

A Multigene Family of Inactivated Cysteine Proteases in *Sarcoptes scabiei*

To the Editor:

Scabies is a disease of worldwide importance caused by burrowing of the ectoparasitic mite *Sarcoptes scabiei* into the lower stratum corneum. Infestation with *S. scabiei* causes an allergic-type skin reaction with visible hypersensitivity lesions and pruritus and results in significant morbidity, primarily due to secondary infections. Little is known about scabies mite molecules involved in immunological responses. House dust mite (HDM) proteins mediate allergic responses in asthma and similar allergenic molecules may be present in scabies mites (Arlan *et al*, 1988). The difficulty of obtaining scabies mites has hindered molecular studies but expressed sequence tag (EST) libraries have now been established and homologs of HDM allergens identified (Mattsson *et al*, 2001; Fischer *et al*, 2003; Harumal *et al*, 2003; Holt *et al*, 2003; Ljunggren *et al*, 2003 and our unpublished observations).

The HDM group 1 and group 3 allergens are cysteine and serine proteases, respectively (Thomas and Smith, 1999). Among the *S. scabiei* ESTs were 24 distinct contigs with HDM group 3 allergens as their closest homolog (Holt *et al*, 2003). Remarkably in all but one of these the active serine was mutated, and in all but two cases the active site aspartic acid and/or asparagine was also mutated. Hence these genes are no longer a family of active proteases. Phylogenetic relationships suggested that inactivation occurred prior to successive gene duplications. It was termed the scabies mite inactivated protease paralog (SMIPP) family. These SMIPPs are derived from serine proteases and they are thus termed the SMIPP-S family. We hypothesized that SMIPP-S molecules retain the capacity to bind peptides/proteins and this mediates their present function. If so, they may be antagonists of other proteases (Holt *et al*, 2003).

Analysis of the *S. scabiei* ESTs has now identified 33 ESTs clustered into 10 contigs, which each has as its most significant BLASTx match HDM group 1 allergens. SignalP strongly predicted signal sequences of 19–26 amino acids, compared with signal sequences of 18 amino acids in HDM group 1 allergens. Sequences of the *S. scabiei* contigs, the HDM group 1 allergens and the sheep scab mite *Psoroptes ovis* group 1 allergen *Pso o 1* were aligned using ClustalW (Fig S1).

Abbreviations: HDM, house dust mite; EST, expressed sequence tag; SMIPP, scabies mite inactivated protease paralog

The putative *S. scabiei* mature proteases ranged from 225 to 251 amino acids long, compared with 222 amino acids for *Der p 1*, and 223 amino acids for *Der f 1* and *Eur m 1*. The cysteine residues involved in disulfide bonding in *Der p 1* (Topham *et al*, 1994) were conserved and the contigs showed 29.6%–46.1% amino acid identity to mature *Der p 1*.

Relationships were estimated using the maximum likelihood algorithm (Fig 1). Five major clades were evident and these relationships were robust, with similar results produced using parsimony and UPGMA/neighbor joining. Clade 1 contained HDM group 1 genes and the *P. ovis* sequence but no *S. scabiei* gene exhibited a monophyletic relationship with these sequences. Clade 2 contained four *S. scabiei* genes. Each has the catalytic residues of papain-like sequences, cysteine 40 and histidine 191 (numbering according to Fig S1). Clade 3 consists of a single *S. scabiei* gene, which is likely to encode an active protease but the phylogeny of this gene was not robust.

The *S. scabiei* genes of clade 4 shared considerable homology with those of clades 2 and 3. In each case, however, the active cysteine has been replaced by a serine, resulting from a G to C transversion in position two of the codon. In three contigs it also involves a C/T–G transversion in position three. In those contigs with the single mutation (Yv6008G08 and Yv4025A02), the active histidine is replaced by a glutamine, resulting from a C/T to A transversion in position three. In those contigs with the double mutation in the active cysteine codon (Yv5009F04, Yv4025C12 and Yv6018B11), the active histidine has been replaced by a leucine resulting from a A to T transversion in position two. These three contigs also have the important glutamine at position 34 replaced by a glutamic acid (Fig S1).

