

Cloning and sequencing of the *Dermatophagoides pteronyssinus* group III allergen, *Der p* III

W.-A. SMITH, K.-Y. CHUA†, M.C. KUO*, B. L. ROGERS* and W. R. THOMAS

The Western Australian Research Institute for Child Health, Perth, Western Australia, Australia and

*ImmuLogic Pharmaceutical Corporation, Waltham, MA, USA

Summary

House dust mites are widely recognized as major factors involved in the triggering of allergic diseases such as asthma. It is now apparent that the group III allergens of the *Dermatophagoides* mite species may play a significant role in a number of house dust mite allergic cases. Natural *Der p* III was isolated by gel filtration of salt precipitated *Dermatophagoides pteronyssinus* extract and as reported previously ran as a doublet of Mr 28 and 30 K on sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE). Natural *Der f* III was isolated by affinity purification with the 5A12 monoclonal antibody. Amino acid sequence data was generated for both these proteins which was used to construct DNA probes to screen a *Dermatophagoides pteronyssinus* cDNA library by hybridization and resulted in the isolation of a recombinant *Der p* III cDNA clone, P3WS1. The 1059 bp cDNA fragment included a 786 bp open reading frame which encodes a pre-pro region of 29 amino acids and a mature protein of 232 amino acids with a calculated Mr 24985. A search of the BLAST protein database has confirmed that the *Der p* III P3WS1 clone is approximately 50% homologous with other trypsin proteins. We have confirmed with both our natural protein sequence and the P3WS1 amino acid sequence data that the group III allergens are trypsin-like proteins.

Clinical and Experimental Allergy, Vol. 24, pp. 220–228. Submitted 28 October 1993; revised 30 November 1993; accepted 2 December 1993.

Introduction

House dust mite allergy is now well recognized as a major factor in the development of allergic diseases such as asthma, allergic rhinitis and atopic dermatitis [1]. Over the last decade there has been considerable progress made in analysing the major group I and II allergens from both the mite species, *Dermatophagoides pteronyssinus* (*D. pteronyssinus*) and *Dermatophagoides farinae* (*D. farinae*), with the construction of lambda (λ) cDNA libraries [2]. Both these groups of allergens have been shown to be reactive with IgE in approximately 80% of mite allergic sera. The group III allergens of the mites also appear significant in a number of mite allergic cases although the

significance of the IgE reactivity remains controversial. The reactivity of this group of allergens has been reported to be as low as 16% using a fluid phase assay [3] and as high as 100% using a RAST assay [4]. Several others have reported IgE reactivity between 60 and 83% [5–7]. The discrepancies in the frequency of IgE reactivity to the group III allergens may be attributable to either the differences in the purity of the allergen preparation or the differences in the sensitivities of the assay techniques.

The group III allergens were originally defined as a 29 kDa protein (*Der f* III) isolated from *D. farinae* [3]. A homologous allergen (*Der p* III) has been demonstrated to exist in *D. pteronyssinus* in predominantly a 31 kDa form, with occasionally 34 kDa and 28 kDa forms existing [4]. Both *Der f* III and *Der p* III have subsequently been defined by both *N*-terminal and some internal sequence analysis and reactivity to a *Der f* III monoclonal antibody [3,4,8,9].

There is significant *N*-terminal sequence homology between the group III allergens and both vertebrate and

Correspondence: Dr W. R. Thomas, The Western Australian Research Institute for Child Health, GPO Box D184, Perth, Western Australia 6001, Australia.

† Present address: Department of Pediatrics, National Taiwan University Hospital, No 7, Chung-Shan South Rd, Taipei 10016, Taiwan, Republic of China.

invertebrate trypsin and chymotrypsin proteins [4,8–10]. Furthermore, Ando *et al.* [10] recently reported the isolation of a trypsin-like protease which they suggested was the *Der f* III allergen. They demonstrated that this *D. farinae* protease could selectively hydrolyse trypsin specific arginyl or lysyl residues in a synthetic trypsin substrate and concluded therefore that *Der f* III was a trypsin-like protease. We now report on the isolation of the cDNA for the *Der p* III allergen.

Materials and methods

D. pteronyssinus cultures

Whole mite bodies were purchased from the Commonwealth Serum Laboratories, Parkville, Australia and the spent medium was a gift from the same source.

Purification of natural *Der p* III

Using a method adapted from Heymann [3], 15 ml of a 50–80% saturated ammonium sulphate precipitate of *D. pteronyssinus* spent growth medium was applied to an upward flowing, 2 cm × 90 cm polyacrylamide P-100 column equilibrated in PBS (Pharmacia, LKB Biotechnology, Uppsala, Sweden). The protein content of the eluted 5 ml fractions was determined by measuring optical density (A280 nm) and analysis by SDS-PAGE. Those fractions containing predominantly bands in the 30 kDa region were pooled, concentrated by polyethylene glycol 6000 (BDH Chemicals, Aust. Pty. Ltd. Kilsyth, Victoria, Australia), dialysed against phosphate buffered saline (PBS) and passed over the column again. This was repeated twice until the analysis by SDS-PAGE determined that the only band detectable was a doublet of approximately Mr 30K.

