House dust mites possess a polymorphic, single domain putative peptidoglycan D,L endopeptidase belonging to the NlpC/P60 Superfamily

Vivian H. Tang, Geoffrey A. Stewart, Barbara J. Chang

Microbiology & Immunology, School of Pathology and Laboratory Medicine, The University of Western Australia, Crawley, WA, Australia

A R T I C L E  I N F O

Article history:
Received 2 June 2015
Revised 8 September 2015
Accepted 10 September 2015

Keywords:
Mites
γ-δ-glutamyl-ε-diamino acid endopeptidase
NlpC/P60 Superfamily
LytFM homologues
Peptidoglycan
Structural modelling

A B S T R A C T

A 14 kDa protein homologous to the γ-δ-glutamyl-ε-diamino acid endopeptidase members of the NlpC/P60 Superfamily has been described in *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae* but it is not clear whether other species produce homologues. Bioinformatics revealed homologous genes in other Sarcoptiformes mite species (*Poroptes ovis* and *Blomia tropicalis*) but not in *Tetranychus urticae* and *Metaseiulus occidentalis*. The degrees of identity (similarity) between the *D. pteronyssinus* mature protein and those from *D. farinae*, *P. ovis* and *B. tropicalis* were 82% (96%), 77% (93%) and 61% (82%), respectively. Phylogenetic studies showed the mite proteins were monophyletic and shared a common ancestor with both actinomycetes and ascomycetes. The gene encoding the *D. pteronyssinus* protein was polymorphic and intronless in contrast to that reported for *D. farinae*. Homology studies suggest that the mite, ascomycete and actinomycete proteins are involved in the catalysis of stem peptide attached to peptidoglycan. The finding of a gene encoding a P60 family member in the *D. pteronyssinus* genome together with the presence of a bacterial promotor suggests an evolutionary link to one or more prokaryotic endosymbionts.

© 2015 The Authors. Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Mites secrete a number of enzymes in their fecal pellets including proteases, carboxydrases and bacteriolytic enzymes, which reflect their digestive processes. With regard to the latter, we have characterised a 14 kDa protein [1] from the house dust mite (HDM) *Dermatophagoides pteronyssinus* which originally showed significant sequence homology with a variety of bacterial proteins such as the P60 proteins from *Listeria monocytogenes* [1,2] and the invasion proteins from mycobacterial species [1]. Such proteins were subsequently shown to be cysteine peptidases, with a catalytic triad similar to that seen in papain [3] (MEROPS Clan CO, Family C40) [4] and this activity is associated with the NlpC/P60 Pfam domain.

Proteins in this family cleave between the second and third residues of the peptidoglycan stem peptide to release the peptide L-Ala-γ-δ-Glu and diaminopimelic acid (DAP) or Lys [5]. In bacteria, they are autolysins that may be cytoplasmic or secreted, and they play an important role in septum formation since gene deletion results in long, connected strings of bacteria [6–8]. However, they may also play a role in eukaryotic cell invasion and, in humans, the Listerial P60 protein induces both B and T cell responses, which help protect against infection [9]. Proteins possessing the NlpC/P60 domain may be single or multidomain proteins and, in the latter, other domains may play a role in targeting the enzyme to bacterial cell walls [10–13], or providing them with additional peptidoglycan cleaving ability [14]. The majority of proteins possessing the NlpC/P60 domain described in the literature are prokaryotic in origin, with only a relatively small number being described thus far in eukaryotes.

The mite 14 kDa protein is a single domain protein and the gene encoding it and a variant have now been designated *lytFM* and *lytFM1*, respectively, based on the terminology used to describe the autolysins from *Bacillus subtilis* [15,16]. Whilst the role of NlpC/P60 family members in bacterial species in peptidoglycan processing is more clearly delineated, that of the proteins in mites and other eukaryotes is yet to be established. As they do not
possess peptidoglycan, it is reasonable to assume that possession of NlpC/P60 proteins and other peptidoglycan-degrading proteins such as lysozyme and amidas are likely to enable eukaryotes to protect themselves against pathogenic bacteria, or to utilise bacteria as a food source or both.

Mites represent one of the most ubiquitous groups of invertebrates, and they can be found in a diverse array of ecological niches. However, it is not clear whether all species of mites possess the NlpC/P60 protein described in the genus Dermatophagoides. In addition, it is possible that the gene encoding the mite NlpC/P60 protein with its relatively high sequence homology to bacterial proteins, entered into the mite genome from mite-associated microorganisms. In this regard, both Gram-positive and Gram-negative bacterial species have been demonstrated in mites by both conventional culture techniques and microbiome studies [17–19], that are capable of producing P60 family members, for example, Bacillus subtilis, Bacillus licheniformis and Bacillus cereus. Whether D. pteronyssinus mites acquired the lytFM gene via horizontal gene transfer either as a single domain protein or one encoding a multi-domain protein, which then lost genetic material encoding non-NlpC/P60 domains, is unclear.