In order to compare the unexpected multiplicity of group 1 allergen homologs expressed in scabies mites with those in HDM, *Dermatophagoides pteronyssinus* cDNA libraries were constructed and 3168 cDNA clones analyzed. Of these, 92 had *Der p 1* as their closest match, of which 91 fell into the same contig, designated Dp7001E02. This clearly corresponds to the published *Der p 1* single gene as it showed 98% amino acid identity (Chua *et al*, 1993). *Der p 1* clones represented 2.9% of transcripts, strikingly higher than any of the transcripts of the *S. scabiei* homologs (all <0.05%). Clone Dp7027H12 differed substantially from *Der p 1* (54% identity). The active histidine is conserved, but the active cysteine is replaced by a serine.

As in the SMIPP-S family of *S. scabiei* (Holt *et al*, 2003), a cysteine protease gene family has been amplified

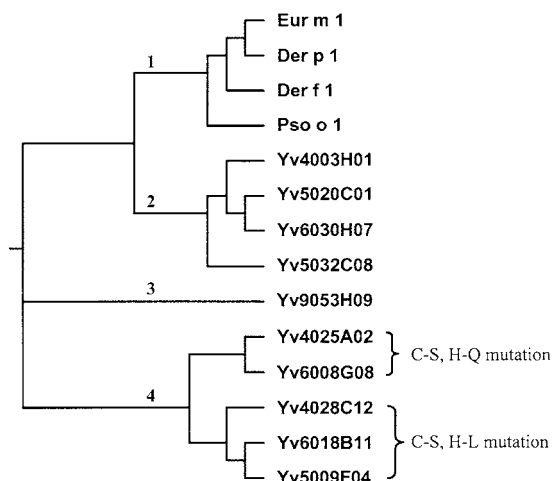


Figure 1
Phylogenetic relationships of the predicted mature protein cysteine protease sequences. The phylogenetic relationships between the *S. scabiei*, *P. ovis* and house dust mite group 1 cysteine proteases were estimated using the maximum likelihood algorithm. Four distinct clades were evident (1–4).

coincidentally with mutations of the active site. Substitution of the active cysteine with a serine would not necessarily abolish activity. Serine and cysteine are functionally interchangeable to some extent, for example, in the cysteine protease picornain 3C of the hepatitis A virus in which mutants with the active cysteine substituted by serine retain catalytic activity (Dessens and Lomonosoff, 1991; Lawson and Semler, 1991; Miyashita *et al*, 1993). In the scabies mite contigs of clade 4, however, the active histidine has also been replaced either by glutamine or leucine. Hence it is not possible to form the thiolate–imidazolium charge relay diad (or a hydroxylate equivalent) that underlies catalysis (reviewed in Sajid and McKerrrow, 2002). Furthermore, the glutamine that helps form the “oxyanion hole” is mutated to glutamic acid in three of the contigs (residue 34—Fig S1). Hence, if these molecules retain peptidase activity, it must be by a previously unknown mechanism. As such, the protein products of these genes are members of the SMIPP family, and are termed SMIPP-C to distinguish them from the SMIPP-S members (Holt *et al*, 2003).

Previously, we postulated that the evolution of SMIPPs may be an adaptation to the parasitic state (Holt *et al*, 2003). If so, they would not be expected to occur in a free-living mite and in accord we did not observe them among HDM ESTs. We hypothesized that SMIPP-S may mediate competitive inhibition of serine proteases of the blood coagulation cascade, complement, or mechanisms such as antagonization of cleavage of protease activated receptors to diminish inflammation. Similar functions may be relevant to SMIPP-C. Although the interactions of SMIPP with host mechanisms remain to be elucidated, the similarities in evolution of these unprecedented families emphasizes their importance.

Deborah C. Holt, Katja Fischer,* Susan J. Pizzutto,† Bart J. Currie,† Shelley F. Walton,† and David J. Kemp*

*Queensland Institute of Medical Research and The Australian Centre for International and Tropical Health and Nutrition, Brisbane, Qld, Australia; †Menzies School of Health Research, Darwin, Northern Territory, Australia

The authors wish to thank Dr Wayne Thomas of the Telethon Institute for Child Health Research and Dr William Cracknell of the Commonwealth Serum Laboratories for providing HDM, the Australian Genome Research Facility for their sequencing service and Mr Danny Wilson for excellent technical assistance. This work was supported by NHMRC (Australia) Program in Medical Genomics Grant 219175 and Project Grant number 137205.