Protein sequence analysis

Purified natural *Der p* III and *Der f* III proteins were subjected to HPEC using a 12% column system (Applied Biosystems, Foster City, CA, USA). HPEC fractions of the appropriate molecular weight were subjected to protein sequence analysis using an Applied Biosystem model 477A gas-phase sequencer with on-line phenylthiohydantoin derivative analysis (model 120).

Preferred codon usage

The codon usage bias for the mature proteins *Der f* I [12], *Der f* II [13], *Der p* I [14] and *Der p* II [15] was determined. The average per cent usage of each triplet codon for each amino acid was determined for each of these four *Dermatophagoides* proteins and the results tabulated in Table 1.

Table 1. Preferred codon usage table for *Dermatophagoides* allergens

Amino acid	Codon	%	Amino acid	Codon	%
Ala A	GCT	45.4	Pro P	CCA	87.2
	GCC	33		CCC	9.6
	GCA	21.6		CCG	3.2
	GCG	0		CCT	0
Cys C	TGT	55.2	Gln Q	CAA	94.4
	TGC	44.8		CAG	5.6
Asp D	GAT	92.3	Arg R	CGT	41.6
	GAC	7.7		CGA	36.5
Glu E	GAA GAG	100 0		CGC	16.5
				AGA	5.4
			CGG	0	
			AGG	0	
Phe F	TTC	82.5	Ser S	TCA	42.5
	TTT	17.5		TCT	18.4
Gly G	GGT	65.2		AGT	11.2
	GGA	24.5		TCG	11.2
	GGC	10.3	AGC	10.2	
	GGG	0	TCC	6.5	
His H	CAT	78.8	Thr T	ACT	39.2
	CAC	21.2		ACA	35
Ile I	ATT ATC ATA	50 48.5 1.5		ACC	18.9
			ACG	6.9	
			Val V	GTT	51.2
				GTC	33.8
Lys K	AAA AAG	99 1		GTA	12.3
Leu L	TTG TTA CTC CTT CTG CTA	64.5 18.5 8 6.5 2.5 0	GTG	2.7	
			Trp W	TGG	100
				Tyr Y	TAT
			TAC		38
			Stop	TAA	75
				TGA	25
Met M	ATG	100		TAG	0
Asn N	AAT	52			
	AAC	48			

Construction of the *D. pteronyssinus* λ gt10 cDNA library

Polyadenylated mRNA (10 μ g) was isolated from 5 g of the *D. pteronyssinus* and 2 μ g used to synthesize cDNA by the RNase H method [16] using a kit (Amersham International, Aylesbury, UK). After the addition of EcoRI restriction enzyme linkers (New England Biolabs, Beverly, USA) the cDNA was ligated to λ gt10 arms and

grown in *Escherichia coli* Y1073 to select for recombinants. Phage DNA was grown in *Escherichia coli* (*E. coli*) JP777 to produce stock for hybridization.

Kinase end-labelling of oligonucleotides

Oligonucleotides were synthesized using an Applied Biosystem PCR mate (Applied Biosystems). Oligonucleotide DNA (20 pm) was end-labelled with (γ - ^{32} P) ATP using T4 polynucleotide kinase (Promega Corp., Madison, Wisconsin, USA) [17]. Labelled oligonucleotide was purified by 15% polyacrylamide gel electrophoresis and subsequently eluted into sterile distilled water.

Isolation of Der p III cDNA clones from the λ gt10 cDNA library

Screening of the library was performed with two oligonucleotide probes designed using the *N*-terminal and internal protein sequence data reported here. The first, P3forward3 (P3F3), was a 38 mer in length with the following nucleotide sequence, 5' TCAGAAAAAGC-TTTGGCTGGTGAATCACCATATCAAAT 3' and the second P3reverse4 (P3R4) was a 41 mer with the following nucleotide sequence, 5' GAATCAACACCAC-CATTAGCAACATCACCACCGCAAATCAT 3'. The library was plated at a 25000 pfu per 150 mm petri dish and the phage lifted onto nitrocellulose (Schleicher & Schuell, Dassel, Germany). Duplicate filter lifts from each plate were denatured and baked [17] for hybridization with a different one of the two probes. Hybridizations were performed in hybridization mix (6 $\times$ sodium chloride, sodium citrate pH 7.0 (SSC), 0.1% Denhardtts, 100 μ g/ml denatured herring sperm DNA) at 42°C with 10⁶ cpm/ml labelled probe overnight. Filters were washed three times at 42°C for 20 min and then at 50°C for 10 min, in a 400 ml wash solution containing 6 $\times$ SSC, 0.1% Triton X-100.

Isolation of DNA from λ gt10 Der p III clones

Phage DNA from the *Der p* III clones was prepared using a polyethylene glycol precipitation procedure [14].

Subcloning and DNA sequencing

Purified *Der p* III phage DNA was digested with EcoRI restriction enzyme (Toyobo) and the released fragment ligated to EcoRI restriction enzyme digested M13mp19 sequencing vector [18]. The recombinant DNA was transformed using *E. coli* TG-1 and sequencing performed using the dideoxynucleotide chain termination method [19] and Sequenase (US Biochemicals). The

oligonucleotide primers used for sequencing included the universal primers, the two primers used for the library screening and two internal primers P3forward5 (P3F5) 5'AAAGCTGTTGGATTACCA3' and P3reverse5 (P3R5) 5'TACATCCGATCCTTTTGC3'. These two internal probes were designed to correspond to nucleotide residues 456–473 (P3F5) and 491–474 (P3R5) (Fig. 1). The primers were used to sequence the isolated clone in both orientations.