In this report, we have searched for lytFM homologues in other mite species using a bioinformatics approach. Similarly, we looked at flanking sequences of lytFM by PCR to determine whether there were remnants of a bacterial heritage, and used gDNA libraries to look for introns. Finally, we have modelled the mite P60 family members to gain information about the possible preferred

2. Materials and methods

All PCR primers were obtained from Invitrogen (Mulgrave, Victoria, Australia). The D. pteronyssinus gDNA library was kindly provided by Dr. Wendy-Anne Smith (Telethon Kids Institute, Perth, Western Australia), respectively. PCR reagents used were HotStarTaq Master Mix (QIAGEN, Hilden, Germany), KAPA HiFi™ PCR Kit (KAPA Biosystems, Boston, USA), Taq buffer (10×, containing 15 mM MgCl2) (Perkin Elmer, New Jersey, USA), the four 2’-deoxynucleotide 5'-triphosphates (Promega, Wisconsin, USA) and Taq DNA polymerase (Ampli Taq Gold) (Applied Biosystems, Victoria, Australia).

2.1. D. pteronyssinus gDNA library and Dermatophagoides farinae cDNA library

Briefly, the library was prepared using purified D. pteronyssinus genomic DNA, digested with Sau 3A I 7–20 kb fragments and packaged into Xho I. The library was propagated and the phage adsorbed to Escherichia coli ER1647 cells at a ratio of 4 × 10⁷ pfu/ml host cells at 37 °C for 30 min. Plaques were allowed to reach semi-confluence (5–6 h at 37 °C) and the phage eluted with SM buffer. The gDNA library was quantified as described previously [20] before being used in PCR.

2.2. PCR primers

The annealing sites of all the primers used in PCR screening of the D. pteronyssinus gDNA library are shown in Fig. 1. The primers included those used previously to amplify lytFM from the D. pteronyssinus cDNA library [1], i.e., GSUTR1 (5’-CATGATATGTTTATGCTATTACCATATTGACCG-3’) and GSR3 (5’-TTTCAACATCGTCAACTTACCC-3’), primers designed to anneal to the right and left arms of the λBlueS-TAR™ vector VP1 (5’-GCATGATATGTTTATGCTATTACCATATTGACCG-3’) and VP2 (5’-GAAAAATAAATGAAATGGGTTCCG-3’), and the 5’ and 3’ ends of lytFM GSP1 (5’-GCTTTAGACAAACGTGAAAGC-3’) and GSP2 (5’-TGTATTGAGCTTGTTTTATGGTCCGG-3’), and GSP2R (5’-GCTTTAGACAAACGTGAAAGC-3’), and GSP1F (5’-GCTTTAGACAAACGTGAAAGC-3’) and FSPR (5’-GCTTTAGACAAACGTGAAAGC-3’) which

Fig. 1. Schematic diagram showing the annealing sites for all the primers used in the PCR screening of the λBlueS-TAR™ gDNA library of D. pteronyssinus. The arrows and numbers above the primers indicate the directions of amplification and the nucleotide positions, respectively, spanned by the primers within the λBlueS-TAR vector, lytFM and 179 bp-3’ sequence. Since the Sau3A I-digested genomic DNA fragments of D. pteronyssinus could be ligated to the Xho I-digested vector arms in either direction, the lytFM gene could exist in orientation 1 (A) or inverted (B).
was designed to anneal to a 179 bp-3’ sequence downstream of lytFM (Fig. 1).