Supplementary Material

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/JID/JID22716/JID22716sm.htm>

Figure S1

ClustalW alignment of the mature protein sequences of group 1 cysteine proteases. Group 1 allergens of house dust mites (*Der p* 1, *Der f* 1, and *Eur m* 1) and the sheep scab mite (*Pso o* 1), aligned with the 10 *S. scabiei* homologs. The cysteine (position 40) and histidine (position 191) of the active site are indicated with white type on black, whereas conserved cysteine residues involved in disulphide bonding are shaded. The sequences of the 10 *S. scabiei* cysteine proteases were submitted to GenBank and were assigned the accession numbers AY525147–AY525156.

DOI: 10.1111/j.0022-202X.2004.22716.x

Manuscript received November 25, 2003; revised January 30, 2004; accepted for publication February 20, 2004

Address correspondence to: David J. Kemp, Queensland Institute of Medical Research and The Australian Centre for International and Tropical Health and Nutrition, Brisbane, Qld, Australia. Email: daveK@qimr.edu.au

References

- Arlian LG, Vyszynski-Moher DL, Gilmore AM: Cross-antigenicity between *Sarcoptes scabiei* and the house dust mite, *Dermatophagoides farinae* (Acari: Sarcoptidae and Pyroglyphidae). *J Med Entomol* 25:240–247, 1988
- Chua KY, Kehal PK, Thomas WR: Sequence polymorphisms of cDNA clones encoding the mite allergen *Der p* 1. *Int Arch Allergy Appl Immun* 101:364–368, 1993
- Dessens JT, Lomonosoff GP: Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology* 184:738–746, 1991
- Fischer K, Holt D, Harumal P, Currie B, Walton S, Kemp D: Generation and characterisation of cDNA clones from *Sarcoptes scabiei* var. *hominis* for an expressed sequence tag library: identification of homologues of house dust mite allergens. *Am J Trop Med Hyg* 68:61–64, 2003
- Harumal P, Morgan MS, Walton SF, *et al*: 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, 2003
- Holt DC, Fischer K, Allen GE, *et al*: 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, 2003
- Lawson MA, Semler BL: Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. *Proc Natl Acad Sci USA* 88:9919–9923, 1991
- Ljunggren EL, Nilsson D, Mattsson JG: Expressed sequence tag analysis of *Sarcoptes scabiei*. *Parasitology* 127:139–145, 2003
- Mattsson JG, Ljunggren EL, Bergstrom K: Paramyosin from the parasitic mite *Sarcoptes scabiei*: cDNA cloning and heterologous expression. *Parasitology* 122:555–562, 2001
- Miyashita K, Kusumi M, Utsumi R, Katayama S, Noda M, Komano T, Satoh N: Site-directed mutagenesis of the putative active site residues of 3C proteinase of coxsackievirus B3: Evidence of a functional relationship with trypsin-like serine proteinases. *Protein Eng* 6:189–193, 1993
- Sajid M, McKerrrow JH: Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol* 120:1–21, 2002
- Thomas WR, Smith W: Towards defining the full spectrum of important house dust mite allergens. *Clin Exp Allergy* 29:1583–1587, 1999
- Topham CM, Srinivasan N, Thorpe CJ, Overington JP, Kalsheker N: Comparative modelling of major house dust mite allergen *Der p* 1: Structure validation using an extended environmental amino acid propensity table. *Protein Eng* 7:869–894, 1994