DNA and protein sequence analysis

Sequence analysis was performed using the MAC VECTOR software (IBI, New Haven, USA). Computation for sequence homology with other proteins was performed at the NCBI using the BLAST network service. The version of BLAST used was BLAST 1.3.1OMP (7-July-1993).

Results

Sequence of the natural group III allergens

The protein isolation procedure produced a *Der p* III sample which ran as a doublet with Mr 30 and 28 K when analysed by SDS-PAGE (Fig. 2). Both bands reacted with polyclonal mouse anti *Der f* III in accord with the interspecies crossreactivity previously reported [6] (data not shown). *Der f* III isolated from a 5A12 monoclonal antibody [3] column exhibited similar characteristics. *N*-terminal sequencing of both natural *Der p* III and *Der f* III extended the known sequence to 31 residues. Comparison of the original sequence data for the group III allergens with our data has highlighted some interesting differences (Fig. 3). In particular, the non-conservative change at residue 8 from a positively charged lysine to a non-polar hydrophobic leucine (*Der f* III, [3]). Two internal peptides of *Der f* III were isolated by HPEC after CnBr digestion of the natural protein. The first peptide as illustrated in Fig. 3 has a high degree of homology with bovine trypsin, but the second is very poorly homologous to bovine trypsin.

Sequence of recombinant Der p III

The new protein sequence information was used in conjunction with the preferred codon usage data (Table 1) to construct two oligonucleotide probes. The P3F3 oligonucleotide was designed to hybridize to nucleotide residues 159–196 and P3R4 to residues 688–648 of the *Der p* III clone (Fig. 1). Only clones which hybridized strongly with both probes were isolated from the λ gt10 cDNA library. The resulting nucleotide sequence for the P3WS1 clone and the deduced amino acid sequence is described in Fig. 1. The complete nucleotide sequence is 1059 bp in length. This includes a 5' non-coding region of 62 bp, a

	CTTTTTTTTTTTTTTTGAAAAACATTGAAAAACATTGAAGCAACTGTGTGTACAACCAGA	59
	-29	
AAG	ATG ATC ATC TAT AAT ATT TTA ATT GTT TTA TTA TTG GCC ATT AAT	107
Met	Ile Ile Tyr Asn Ile Leu Ile Val Leu Leu Leu Ala Ile Asn	
	-11	
ACA	TTG GCT AAT CCA ATT CTA CCA GCA TCA CCA AAT GCA ACT ATT GTT	155
Thr	Leu <u>Ala Asn</u> Pro Ile Leu Pro Ala Ser Pro Asn Ala Thr Ile Val	
	10	
GGT	GGT GAA AAA GCA TTA GCT GGT GAA TGT CCA TAT CAG ATT TCA TTA	203
Gly	Gly Glu Lys Ala Leu Ala Gly Glu Cys Pro Tyr Gln Ile Ser Leu	
	20	
CAA	TCA AGT AGT CAT TTT TGT GGT GGT ACT ATT CTT GAT GAA TAT TGG	251
Gln	Ser Ser Ser His Phe Cys Gly Gly Thr Ile Leu Asp Glu Tyr Trp	
	40	
ATT	TTA ACA GCT GCA CAT TGT GTT GCC GGA CAA ACA GCA AGT AAA CTT	299
Ile	Leu Thr Ala Ala <u>His</u> Cys Val Ala Gly Gln Thr Ala Ser Lys Leu	
	60	
TCA	ATT CGT TAC AAT AGT TTA AAA CAT TCA TTA GGT GGT GAA AAA ATT	347
Ser	Ile Arg Tyr Asn Ser Leu Lys His Ser Leu Gly Gly Glu Lys Ile	
	70	
TCT	GTT GCT AAA ATT TTT GCA CAT GAA AAA TAT GAT AGT TAT CAA ATT	395
Ser	Val Ala Lys Ile Phe Ala His Glu Lys Tyr Asp Ser Tyr Gln Ile	
	90	
GAT	AAT GAT ATT GCA TTG ATT AAG CTT AAA TCA CCT ATG AAA TTA AAT	443
Asp	Asn <u>Asp</u> Ile Ala Leu Ile Lys Leu Lys Ser Pro Met Lys Leu Asn	
	100	
CAG	AAA AAT GCC AAA GCT GTT GGA TTA CCA GCA AAA GGA TCG GAT GTA	491
Gln	Lys Asn Ala Lys Ala Val Gly Leu Pro Ala Lys Gly Ser Asp Val	
	120	
AAA	GTT GGT GAT CAA GTT CGT GTT TCT GGT TGG GGT TAT CTT GAA GAA	539
Lys	Val Gly Asp Gln Val Arg Val Ser Gly Trp Gly Tyr Leu Glu Glu	
	140	
GGA	AGT TAT TCA TTA CCA TCT GAA TTA AGA CGT GTT GAT ATT GCT GTT	587
Gly	Ser Tyr Ser Leu Pro Ser Glu Leu Arg Arg Val Asp Ile Ala Val	
	150	
GTA	TCA CGT AAA GAA TGT AAT GAA TTA TAT TCA AAA GCT AAT GCT GAA	635
Val	Ser Arg Lys Glu Cys Asn Glu Leu Tyr Ser Lys Ala Asn Ala Glu	
	170	
GTT	ACT GAT AAT ATG ATT TGT GGT GGT GAT GTT GCA AAT GGT GGT AAA	683
Val	Thr Asp Asn Met Ile Cys Gly Gly Asp Val Ala Asn Gly Gly Lys	
	180	
GAT	TCT TGT CAA GGT GAT TCT GGT GGA CCG GTT GTT GAT GTT AAA AAT	731
Asp	Ser Cys Gln Gly Asp <u>Ser</u> Gly Gly Pro Val Val Asp Val Lys Asn	
	200	
AAT	CAA GTT GTT GGT ATT GTT TCA TGG GGT TAT GGT TGT GCA CGT AAA	779
Asn	Gln Val Val Gly Ile Val Ser Trp Gly Tyr Gly Cys Ala Arg Lys	
	220	
GGT	TAT CCA GGT GTT TAT ACA CGT GTT GGT AAT TTT ATC GAT TGG ATT	827
Gly	Tyr Pro Gly Val Tyr Thr Arg Val Gly Asn Phe Ile Asp Trp Ile	
	230	
GAA	TCA AAA CGT TCA CAG TGATTGATAAAAAACAATTTTCGACATATAAAAAATTGGCAA	884
Glu	Ser Lys Arg Ser Gln *	
	ATGATGCCATTATATGTTATCATTATGTTTCTGGTGATTCTTTTCTACTTTTATCATTTTAT	947
	TTTATATTCGAGCTGAAAAACAAACAATCATGATTGTATAGGGATTGTGTGTTTTCGAATT	1010
	TAAAAAATCCAAATATAATAAATAAATCATTTCACAAAAAAG	1059