2.3. PCR screening of the *D. pteronyssinus* gDNA library

PCR to amplify lytFM from the gDNA library was performed with the HotStarTaq Master Mix kit under the conditions: 95 °C for 15 min, 94 °C for 1 min, 37 °C for 45 s, 68 °C for 1 min and 72 °C for 10 min for 35 cycles. PCR screening of the gDNA library to elucidate the sequences flanking the 5’ and 3’ ends of lytFM was performed using the HotStarTaq Master Mix kit at the annealing gradient temperatures of 29.6, 29.9, 31.2, 33.4, 36.6, 38.6, 39.8 and 40.1 °C (29–40 °C) or 40, 41.4, 43.6, 46.9, 51.4, 54.6, 56.8 and 58 °C (40–58 °C). For PCR performed using the KAPA HiFi™ PCR Kit, the conditions applied were 95 °C for 5 min, 98 °C for 20 s, 29–40 °C or 40–58 °C for 45 s, 72 °C for 2.5 min and 72 °C for 5 min for 30 cycles. For PCR performed using the conventional reagents, the samples prepared contained 1× Taq buffer (containing 1.5 mM MgCl2), 0.5–5.5 mM supplementary MgCl2, 0.1 µM forward and reverse primers, 200 µM of each of the four dNTPs, 2.5 units of AmpliTaq Gold and 100 ng of template to a total volume of 25 µl, and the PCR conditions applied were 95 °C for 15 min, 94 °C for 1 min, 37 °C for 45 s, 68 °C for 1 min and 72 °C for 10 min for 30 cycles. All amplifications were performed in a Bi-RAD iCycler IQ PCR Thermal Cycler (96 wells) (Bio-RAD, Hercules, California, USA).

2.4. Preparation of PCR products for DNA sequencing and analysis of sequencing data

DNA sequencing was performed by the Lotterywest State Biomedical Facility Genomics at the Royal Perth Hospital, Perth, Australia. The sequencing data were compared with sequences in the NCBI nucleotide database or the European Nucleotide Archive using BLASTN or translated into amino acid sequences using the ExPaSy translation tool (University of Geneva, Switzerland) and the sequences compared with those in the protein and translated nucleotide sequence databases of NCBI or EBI using BLASTP and TBLASTN, respectively.

2.5. Bioinformatic analyses

Various searches of the available databases were undertaken to determine whether the presence of the lytFM sequence had been previously reported in other mite species. In addition, appropriate nucleotide sequences were analysed for the presence of prokaryotic promoters using the Neural Network Promoter Prediction (NPPP) program at the Berkeley Drosophila Genome Project (BDGP) (http://fruitfly.org/seq_tools/promoter.html) and the program Softberry BPROM (Softberry, Inc., Mount Kisco, New York). The latter program predicts bacterial sigma 70 promoter sequences with about 80% accuracy and specificity (http://linux1.softberry.com/berry.phtml). The likely presence of leader sequences was determined using the eukaryote non-TM data set and SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), and the theoretical isoelectric points (pI) determined using the ProtParam tool at ExPaSy (www.expsay.org). PHYRE 2.0 was used to model the structure of the mite protein [21].

Phylogenetic analysis of the mite proteins, and bacterial and eukaryotic homologues were performed using the Phylogeny.fr pipeline [22], with multiple sequence alignment performed using MUSCLE [23] after using the MaxAlign software (www.cbs.dtu.dk) and manual curation [24]. Sequences possessing the invariant Tyr and the Cys, His and a polar residue comprising the catalytic triad [5] were included. The approximate likelihood ratio test was used for branches instead of bootstrapping as described [25].

3. Results

3.1. The lytFM gene in *D. pteronyssinus* is polymorphic

A search was performed to ascertain whether lytFM or its variant, lytFM1, were curried in any of nucleotide databases. Three *D. pteronyssinus* cDNA clones (ESTs, GenBank: EX163010, EX163553 and EX162735) were identified. In addition, one of these (GenBank: EX163010) was found to contain a 179 bp 3’ flanking sequence (data not shown). A number of nucleotide polymorphisms were detected in all 5 sequences (Table 1), most of which were silent. However, polymorphisms in the 3’ region of the gene result in amino acid substitutions giving rise to another variant (LytFM2) in addition to the two originally described. The changes are predicted to alter the pl of the proteins; 8.57 for LytFM, and 8.92 for the remaining four sequences. A *D. pteronyssinus* gDNA library was then screened for the presence of lytFM variants with PCR conditions previously used for the successful amplification of the gene from the *D. pteronyssinus* cDNA library [1]. However, both primer pairs used, GSUTR1/GSR3 and GS2/GSR3 (Fig. 1) resulted in amplification of the lytFM isoform, but not the lytFM1 or lytFM2 isoforms (data not shown). In addition, the PCR product obtained was found to be intronless, as the sequence obtained was identical to that derived from the cDNA library (data not shown).

