNS³⁰ omitting 5 % of the model. All model building and structural validation was done using COOT.³¹ Water molecules were added to the model using ARP/wARP³² when the R_{free} reached 25%. Solvent molecules were retained only if they had acceptable hydrogen bonding geometry contacts of 2.5 – 3.5 Å with protein atoms or with existing solvent and were in good $2F_o - F_c$ and $F_o - F_c$ electron density. A summary of statistics is provided in Table 1.

Structural analysis

Structures were superimposed using the program MUSTANG.³¹ Disordered surface exposed residues were modelled using the repair a model function of the WHAT-IF server $\ddagger$, this step is important for subsequent electrostatic and sequence conservation calculations. Electrostatic potential molecular surfaces were created using CCP4MG (calculated by the Poisson-Boltzmann solver within CCP4MG).^{33,34} The probe radius used was 1.4 Å. Sequence conservation was mapped onto the structures using the CONSURF Server.^{35,36} Pymol $\S$ was used to produce all other structural representations.

Protein Data Bank accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 3H7O (SMIPP-S-I1) and 3H7T (SMIPP-S-D1)).

Acknowledgements

A.M.B is a National Health and Medical Research Council of Australia Senior Research Fellow. J.C.W.

$\ddagger$ <http://swift.cmbi.kun.nl/WIWWWI/>

$\S$ <http://pymol.sourceforge.net/>

is an Australian Research Council Federation Fellow and honorary NHMRC Principal Research Fellow. D. J. K. is an NHMRC Senior Principal Research Fellow. J.A.I. is an NHMRC CJ Martin Postdoctoral Fellow. This work was supported by the NHMRC, the Australian Research Council and the State Government of Victoria (Australia). We thank the staff at IMCA-CAT (The Advanced Photon Source, Chicago) for technical assistance. We thank Professor Mick Foley and Dr Joanne Casey from La Trobe University, Melbourne for providing the phage display peptide library and expert advice.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2009.04.082](https://doi.org/10.1016/j.jmb.2009.04.082)

References

- Arlian, L. G. (1989). Biology, host relations, and epidemiology of *Sarcoptes scabiei*. *Annu. Rev. Entomol.* **34**, 139–161.
- Elgart, M. L. (1990). Scabies. *Dermatol. Clin.* **8**, 253–263.
- Green, M. S. (1989). Epidemiology of scabies. *Epidemiol. Rev.* **11**, 126–150.
- Currie, B. J. & Carapetis, J. R. (2000). Skin infections and infestations in Aboriginal communities in northern Australia. *Australas J. Dermatol.* **41**, 139–143.
- Currie, B. J., Harumal, P., McKinnon, M. & Walton, S. F. (2004). First documentation of in vivo and in vitro ivermectin resistance in *Sarcoptes scabiei*. *Clin. Infect. Dis.* **39**, e8–e12.
- Taplin, D., Porcelain, S. L., Meinking, T. L., Athey, R. L., Chen, J. A., Castillero, P. M. & Sanchez, R. (1991). Community control of scabies: a model based on use of permethrin cream. *Lancet*, **337**, 1016–1018.
- Arlian, L. G., Rapp, C. M. & Morgan, M. S. (1995). Resistance and immune response in scabies-infested hosts immunized with *Dermatophagoides* mites. *Am. J. Trop. Med. Hyg.* **52**, 539–545.
- Mellanby, K. (1944). The development of symptoms, parasitic infection and immunity in human scabies. *Adv. Parasitol.* **35**, 197–206.
- Walton, S. F., Holt, D. C., Currie, B. J. & Kemp, D. J. (2004). Scabies: new future for a neglected disease. *Adv. Parasitol.* **57**, 309–376.
- Dougall, A., Holt, D. C., Fischer, K., Currie, B. J., Kemp, D. J. & Walton, S. F. (2005). Identification and characterization of *Sarcoptes scabiei* and *Dermatophagoides pteronyssinus* glutathione S-transferases: implication as a potential major allergen in crusted scabies. *Am. J. Trop. Med. Hyg.* **73**, 977–984.
- Harumal, P., Morgan, M., Walton, S. F., Holt, D. C., Rode, J., Arlian, L. G. *et al.* (2003). Identification of a homologue of a house dust mite allergen in a cDNA library from *Sarcoptes scabiei* var. *hominis* and evaluation of its vaccine potential in a rabbit/*S. scabiei* var. *canis* model. *Am. J. Trop. Med. Hyg.* **68**, 54–60.
- Holt, D. C., Fischer, K., Allen, G. E., Wilson, D., Wilson, P., Slade, R. *et al.* (2003). Mechanisms for a novel immune evasion strategy in the scabies mite *Sarcoptes scabiei*: a multigene family of inactivated serine proteases. *J. Invest. Dermatol.* **121**, 1419–1424.
- Hubbard, S. J., Campbell, S. F. & Thornton, J. M. (1991). Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.* **220**, 507–530.
- Deperthes, D. (2002). Phage display substrate: a blind method for determining protease specificity. *Biol. Chem.* **383**, 1107–1112.
- Hekim, C., Leinonen, J., Narvanen, A., Koistinen, H., Zhu, L., Koivunen, E. *et al.* (2006). Novel peptide inhibitors of human kallikrein 2. *J. Biol. Chem.* **281**, 12555–12560.
- Coley, A. M., Campanale, N. V., Casey, J. L., Hodder, A. N., Crewther, P. E., Anders, R. F. *et al.* (2001). Rapid and precise epitope mapping of monoclonal antibodies against *Plasmodium falciparum* AMA1 by combined phage display of fragments and random peptides. *Protein Eng.* **14**, 691–698.
- Casey, J. L., Coley, A. M., Anders, R. F., Murphy, J., Humberstone, K. S., Thomas, A. W. & Foley, M. (2004). Antibodies to malaria peptide mimics inhibit *Plasmodium falciparum* invasion of erythrocytes. *Infect. Immun.* **72**, 1126–1134.
- Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Armon, A., Graur, D. & Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* **307**, 447–463.
- Bell, R. E. & Ben-Tal, N. (2003). In silico identification of functional protein interfaces. *Comput. Funct. Genomics*, **4**, 420–423.
- Bergström, F. C., Reynolds, S., Johnstone, M., Pike, R. N., Buckle, A. M., Kemp, D. J. *et al.* (2009). Scabies mite inactivated serine protease paralogues inhibit the human complement system. *J. Immunol.* **182**, 7809–7817.
- Adda, C. G., Tilley, L., Anders, R. F. & Foley, M. (1999). Isolation of peptides that mimic epitopes on a malarial antigen from random peptide libraries displayed on phage. *Infect. Immun.* **67**, 4679–4688.
- Leslie, A. G. W. (1992). *Joint CCP4 + ESF-EAMCB Newsletter Protein Crystallogr.* **26**.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D*, **62**, 72–82.
- CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Brunger, A. T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr. D*, **49**, 24–36.
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr. D*, **61**, 458–464.
- Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. (2005). FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res.* **33**, W284–W288.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

31. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.
32. Morris, R. J., Perrakis, A. & Lamzin, S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244.
33. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D*, **59**, 1131–1137.
34. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P. *et al.* (2004). Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D*, **60**, 2288–2294.
35. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302.
36. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.