Fig. 1. Complete nucleotide and deduced amino acid sequence of *Der p* III P3WS1. The nucleotide residues of the cDNA are numbered at the end of each row, with the amino acid residue numbers above each line. The positive numbering commences at the start of the mature protein. The stop codon is represented as an *. The polyadenylation signal, AATAAA between nucleotide residues 1024–1029, is underlined. The Met residue at the translation initiation site is highlighted in bold type. The residues involved in the catalytic site of the mature protein are underlined and those in the substrate specificity of the protein are in bold type.

211 bp 3' untranslated region and an open reading frame of 786bp with a stop codon (TGA) at nucleotide residues 846–848. There is no poly A tail but there does appear to be a polyadenylation signal (AATAAA). The open reading frame includes a 29 amino acid (Mr 3000) pre-pro region and starting at the *N*-terminal Ile, a mature protein of 232 amino acids with a calculated Mr 24 985 and pI of

8.5. The Met residue (ATG) at amino acid position -29 is the most likely translation initiation site. There are two reasons for this: firstly, the subsequent sequence codes for a classical signal peptide of 18 amino acid residues with predominantly hydrophobic residues present. Secondly, the immediate sequence prior to this ATG codon, 5'GAAAGATG3', conforms loosely to the Kozak con-

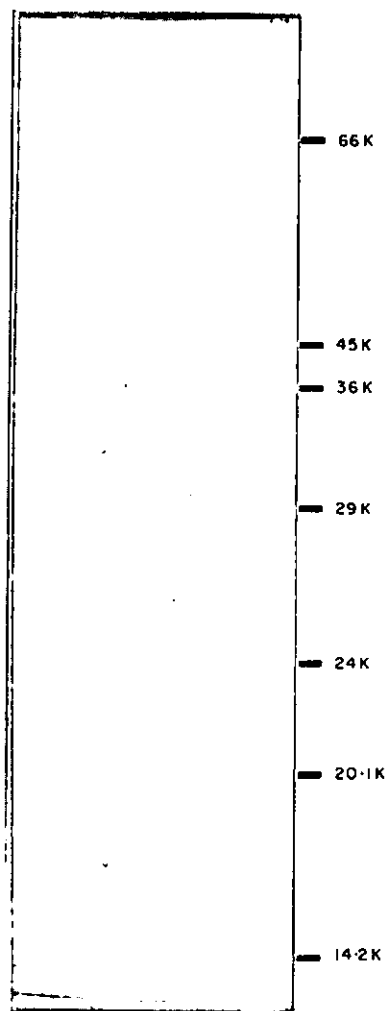


Fig. 2. SDS (13%) polyacrylamide gel electrophoresis of a natural *Der p* III preparation isolated from *D. pteronyssinus* spent mite medium by salt precipitation and subsequent size exclusion chromatography. Lanes 1, 2 and 3 contain 20, 30 and 40 μ g of protein respectively. It is evident that the *Der p* III protein exists as a doublet of 30 and 28 kDa.

sensus sequence (CCACCATG) for the eukaryotic translation initiation site with the crucial purine (most often A) at the -3 position (underlined) [20].