3.2. lytFM homologues in other mite species

A partial [26] and full-length sequence [19] of the lytFM homologue from *D. farinae* have recently been described. In this regard, an intron has been identified in the lytFM homologue of *D. farinae*, in contrast to the intronless gene reported here for *D. pteronyssinus*. The intron is between the sequence encoding the first 16 residues of the leader sequence and the sequence encoding the next four residues (Fig. 2) using the reported unplaced genomic scaffold (KN266412). A search of the available databases identified lytFM homologues in *P. ovis* (FR749374) and *B. tropicalis* (CB282085). An alignment of the deduced amino acid sequences for the four

| Table 1 |
|---|
| Polymorphisms found in the lytFM gene of *D. pteronyssinus*. |
| Sequence | aNucleotide position (amino acid residue change) |
| 12 | 69 | 72 | 105 | 138 | 165 | 396 | 414 | 415 | 417 | 418 | 420 |
| LytFM | C | C | T | T | T | C | T(N) | A | T(W) | G | A(S) | T |
| LytFM1 | C | C | T | T | T | C | G(K) | A | G(G) | A | G(G) | A |
| bEX163553 | – | T | G | A | T | T | G(K) | C | T(W) | G | A(S) | T |
| bEX163010 | T | C | T | A | A | C | G(K) | C | T(W) | G | A(S) | T |
| bEX162735 | T | A | G | A | T | T | A(K) | A | T(W) | G | A(S) | T |

a Nucleotide position based on LytFM sequence.

b Sequences obtained from the European Nucleotide Archive.

c Protein with this polymorphism designated LytFM2.

V.H. Tang et al. / FEBS Open Bio 5 (2015) 813–823 815
mote species of LytFM homologues is shown in Fig. 2, where significant homology was detected across the mature proteins compared to LytFM for the mite species, D. farinae (82% identity, 96% similarity), P. ovis (77% identity, 93% similarity) and B. tropicalis (61% identity, 82% similarity). The percent GC contents of the nucleotide sequences were 38.9, 39.3, 38.5, 36.7, 58, and 81% for D. pteronyssinus, D. farinae, P. ovis, B. tropicalis, and D. pteronyssinus homologues, respectively. Two conservative polymorphisms were detected when the full-length [19] and partial [26] D. farinae sequences were compared, namely Lys for Arg at residue position 29 and Ile for Leu at position 45 (mature D. pteronyssinus numbering). Fig. 2 shows that most signal peptide cleavage sites were predicted for P. ovis, D. pteronyssinus and D. farinae homologues (Fig. 2) but not for B. tropicalis. The cleavage positions predicted were between 20 and 21 for P. ovis and D. pteronyssinus, and between 16 and 17 for D. pteronyssinus using the eukaryotic data set. However, the cleavage position for D. pteronyssinus LytFM was previously determined to be between residues 20 and 21 (Fig. 2) using N-terminal sequencing (1), which coincided with that predicted using the Gram-negative data set. Using these data, the predicted pl’s for the mature D. pteronyssinus, P. ovis and B. tropicalis LytFM homologues were 9.08, 9.55 and 8.86, respectively. A search of genomic data obtained for two other mite species, namely the herbivorous mite D. pteronyssinus and B. tropicalis respectively. A search of genomic data obtained for two other mite species, D. pteronyssinus (GenBank: KN2666412), P. ovis (GenBank: FR749374) and B. tropicalis (GenBank: CB282085) were translated into amino acid sequences using the ExPaSy translation tool and the alignment was performed using MAFI[34]. Conserved, consensus residues are highlighted in boxshades and chemically similar substitutions (:) and weakly similar substitutions (.) are marked as indicated. The predicted (SignalP 4.1) or demonstrated N-terminal residues are shown (+1)[1]. The open triangle indicates the exon–intron boundary identified in the leader sequence of the D. farinae homologue [19]. The catalytic triad residues are indicated by open arrows and the residues comprising the S1 and S’ site are indicated by solid arrows and dots, respectively [5].

3.3. Phylogenetic analysis of mite LytFM homologues

A BLAST search of the mite LytFM homologue against eukaryote and prokaryote sequences was performed. With regard to eukaryotes, a significant proportion of the available sequences were from the ascomycetes although homologues were detected in wasps, ants, trichomonads, protozoa and amoebae. Fig. 3 shows the phylogenetic relationship between the homologues and it can be seen that the mite LytFM proteins form a monophyletic group. The mites and some ascomycetes and actinomycetes share a common ancestor. The mite homologues and its homologues from D. farinae (GenBank: KN2666412), P. ovis (GenBank: FR749374) and B. tropicalis (GenBank: CB282085) were translated into amino acid sequences using the ExPaSy translation tool and the alignment was performed using MAFI[34]. Conserved, consensus residues are highlighted in boxshades and chemically similar substitutions (:) and weakly similar substitutions (.) are marked as indicated. The predicted (SignalP 4.1) or demonstrated N-terminal residues are shown (+1)[1]. The open triangle indicates the exon–intron boundary identified in the leader sequence of the D. farinae homologue [19]. The catalytic triad residues are indicated by open arrows and the residues comprising the S1 and S’ site are indicated by solid arrows and dots, respectively [5].