Figure 1

	1	11	21	31	41	51
<i>Eur m 1</i>	TYAC	SINS	--VSLP	SELDL	RLSL	RTVTPIRM---
<i>Der p 1</i>	TNAC	SIN	--GNAP	AEIDL	RQMR	TVTPIRM---
<i>Der f 1</i>	TSAC	RINS	--VNVP	SELDL	RLSL	RTVTPIRM---
<i>Pso o 1</i>	AQAC	QIGA	--VNIP	NEIDL	RALG	YVTKIKN---
Yv4003H01	LIAC	RFNQ	--THFQ	SEIDL	RKCG	FVTPVKD---
Yv5020C01	-YAC	DFLP	--GQYP	AQIDL	RDIH	LTSIKN---
Yv6030H07	YGFC	NVTE	--TSIF	PEIDL	RKDN	VLTPIRE---
Yv5032C08	LFSC	DKFE	--GVKL	GELDL	RKEG	RVTKIRE---
Yv9053H09	YTIC	RTTSK	--GDLP	EKFDL	RDLG	YLSKIRN---
Yv4025A02	FRFC	NPPPR	--PKLP	KEFDL	RKLK	VIPPVRN---
Yv6008G08	FRFC	NPPPR	--PKLP	KEFDL	RKLK	VIPPVRN---
Yv4028C12	LQFC	KVPKVY	QALPTW	FDLRN	LVLV	TPTRDNSTEKQCKASWAFGP
Yv6018B11	LQFC	EIPKVF	QKLPKW	FDLRN	LVLV	TPTRDNSTEKKCKASWAFGP
Yv5009F04	LQFC	KIPKVF	QKLPKW	FDLRN	LEFV	TPTRDNSTESKCKASWAFGP
	61	71	81	91	101	111
<i>Eur m 1</i>	--RNMS	LDLAEQ	ELVDCA	-SQNG	CHG-DTIP	RGIEYIQQNGVVQEHYYPYVAREQ--SCH
<i>Der p 1</i>	--RNQS	LDLAEQ	ELVDCA	-SQHG	CHG-DTIP	RGIEYIQHNGVVQESYYRYVAREQ--SCR
<i>Der f 1</i>	--RNTS	LDLSEQ	ELVDCA	-SQHG	CHG-DTIP	RGIEYIQQNGVVEERSYPYVAREQ--RCR
<i>Pso o 1</i>	--DNVS	LDLSEQ	ELVDCA	-SQHG	CGG-DTVL	NGLRYIQKNGVVEEQSYPYKAREG--RCQ
Yv4003H01	QVSP	WKFL	SEQELV	DCA-SPHG	CDG-DKMS	SVGFGYIEHKGVLSDQYPYIARVQ--PCQ
Yv5020C01	QVSP	CKFSL	SEQNLID	CA-SKRG	CDG-EKQST	GYSYRFLQQNGTCESSRYPYVAKVQ--QCK
Yv6030H07	QAPL	NKWT	LSEQQL	VDCA-SPKG	CDG-EKPT	TGFKYLLEKGVTTGDRYPYVGKVQ--PCR
Yv5032C08	RWST	NDW-LSP	QQLVDCA	-SDHG	CNG-EKTST	GLEYVQHKGIVKEGVYPYKARVG--VCK
Yv9053H09	HNRT	RKHHF	SEQELV	DCA-SPNTE	GCGSG-NIIS	NGLKYVQLRGVVKSANYPYEEQRNPLGCR
Yv4025A02	-LPHR	HQFQ	LSTQEL	VDCA-GNQG	CRGGVDVTQAF	SYLMEKGVVTEFEYPYAKRG--KCQ
Yv6008G08	-LPHR	HQFQ	LSTQEL	VDCA-GNQG	CRGGVDVTQAF	SYLMEKGVVTEFEYPYAKKG--ICH
Yv4028C12	-IASD	AFSL	SAQNLID	CA-GYQG	CNGGVDVIEAF	NYLQEKGVVKEENYKYKGKKH--ACR
Yv6018B11	-IASD	SFFL	SPQNLID	CA-GYQG	CDGGVDVIEAF	NMHHKGIVKEEFYKYKAKKH--HCR
Yv5009F04	-IASD	SFFL	SPQNLID	CA-GYQG	CDGGVDVIEAF	NMHHKGIVKEEFYKYRAKKH--RCR
	121	131	141	151	161	171
<i>Eur m 1</i>	-RPNA	QRY--GLK	NYCQIS	PPDSNKIR	QALTQTHT	AVAVIIGIKDLNAFRHYDGRITIMQH
<i>Der p 1</i>	-RPNA	QRF--GIS	NYCQIY	PPNVNKIRE	ALAQTHSA	IAVIIGIKDLDAFRHYDGRITIIQR
<i>Der f 1</i>	-RPNS	QHY--GIS	NYCQIY	PPDVVKQIRE	ALQTHTA	IAVIIGIKDLRAFAQHYDGRITIIQH
<i>Pso o 1</i>	-RPNA	KRY--GIK	DLCCQIY	PPNGDKIR	TYLATKQA	ALSVIIGIRDLDSFRHYDGRITILQS
Yv4003H01	-HCFG	PKF--RIGG	YCIIY	PPDKTKIK	VAMTVVQS	AVSAVLLIEDLASFKHYDGKSVISS
Yv5020C01	-HPFG	PNY--KIR	DFCMIH	PENPTIVK	TVLAKF	KAASVSTSLQILDMDKFRHYDGKSVIIN
Yv6030H07	-PPIG	SY--KIR	TFCWY	PPDPKIKI	QVLLTN	RKAALTAVMKITNYMQRHYDGKSVIET
Yv5032C08	-HPCG	PYY--HIK	SFCGIS	PPDPQVKI	ALS	SKTRSALSASIIYDV
Yv9053H09	-GAKG	TRH--TIK	DHCII	FPPSVKRI	QRL	LLKERETAIAAMIEVND
Yv4025A02	EKKF	RKYYH	VKIKDY	CAICPHT	VPILK	SFIYHYKRPLTTILHIRNRRAYNRLGIYAVEDD
Yv6008G08	ARRYR	KYYH	VKIKDY	CAICPHT	VPILK	SFIYHYKRPLTTILHIRNRLKAYNRIGIYAVEDD
Yv4028C12	WRLF	YRRY--KIK	SYCAIC	PATVETIK	KFLFKH	KTALTTVISVRNLTA
Yv6018B11	KRLR	ARRY--KIR	SYCAIC	PATIFTIK	KFLFKH	KTALTTVISVRNLTA
Yv5009F04	KRLR	ARRY--KIR	SYCAIC	PATIFTIK	KFLFKH	KTALTTVISVRNLTA