A search of the protein database showed that the *Der p* III clone P3WS1 is highly homologous to both vertebrate and invertebrate trypsins and in particular there are clusters of residues that are highly conserved amongst trypsin proteins from evolutionary diverse origins (Fig. 4).

Discussion

The translated *N*-terminal sequence of the *Der p* III P3WS1 clone was identical to the 18 residue *Der p* III sequence reported by Stewart *et al.* where the allergen was isolated on the basis of its substrate specificity [4]. Our own protein sequencing data differ from the protein sequence for the P3WS1 clone as a result of the substitutions of Gln/Glu at residue 11, a Ser/Cys substitution at residue 12 and a Ser/Cys substitution at residue 25. These substitutions, like those reported by Ando *et al.* [10] for the *Der f* III *N*-terminal sequence may be due to the existence of *Der p* III in different isoforms as will be discussed in more detail later. Analysis of the deduced *Der p* III P3WS1 amino acid sequence confirmed that the position of the *Der f* III CnBr peptide 1 (MI(C)GGD-VANGGVDS(E)QD) was correctly predicted to be from amino acid residues 167–184 and has 88% identity with the recombinant *Der p* III sequence in this region (MICGGDVANGGKDSCQD). Interestingly, there are two identical non-conservative changes that exist when the natural *Der f* III peptide one sequence is aligned with both the bovine trypsin and the P3WS1 clone sequence. There is a change from a non-polar hydrophobic Val at 178 in *Der f* III to a positively charged Lys residue in both the bovine and P3WS1 clone sequences. The other change is a Glu residue at 181 in the *Der f* III substituted for a Cys residue in both the bovine trypsin and P3WS1 clone sequence. This cysteine forms one of the disulphide bridges in both these and most other trypsin proteins. The predicted alignment of the second *Der f* III peptide ((M)TL DQTNA(K)PVPL(P)(T)(S)) produced by the CnBr digestion was from amino acid residues 95–110(MKLNQKNAKAVGLPAK). While the identity between this fragment and bovine trypsin is only 25%, there is 63% identity with the P3WS1 clone in this region. These differences between the *Der p* III and *Der f* III protein sequences are consistent with the 20% sequence variation found between other homologous allergens from this species such as the *Der p* I and *Der f* I; and *Der p* II and *Der f* II.

Stewart *et al.* [4] reported the isolation of small Mr degradation products resulting from the autolysis of trypsin isolated from *D. pteronyssinus*. The products were sequenced and the peptides aligned to the crayfish trypsin sequence. Peptide 2 (NNQVXGI) was aligned to a region corresponding to amino acid residues 98–105 in Fig. 1. Our nucleotide sequence data seems to suggest that this is not the correct alignment position but in fact the correct position is between residues 194–200 (NNQVVGI).

The complete sequence for the *Der p* III P3WS1 clone encodes a protein of 261 amino acids. *Der p* III, like all known trypsins, is synthesized as a pre-pro zymogen. The

N-terminal sequence

	1	20
<i>Der f</i> III ext.	IVGGVKAK(A)GD S PYQI S LQS S S HF X G(G)SILD	
	(C)	
<i>Der f</i> III*	L A .. X ..E. . .EV	
<i>Der p</i> III ext.	E..L A .Q(.)(.)...(.)..(S) ..(.)T..	
<i>Der p</i> III*	EX.L A .E X	

Fig. 3. Protein sequence homology of the natural group III proteins. Numbering of the *N*-terminal sequence is based on the extended *Der f* III (*Der f* III ext.) sequence. The new *Der f* III is compared with the previously published *N*-terminal sequences for *Der f* III# [3], *Der p* III# [4] and our new extended *Der p* III (*Der p* III ext.) data. Numbering for sequence alignments of the peptides produced by CnBr digestion of *Der f* III is based on the *Der p* III P3WS1 deduced amino acid sequence. Identical amino acid residues indicating homology with *Der f* III are indicated by an (.)

Sequence alignments of the internal peptides

	167	184
<i>Der f</i> III (peptide 1)	MI (C)GGDVANGGVDS (E)QGD	
Bovine trypsin	.F . A.YLE...K.. C ...	
	95	110
<i>Der f</i> III (peptide 2)	(M)TLDQTNA (K)PVPL (P) -- (T) (S)	
Bovine trypsin	A S.NSRV. - SIS. . -- . .	

Fig. 4. Comparative analysis of the amino acid sequences of trypsin proteins from various species. The amino acid sequence for *Der p* III P3WS1 is compared with bovine trypsin [28], crayfish trypsin [27] and Atlantic salmon trypsin [32]. Those residues which are highly conserved amongst all four sequences are represented in bold type. The boxed cysteine residues are those residues involved in the formulation of disulphide linkages, the linkages being represented by the solid lines joining two boxes. Only those disulphide linkages that are formed by *Der p* III P3WS1 are indicated. The numbering of the residues is based on the *Der p* III P3WS1 deduced amino acid sequence.