3.4. The mite LytFM proteins are predicted to be peptidoglycan D,L endopeptidases

Members of the P60 family possess either γ-D-glutamyl-L-diamino acid or γ-D-glutamyl-L-lysyl endopeptidase activity [5]. Recently, the P60 protein from B. cereus (BcYkFC) has been crystallised with the bound reaction product, L-Ala-γ-D-Glu, thus establishing the residues involved in catalysis [5]. Fig. 2 shows that most of the residues comprising the substrate binding and catalytic site are present in the homologous mite proteins, indicating they are endopeptidases. However, some specific residues contributing to the S2 site of the binding cleft in the B. cereus protein are missing from the mite homologues; these being residues directly interacting with the free amino group of the L-Ala residue in the stem peptide. These include Asp256 (BcYkFC numbering), Glu83 and Tyr118.
with the latter two found within the SH3b domain (data not shown) rather than in the P60 domain [5]. All of the proteins shown in Fig. 3 share the catalytic residues involved in cleavage and only one, namely, the protein from *Penicillium oxalicum* has a corresponding residue to the Asp256, although it does not possess a SH3 domain.

The availability of several three dimensional structures of NlpC/P60 Superfamily members has enabled the structure of the mite proteins to be modelled using PHYRE 2.0. For each mite protein, the highest-ranking model was obtained using the NlpC/P60 domains of *Mycobacterium tuberculosis* RipA and RipB [27]. Each of the homologues was modelled with 97% coverage and 100% confidence and Fig. 5A and B show cartoon and surface models respectively, for the *D. pteronyssinus* homologue. Fig. 5C shows the *D. pteronyssinus* model superimposed on the *B. cereus* BcYkFC domain template, highlighting the location of a relatively large insertion between positions 22 and 38 in the mature *D. pteronyssinus* LytFM sequence in the N-terminal region seen in some actinomycete and ascomycete proteins, but absent from some Firmicutes and from other eukaryotic sequences.

3.5. PCR analysis of a *D. pteronyssinus* gDNA library reveals the presence of a prokaryotic promoter, 3′ to the lytFM gene

A reverse primer, FSPR based on the 179 bp 3′ sequence of the *D. pteronyssinus* LytFM cDNA clone EX163010 was used in conjunction with a lytFM forward primer, GSP2 to amplify nucleotide sequences flanking the 3′ end of lytFM in the *D. pteronyssinus* gDNA library (Fig. 1). An amplicon (GenBank: KT595670) was obtained and shown to be completely homologous with the 179 bp 3′ sequence (data not shown) indicating that it is a consistent feature and that no intron splicing occurs in the region adjacent to the 3′ end of the gene (data not shown). Although the 179 bp-3′ sequence did not match any sequence in the nucleotide or protein databases, the

**Fig. 3.** Homology between the mite LytFM homologues and selected eukaryote and prokaryote NlpC/P60 proteins. The sequences of mature LytFM (*D. pteronyssinus*) and its homologues *D. farinae* (GenBank: KN266412), *P. ovis* (GenBank: FR749374) and *B. tropicalis* (GenBank: CE82086) were compared with eukaryote and prokaryote NlpC/P60 proteins and homology determined using MAFFT [34]. The alignments were then shaded using boxshade (http://www.ch.embnet.org/software/BOX_form.html). Accession numbers are those cited in Fig. 4. The Asp256 residue in the *B. cereus* BCYKFC and its equivalent in the *P. oxalicum* protein that directly interact with the free amino group of the \(\gamma\)-Ala residue in the stem peptide during catalytic cleavage [5] are highlighted in red.
bacterial promoter prediction softwares indicated one between nucleotides 76 and 125 (NNPP score of 0.97, threshold 0.80) and another between nucleotides 97 and 129 (BPROM score 34, threshold 0.2) (data not shown).

PCR screening of the D. pteronyssinus gDNA library reveals homologies between the 5′ and 3′ flanking sequences of the lytFM gene with both eukaryotic and prokaryotic sequences, respectively. Attempts to identify sequences 5′ and 3′ to the lytFM gene were undertaken to determine possible ancestry. PCR was performed using a forward vector sequence primer, VP1 and a reverse lytFM gene sequence primer, GSP1 to amplify sequences upstream from the start of lytFM (Fig. 1). An amplicon was obtained that contained both vector and lytFM sequences (data not shown), as well as a sequence (GenBank: AKG95503) with homology to eukaryotic and prokaryotic hydroxysteroid dehydrogenases (Table 2). A similar screening approach using a lytFM forward primer, GS2 in conjunction with the reverse vector primer, VP2 (Fig. 1) amplified sequences downstream from the 179 bp-3′ sequence, some of which (GenBank: KT959671 and KT959672) showed high sequence homology with a hypothetical gene in Escherichia coli (Table 3).