	181	191	201	211	221	231
<i>Eur m 1</i>	DNG--YQPNYH	AVNIVGYG--	NTQGV	DYWIVRNS	WDTTW	GDNGYGYFAANIN--LMMIEQ
<i>Der p 1</i>	DNG--YQPNYH	AVNIVGYS--	NAQGV	DYWIVRNS	WDTNW	GDNGYGYFAANID--LMMIEE
<i>Der f 1</i>	DNG--YQPNYH	AVNIVGYG--	STQGD	DYWIVRNS	WDTTW	GDSGYGYFQAGNN--LMMIEQ
<i>Pso o 1</i>	DNG--GKRNFH	AINIV-----				
Yv4003H01	ESKR-SKTYGH	GVNIVGYG--	SKYGQ	EVWIVRNS	WGTTW	GDKGYAYFAQNST--VMKLTK
Yv5020C01	DFK--AKPLNH	AVNIVGYG--	PKFGR	QAWIVRNS	WGTSG	WDKGYAYVSQDSS--IFGITK
Yv6030H07	EVRE-GKTLSH	AVNIVGYG--	KFFGK	DAWIVRNS	WGTSG	WDKGYCYVSMNSQ--VFRLL
Yv5032C08_	DGKLIGEQLSH	AINIVGYT--	QRNGV	EVWIVRNS	WGETW	GDHGYGYFKILPKRGVMGITR
Yv9053H09	DAS-TGKKFYH	AVNIVGFK----	TTYWI	IRNSW	GIEWG	DKGYGYVNFGLY----QIEE
Yv4025A02	FGN--KVRQQQ	VVNIVGW	GYHHRGNISY	WIVKNS	MGLHW	GKHGYAFVDIDSD--AFEIRK
Yv6008G08	FGK--KVQHQQ	VVNIVGW	GYHHRGNISY	WIVKNS	MGLHW	GKHGYAFVDIDSD--AFEIRK
Yv4028C12	EGP--SVQRR	LVVNIGG	WRHDDATDR	DYWVVK	NSWGR	KWGMGGYGFVDAEEN--LFGILD
Yv6018B11	EGP--RVQRR	LVVNIVG	WRHDDATDR	DYWVVK	NSWGR	KWGMGGYGFVDAEEN--PFHILD
Yv5009F04	EGT--RVQRR	LVVNIVG	WRHDDATDR	DYWVVK	NSWGR	KWGMGGYGFVDAEEN--PLHILD

	241	251	261
<i>Eur m 1</i>	YPYVVML-----		
<i>Der p 1</i>	YPYVVIL-----		
<i>Der f 1</i>	YPYVVIM-----		
<i>Pso o 1</i>	-----		
Yv4003H01	NVYMAWLH-----		
Yv5020C01	RVYVAWL-----		
Yv6030H07	LVYSASVV-----		
Yv5032C08	MVIVADLGKETVA-----		
Yv9053H09	NAFTVTL-----		
Yv4025A02	NNFYVRLGRPEF-----		
Yv6008G08	NNFYVRLGRPEF-----		
Yv4028C12	QNYLLSVKDEVIPFEAEPDDS-		
Yv6018B11	HNYLVVRVKEDVIPFENEPEDFE		
Yv5009F04	HNYLVVRVKEDVIPFENEPEDFE		