	1	20	40
<i>Der p</i> III	IVGGEKALAGECPYQIS	-----LQSSSHFCGGTILDEYWILTA	AAHCV-----AG
Bovine	IVGGYTCGANTVPYQVS	-----LNSGYHFCGGSLINSQWVVS	AAHC-----YKSG
Crayfish	IVGGTDAVLGEFFPYQLSFQETFLGFSFHFC	ASIIYNNENYAITAGHC	VYGGDDYENPSG
Salmon	IVGGYECRKNSASYQAS	-----LQSGYHFCGGSLISSTWVVS	AAHC-----YKSR
	60	80	
<i>Der p</i> III	-QTASKLSIRYNSLKHSLGGE-KI-SVAKIFAHEKYDSYQIDNDIALIKLKSPMKLN		
Bovine	IQVR--LGED-NI-NVVEGNEQFI-SASKSIVHPSYNSNTLNNDIMLILKLSAASLN		
Crayfish	LQIV--AGEIDMS--VNEGSEQTI-TVSKRIILHENFDYDLLDNDISLKLKLSGLTFN		
Salmon	LQVR--LGEH-NI-AVNEGTEQFIDSV-KVIMHPSYNSRNLDNDIMLILKLSKPASLN		
	100	120	140
<i>Der p</i> III	QKNKAVGLPAKGSVDKVGQVVRVSGWGYLEEGSYSLPSELRRVDIAVVSRRKQNEL		
Bovine	SRVA-SISLFP--TSCASAGTQCLISGWNKSSGTSYPDVLKCLKAPILSDSSCKSA		
Crayfish	NNVA-PIALP--AQGHATGNVIIVTGWGTTEGGNT-PDVLQKVTVPLVSDAEQRDD		
Salmon	SYVS-TVALP--SSCASSGTRCLVSGWGNL-SGSSSYPDTLRCLDLPLSSSSCNLSA		
	160	180	200
<i>Der p</i> III	YSKANAETDNDMIGGVDVANGGRKDS	CGGDSGGFPVVDVKN----	NQ-VVGIVSWGYGCC
Bovine	Y---PGQITSNMFACAGYLE-GGRKDS	CGGDSGGFPVVC	SGK-----LQ---GIVSWGSGCC
Crayfish	YG--ADEIFDSMFCAGVPE-GGRKDS	CGGDSGGPLAASDTGSTYLA---	GIVSWGYGCC
Salmon	Y---PGQITSNMFACAGFME-GGRKDS	CGGDSGGFPVVC	CNQ-----LQ---GVVSWG-GC
	220		
<i>Der p</i> III	ARKGYPGVYTRVGNFIDWIESKRSQ		
Bovine	AQKNKPGVYTKVCNYSWIKQTASN		
Crayfish	ARPGYPGVYTEVSYHVDWIKANAV		
Salmon	AQRNKPGVYTKVCNYSWISSTMSN		

cleavage site for the signal peptide is postulated to be between amino acid residue -12 and -11 (as underlined in Fig. 1) as the residues around this site conform to the amino acid constraints outlined by von Heijne [21]. von Heijne proposed that small neutral amino acids are

strongly preferred at positions -1 and -3 from the cleavage site, such as the Ala and Tyr seen in the *Der p* III cDNA. Pro residues are never found in positions +1 to -3 thus excluding any other sites. Most mammalian trypsin proteins have been reported to have pre or signal

peptides of 15 or 16 amino acids. The *Der p* III pre peptide is 18 amino acids in length. Whilst the difference in length is relatively small, it does seem that the length of the signal peptide may be attributable to the phylogenetic diversity of the species from which the trypsin originated. This is most apparent in *Streptomyces griseus* (*S. griseus*) with a 32 amino acid pre peptide [22]. It has also been reported that for mammalian trypsins the signal peptide contains two specific clusters of two and then four hydrophobic residues [23,24]. There is an abundance of hydrophobic residues within this region of the *Der p* III cDNA but they are not arranged into highly conserved clusters. The same is true of the signal peptides of trypsins from both *Drosophila melanogaster* (*D. melanogaster*) [25] and *S. griseus* [22]. It is possible therefore that the highly conserved clusters of hydrophobic residues are not a characteristic of all trypsin proteins but rather that of mammalian trypsins.

Downstream from the signal peptide is the 11 residue pre-activation peptide from amino acid residues -11 to -1. Comparison of this pro region with other pro regions from trypsins of various species indicates that the *Der p* III pro peptide is unusual. Trypsin proteins from vertebrates have been shown to have an octapeptide and/or a hexapeptide pro region with four contiguous aspartyl residues followed by a lysine residue at the -1 position [24]. It is the carboxyl region of this lysine residue which is cleaved during the activation of the trypsinogen to trypsin. The *Der p* III lacks this polyaspartyl-lysine sequence. The trypsinogen therefore can not be activated by the usual mechanism. There are a few other trypsins which are similar to *Der p* III in this respect, *D. melanogaster* [25], *S. griseus* [22] and most interestingly, a 32 kDa anionic trypsin from the rat pancreas [26]. This group showed that enterokinase, an enzyme which has a highly specific affinity for the polyaspartyl-lysine residues in trypsin proteins, had no effect on the trypsin that they isolated. They suggested therefore that a different mechanism was being used for the activation process. The absence of any homology of the *Der p* III pro enzyme region with any of the known sequences may imply a unique activation mechanism.