4. Discussion

Homologues and variants of the lytFM gene have now been shown to be present in D. pteronyssinus, D. farinae, P. ovis, and B. tropicalis but not in T. urticae or M. occidentalis. With regard to D. pteronyssinus, five cDNA sequences have now been reported, and a comparative analysis indicates that the lytFM gene is polymorphic. Although most of the polymorphisms are silent, those in the 3′ region of the gene alter the pI of the LytFM1 and LytFM2 isoforms. In addition to identifying polymorphisms in D. pteronyssinus, a conservative polymorphism in the D. farinae homologue was also identified using the reported unplaced genomic scaffold (KN266412) [19]. These findings are consistent with those seen with other mite proteins, in particular the cysteine protease, trypsin and the MD-related group 2 protein [28,29].

This study also showed that there were no introns in the D. pteronyssinus lytFM gene which contrasts with that seen in D. farinae [19], where a single intron was observed in the leader sequence of the gene rather than in the region encoding the mature protein. In addition, the D. pteronyssinus, D. farinae and P. ovis
LytFM proteins are secreted, given the presence of a leader sequences. The predicted cleavage site for *D. pteronyssinus*, however, did not correspond with that observed experimentally [1] using the eukaryotic data set but did so with the Gram-negative data set. In contrast, the *B. tropicalis* protein did not appear to possess a leader sequence but confirmatory sequence data are required before it can be concluded that the latter is a cytoplasmic protein cleaving murein peptides intracellularly.

The *lytFM* gene is present in Sarcoptiformes mites occupying different ecological niches. In this regard, both *Dermatophagoides* spp are widely distributed in temperate climates whereas *B. tropicalis* is associated with the tropics. In contrast, *P. ovis* is a parasitic mite infesting the skin of animals such as rabbits and sheep. Interestingly, a NlpC/P60 homologue is absent from the genome of the hematophagous tick *Ixodes scapularis* [4], and is yet to be described in the human skin-dwelling mite *Sarcoptes scabei*.

Phylogenetically, the greatest similarity was found to occur between the two *Dermatophagoides* spp and *P. ovis*, rather than with *B. tropicalis* and this is consistent with the relationships observed using homologues of the mite cysteine protease allergens (e.g. Der p 1) and the MD-related allergens (e.g. Der p 2) (data not shown). However, the degree of similarity between the *D. pteronyssinus* LytFM protein and the *B. tropicalis* homologue (61% identity) is higher than that seen with the *B. tropicalis* homologues of Der p 1 (35%) and Der p 2 (42%). The mite proteins showed some homology with homologues from other eukaryotes, and share a common ancestor with the actinomycetes and ascomycetes, with the Firmicutes, trichomonads, wasp, fruit fly, moths and protozoa being more distantly related.

Several of the eukaryote proteins including the mite LytFM homologues are single domain proteins as revealed by a Pfam search ([5,27,30] data not shown), although three proteins (*Aspergillus fumigatus* GI 846917091 and GI 70997053, and *Neosartorya udagawae*, GI 849274716) were shown to possess a SH3_3 domain and one protein was a multidomain protein (Chemotaxis protein LafT, *Beauveria bassiana* GI 701777570). In addition, disulphide bond arrangement in the mite homologues is clearly different from that likely to occur in the ascomycetes. Here, there were distinct differences with cysteines being absent from the actinomycetes and a number of arrangements for the ascomycetes particularly in a C terminal insertion absent from the mite sequences.

The availability of domain and enzymatic substrate data ([5,27,30]) enables deductions to be made regarding the possible peptidoglycan substrates susceptible to the mite enzymes as well
Fig. 4. Phylogenetic relationships of mite LytFM homologues and selected eukaryote and prokaryote NlpC/P60 proteins. The phylogenetic tree was created using Phylogeny.fr pipeline. The approximate likelihood ratio test was used and the confidence values shown on the branches.
as the actinomycetes and other eukaryotes. In this regard, structural data show that the \textit{B. cereus} BcYkfC protein and catalytic residues are highlighted in magenta and yellow, respectively. The bacillus protein comprises both the P60 domain as well as two SH3b domains, one of which contributes to the S2 site, as well as sterically restricting access to larger substrates containing peptidoglycan [5]. The first two residues constituting the catalytic triad (Cys and His) are invariant throughout the sequences including the mite proteins but the third residue varies slightly between members, and may be His, Glu, Asn, Gln. The Asp interacting with the \(\gamma\)-D-Glu residue of the stem peptide in the \textit{B. cereus} P60 domain is invariant in the mite, ascomycete and actinomycete homologues whereas the Asp\textsubscript{256} residue in

![Fig. 5. Model of the mite LytFM protein. (A) A model of \textit{D. pteronyssinus} LytFM generated by PHYRE2, using \textit{M. tuberculosis} RipA (PDB: 2XIV) and \textit{B. cereus} BcYkfC (PDB: 3H41). The 13-residue insertion compared to the \textit{B. cereus} BcYkfC protein and catalytic residues are highlighted in magenta and yellow, respectively. (B) A surface representation of the LytFM showing the putative substrate-binding groove with the Cys\textsubscript{42}, His\textsubscript{93} and Asn\textsubscript{105} catalytic residues revealed and shown in magenta, red and blue, respectively. The 13-residue insertion found in LytFM and the \textit{M. tuberculosis} protein is highlighted in pink. (C) The LytFM model (green) is superimposed on the \textit{B. cereus} BcYkfC template (light brown) to highlight the relative location of the 13-residue insertion (magenta) in the former. A 4-residue deletion in the mite LytFM homologues and \textit{M. tuberculosis} protein relative to the \textit{B. cereus} BcYkfC is shown in blue. The catalytic residues are highlighted in yellow.]

### Table 2

| GenBank accession number | Amino acid sequences bearing homology                                                                 | Query coverage (%) | E value | Maximum identity (%) |
|--------------------------|------------------------------------------------------------------------------------------------------|--------------------|---------|----------------------|
| KFM77175                 | Short-chain dehydrogenase reductase family 4E member from \textit{Stegodyphus mimosarum} (spider)     | 97                 | 3e−22   | 49                   |
| XP_002434360             | Putative 3 hydroxysteroid dehydrogenase from \textit{Ixodes scapularis} (tick)                      | 96                 | 1e−17   | 42                   |
| XP_003743458             | Predicted 3 BETA-hydroxysteroid dehydrogenase/delta 5 → 4-isomerase type 7-like hydroxysteroid dehydrogenase from \textit{Metaseiulus occidentalis} (mite) | 96                 | 4e−15   | 39                   |
| XP_003743458             | 3 beta-hydroxysteroid dehydrogenase/delta 5 → 4-isomerase type 1 from \textit{Zootermopsis nevadensis} (dampwood termite) | 96                 | 1e−14   | 42                   |
| XP_008216462             | 3 beta-hydroxysteroid dehydrogenase/isomerase type 4 from \textit{Nasonia vitripennis} (wasp)       | 96                 | 6e−14   | 41                   |
| XP_009051883             | Hypothetical protein LOTGIDRAFT 114779 from \textit{Lottia gigantea} (sea snail)                     | 96                 | 8e−13   | 39                   |
| XP_011181291             | Predicted 3 beta hydroxysteroid dehydrogenase/delta 5 → 4 isomerase type 3 from \textit{Bactrocera cucurbitae} (melon fly) | 96                 | 1e−12   | 44                   |
| WP_012470322             | 3 beta-hydroxysteroid dehydrogenase from \textit{Geobacter lovleyi}                                 | 87                 | 2e−9    | 41                   |

\[ The PCR product was amplified using the primer set VP1/GSP1 and found to align with the 5’ end of \textit{lytFM} within the \textit{\lambda}BlueSTAR\textsuperscript{TM} \textit{D. pteronyssinus} gDNA library. \]
the P60 domain of the bacillus protein along with Glu83 and Tyr18
from the SH3b domain that interact with the free amino group of
the γ-Ala residue are missing. It is also missing from the two SH3
domain containing ascomyces homologues. In addition, the Tyr
involved in catalysis is also invariant [5].

In contrast to the bacillus protein, enzymatic studies show that
although the mycobacterial RipA and RipB endopeptidases have
the same cleavage specificity, they can accommodate peptidogly-
can attached to stem peptide as they lack the equivalent SH3b
domain and possess two cavities, one at each end of the substrate
binding groove; the first allows for interaction with the carbohy-
drate components of peptidoglycan and the second allows for
interaction with the residues beyond the γ-D-Glu of the stem peptide
[27]. Similarly, Lactobacillus casei Lc-P75 protein cleaves the mur-
en tetrapeptide of lysine type peptidoglycans [30]. Given the
sequence homology with the mycobacterial proteins, it is likely
that the mite and other eukaryote proteins will cleave peptide
substrates attached to peptidoglycan.