A search of the protein database has confirmed that *Der p* III P3WS1 is highly homologous to both vertebrate and invertebrate trypsins. The *Der p* III allergen is not a mite chymotrypsin [7] as comparison of the *N*-terminal protein sequences showed that they differ by as much as 50%. Figure 4 shows a comparison of the amino acid sequences of *Der p* III P3WS1, crayfish, bovine and salmon trypsins. The salmon has the highest degree of identity with *Der p* III with 47%, closely followed by bovine and crayfish trypsin with 45% and 44% respectively. The residues in bold type are those that are most

highly conserved amongst all trypsin proteins [27]. There are seven cysteine residues; six of these residues are predicted to form disulphide bridges, 25-41, 152-169 and 181-207 (Fig. 4) [28]. The possible role of the seventh cysteine is discussed in detail later. Most importantly the *Der p* III protein contains the highly conserved residues involved in the catalytic activity and substrate specificity of the trypsin proteins. The conserved His 40, Asp 85 and Ser 185 (underlined, Fig. 1) represent the charge relay system which comprises the active catalytic site of the trypsin enzyme. The positively charged Asp residue at 179 represents the trypsin specificity site. Most serine proteases have a neutral amino acid at this residue [29]. The Gly residues at 204 and 214 in combination with the Asp 179 are responsible for accommodating the bulky positively charged side chains of lysine and arginine, the substrate residues cleaved by trypsins. Amino acid residues 172-178 and 208-212 form the S1 binding pockets which are involved in the binding of the substrate to the trypsin molecule. The areas immediately surrounding all of these important residues are the most conserved regions of the *Der p* III protein. All these results are consistent with the fact that among trypsin proteins from different species structural variations occur in those regions of the molecule which are not important for catalytic activity [30].

Most trypsin proteins generally have a common pairing of cysteine residues to form six disulphide bridges, with two of these bridges, Cys 7-141 and Cys 113-219 (Fig. 4) being unique to this group of proteins. *Der p* III like both the crayfish [27] and *S. griseus* [22] are missing these two unique links, but more importantly *Der p* III is unique as it contains an extra unpaired cysteine residue, Cys 12. We propose that this residue may be equivalent to Cys 7 in the bovine trypsin which unlike Cys 141, 113 and 219 was present prior to the evolutionary divergence of the mite trypsin.

The isolation of the cDNA for *Der p* III is the first primary sequence determination for the group III *Dermatophagoides* allergens. Comparison of the sequence with other vertebrate and invertebrate serine proteases has confirmed the analysis with substrate binding experiments [4,8], indicating that the group III allergens are trypsin-like proteins. Trypsin proteins are secreted as pre-pro zymogens by the pancreatic acinar cells of all vertebrate and invertebrate species. Invertebrate trypsins have been reported to have a MW ranging from 20 to 30 kDa [31]. The *Der p* III P3WS1 trypsin has a calculated Mr of 24 985 while the natural protein we isolated existed as a duplicate of 28 and 30 kDa as estimated by SDS-PAGE. There have been other reports of multiple protein bands being isolated for both *Der f* III [6] and *Der p* III [4]. These results are not unusual as trypsin proteins are

known to exist in several isoforms. Gendry and Launay [26] isolated duplicate protein bands of an anionic trypsin-like protein from rat pancreas. They demonstrated that the higher 32 kDa band represented the inactive trypsin and the lower 30 kDa band represented the active form of the trypsin which could autocatalyse the cleavage of the 32 kDa band. It is therefore possible that the group III allergens exist as an inactive trypsinogen and the 28 kDa band the active trypsin although there was no evidence of this in our amino acid sequence and the calculated MW is within the variation found for SDS-PAGE estimations. The doublet found by SDS-PAGE could also possibly arise by either the resistance of the disulphide bonds in some molecules to reduction, affecting the binding of SDS or the number of disulphide bonds present as a reflection of the existence of different isoforms. The number of isozymes for invertebrate trypsins has been reported to range from 1 to 12 [23]. Stewart *et al.* [4] proposed that nine major isoforms of *Der p* III existed, with *p* I ranging from 4 to > 8. It is apparent, therefore, that we now need to confirm whether in fact *Der p* III does exist in different isoforms and how these polymorphic isoforms differ from our original clone.

Acknowledgments

This work was supported by the National Health and Medical Research Council of Australia and ImmuLogic Pharmaceutical Corporation. We thank Liam O'Connor, Wendy Nachajko and Andrew W. Brauer for their technical assistance.