In searching for bacterial ancestry, 5’ and 3’ sequences were
analysed. With regard to 3’ sequences, the 179 bp flanking
sequence downstream of the 3’ end of lytFM2 was found to contain
several putative prokaryotic promoter sequences although prom-
oter activity was not assessed. However, these sequences might
suggest an early capture of bacterial sequences during mite evolu-
tion, and further study of the 3’ region identified a hypothetical
protein from E. coli. The weak homology of this sequence with the
corresponding genes in E. coli K-12 from which ER1647 (the
host in which the D. pteronyssinus gDNA library was propagated)
was derived by a series of mutations (data not shown) [31–33]
indicates that they were unlikely to be chromosomal contaminants
but generated by bacterial sequences in the gDNA library. If con-
firmed, these may have arisen from endosymbionts incorporated
during preparation of the library. With regard to the 5’ flanking
region, a sequence with high homology to a putative arachnid
3-hydroxysteroid dehydrogenase was identified, suggesting it was
mite-derived.

In summary, the data presented reveal that the mite LytFM
proteins are single domain proteins, that are likely to cleave stem
peptide bound to N-acetylmuramic acid in peptidoglycan. They
are found in several families of Sarcoptiformes mites but are
absent from the herbivorous Trombidiformes mite T. urticae and
the Parasitiformes predatory mite M. occidentalis. They are present
amongst dust mite species and are likely to act in concert with
other enzymes to completely degrade peptidoglycan for either
defence or nutrition. Their presence raises the possibility that
inhalation of mite fecal material containing peptidoglycan
breakdown products may be pro-inflammatory via activation of
NOD (nucleotide-binding oligomerisation domain-containing
protein)-like receptors. Finally, the finding of a gene encoding a
P60 family member in the D. pteronyssinus genome together with
the presence of a bacterial promoter suggests an evolutionary link
to one or more prokaryotic endosymbionts.

Acknowledgements

This study was supported by funds from the Australian National
Health and Medical Research Council. We are grateful to Associate
Professor Martha Ludwig (University of Western Australia) for
advice on the PCR experiments, and Professor Wayne Thomas
and Dr Wendy-Anne Smith (Telethon Kids Institute) for kindly
supplying the D. pteronyssinus gDNA library. VHT planned and
performed the experimental work, analysed the data and
wrote the manuscript. GAS provided all the experimental
reagents and materials, advised in the planning of experiments,
contributed to data analysis and edited the manuscript. BJC
collected to the planning of experiments, oversaw the progress
of the experimental work, contributed to data analysis and edited the
manuscript. The authors declare no competing interests.

References

[1] Mathaba, L.T., Pope, C.H., Lenzo, J., Hartofilis, M., Peake, H., et al. (2002)
Isolation and characterisation of a 13.8-kDa bacteriolytic enzyme from house
dust mite extracts: homology with prokaryotic proteins suggests that the
enzyme could be bacterially derived. FEMS Immunol. Med. Microbiol. 33,
77–88.

[2] Hubert, A., Kuhn, M., Coebel, W. and Kohler, S. (1992) Structural and functional
properties of the p56 proteins from different Listeria species. J. Bacteriol. 174,
8166–8171.

[3] Aramini, J.M., Rossi, P., Huang, Y.J., Zhao, L., Jiang, M., et al. (2008) Solution
NMR structure of the NlpC/P60 domain of lipoprotein Spr from Escherichia coli;
structural evidence for a novel cysteine peptidase catalytic triad. Biochemistry
47, 9715–9717.

[4] Rawlings, N.D., Waller, M., Barrett, A.J. and Bateman, A. (2014) MEROPS: the
database of proteolytic enzymes, their substrates and inhibitors. Nucl. Acids
Res. 42, D551–D559.

[5] Xu, Q., Abdulbek, P., Astakhova, T., Axelrod, H.L., Bakolitsa, C., et al. (2010)
Structure of the gamma-a-glutamyl-diamino acid endopeptidase YkC from
Bacillus cereus in complex with γ-Ala-gamma-γ-Glu: insights into substrate
recognition by NlpC/P60 cysteine peptidases. Acta Crystallogr., Sect. F: Struct.
Biol. Cryst. Commun. 66, 1354–1364.

[6] Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R. and Garcia, J.L. (1999) LytB, a
novel pneumococcal murein hydrolase essential for cell separation. Mol.
Microbiol. 31, 1275–1281.

[7] Takahashi, J., Komatsu, H., Yamada, S., Nishida, T., Labischinski, H., et al.
(2002) Molecular characterization of an atn null mutant of Staphylococcus
aureus. Microbiol. Immunol. 46, 601–612.

[8] Vollmer, W., Joris, B., Chartier, P. and Foster, S. (2008) Bacterial peptidoglycan
(murein) hydrolases. FEMS Microbiol. Rev. 32, 259–286.