References

- 1 Platts-Mills TAE, Chapman MD. Dust mites: Immunology, allergic disease, and environmental control. *J Allergy and Clinical Immunol* 1987; 80:755-75.
- 2 Thomas WR, Chua K-Y, Stewart G, Greene, W. Allergy to the house dust mite. *Today's Life Science* 1990; 2:20-8.
- 3 Heymann PW, Chapman MD, Aalberse RC, Fox JW, Platts-Mills TAE. Antigenic and structural analysis of group II allergens (*Der f* II and *Der p* II) from house dust mites (*Dermatophagoides* spp). *J Allergy Clin Immunol* 1989; 83:1055-67.
- 4 Stewart GA, Ward LD, Simpson RJ, Thompson PJ. The group III allergen from the house dust mite *Dermatophagoides pteronyssinus* is a trypsin-like enzyme. *Immunology* 1992; 75:29-35.
- 5 Tovey ER, Baldo BA. Comparison by electroblotting of IgE-binding components in extracts of house dust mite bodies and spent mite culture. *J Allergy Clin Immunol* 1987; 79:93-102.
- 6 Thomas WR, Chua K-Y, Smith W-A. Molecular polymorphisms of house dust mite allergens. *Exp Appl Acarol* 1992; 16:153-64.
- 7 Yasueda H, Mita H, Shida T, Ando T, Sugiyama S, Yamakawa H. Allergens from *Dermatophagoides* mites with chymotryptic activity. *Clin Exp Allergy* 1993; 23:384-90.
- 8 Ando T, Homma R, Ino Y *et al.* Is a trypsin-like protease of mites a *Der f* III allergen? *Arerugi* 1992; 41:704-7.
- 9 Stewart GA, Thompson PJ, Simpson RJ. Protease antigens from house dust mite. *Lancet* 1989; 2:154-155 (plus corrections 462).
- 10 Ando T, Homma R, Ino Y *et al.* Trypsin-like protease of mites: purification and characterization of trypsin-like protease from mite faecal extract *Dermatophagoides farinae*. Relationship between trypsin-like protease and *Der f* III. *Clin Exp Allergy* 1993; 23:777-84.
- 11 Kohmoto S, Kodera Y, Takahashi K *et al.* Activation of the kallikrein-kinin system in human plasma by a serine protease from mites. *J Clin Biochem Nutr* 1991; 10:15-20.
- 12 Dilworth RJ, Chua K-Y, Thomas WR. Sequence analysis of cDNA coding for a major house dust mite allergen *Der f* I. *Clin Exp Allergy* 1991; 21:25-32.
- 13 Trudinger M, Chua K-Y, Thomas WR. cDNA encoding the major mite allergen *Der f* I. *Clin Exp Allergy* 1991; 21:33-7.
- 14 Chua K-Y, Stewart GA, Thomas WR *et al.* Sequence analysis of cDNA coding for a major house dust mite allergen, *Der p* I. Homology with Cysteine proteases. *J Exp Med* 1988; 167:175-82.
- 15 Chua K-Y, Doyle CR, Simpson RJ *et al.* Isolation of cDNA coding for the major mite allergen *Der p* II by IgE plaque immunoassay. *Int Arch Allergy Appl Immunol* 1990; 91:118-23.
- 16 Gubler U, Hoffman BJ. A simple and very efficient method for generating cDNA libraries. *Gene* 1983; 25:263-9.
- 17 Maniatis T, Fritsch EF, Sambrook J. *Molecular Cloning. A laboratory manual*. 2nd Edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989.
- 18 Messing J. New M13 vectors for cloning. *Method Enzymol* 1983; 101:20-78.
- 19 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74:5463-7.
- 20 Kozak M. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNA's. *Nucleic Acids Research* 1984; 12:857-72.
- 21 von Heijne G. Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem* 1983; 133:17-21.
- 22 Kim JC, Cha SH, Jeong ST, Oh SK, Byun SM. Molecular cloning and nucleotide sequence of *Streptomyces griseus* trypsin gene. *Biochem Biophys Res Commun* 1991; 181:707-13.
- 23 MacDonald RJ, Stary SJ, Swift GH. Two similar but nonallelic rat pancreatic trypsinogens. *J Biol Chem* 1982; 257:9724-32.
- 24 Le Huerou I, Wicker C, Guilloteau P, Toullec R, Puigserver A. Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic trypsinogen. Structural identity within the trypsin family. *Eur J Biochem* 1990; 193:767-73.

- 25 Davis CA, Riddell DC, Higgins MJ, Holden JJ, White BN. *Drosophila melanogaster* α -gene encoding trypsin-like enzyme. Nucleic Acid Research 1988; 13:6605-15.
- 26 Gendry P, Launay J-F. Purification and characterization of a trypsin-like protein from rat pancreas. Biochemica et Biophysica Acta 1988; 955:243-9.
- 27 Titani K, Sasagawa T, Woodbury RG et al. Amino acid sequence of Crayfish (*Astacus fluviatilis*) trypsin I_r. Biochemistry 1983; 22:1459-65.
- 28 Hartley BS. Homologies in serine proteinases. Phil Trans Roy Soc Lond B 1970; 257:77-87.
- 29 Keil B. The Enzymes. New York: Academic Press, 1971:249-75.
- 30 Vithayathil AJ, Buck F, Bier M, Nord FF. On the mechanism of enzyme action. LXXII. Comparative studies on tryptins of various origins. Arch Biochem Biophys 1961; 92:532-40.
- 31 Graf R, Briegel H. Isolation of trypsin isozymes from the mosquito *Aedes aegypti* (L). Insect Biochem 1985; 15:611-8.
- 32 Male R, Lorens JB, Smals AO, Jensen MF, Torrissen KR. Cloning and characterization of cationic and anionic forms of trypsin from atlantic salmon (submitted).

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.