

Wendy-Anne Smith
Wayne R. Thomas

TVW Telethon Research Institute for Child
Health, Perth, Australia

Comparative Analysis of the Genes Encoding Group 3 Allergens from *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*

Key Words

Der f 3
Der p 3
Sequence polymorphisms
Dermatophagoides
Trypsin-like proteins
House dust mite

Abstract

Group 3 allergens of the genus *Dermatophagoides* represent one of the major groups of house dust mite allergens. The cDNA sequence data for Der p 3, in combination with both N-terminal amino acid sequences and substrate affinity data, have confirmed that the group 3 allergens are trypsin-like proteases. Using the information from the Der p 3 P3WS1 cDNA clone, genes encoding both Der f 3 and Der p 3 have now been amplified by the polymerase chain reaction and analysed. Two Der f 3 clones and three Der p 3 genomic clones were sequenced. Each of the clones contained a single small intron and encoded a mature protein of 233 amino acids. The nucleotide sequence was identical for both Der f 3 clones. There was 81% identity between the Der f 3 sequence and the original Der p 3 P3WS1 clone. The calculated molecular weight of Der f 3 was 25.27 kDa compared to 24.98 kDa for Der p 3. All the amino acid residues required for the catalytic activity and the substrate specificity were conserved between the two homologues. The coding sequences of two of the three Der p 3 genomic clones were identical to the original Der p 3 P3WS1 clone with the third having nucleotide changes resulting in four non-conservative amino acid substitutions in the mature protein. These substitutions resulted in a molecule with a slightly larger molecular weight and a more acidic pI value than the original Der p 3 clone. This third Der p 3 genomic clone is, therefore, an isoform of the Der p 3 P3WS1 clone and is classified as an isovariant of the allergen. The nucleotide sequence data presented are the first reported for Der f 3. The Der f 3 gene, like the Der p 3 gene, encoded a trypsin-like protease, but with a slightly larger molecular weight.

Introduction

House dust mites of the *Dermatophagoides* genus play a significant role in triggering allergic diseases such as asthma and allergic rhinitis. Group 3 allergens represent one major component of the *Dermatophagoides* species believed to trigger this allergic reaction. The frequency of IgE reac-

tivity towards the group 3 allergens, Der f 3 and Der p 3, has been reported to be approximately 70–80% [1–3].

The group 3 allergens were originally defined by Heymann et al. [4] with the isolation of a 29 kDa protein from the *Dermatophagoides farinae* mite. Furthermore, Ando et al. [5] reported the isolation of a trypsin-like protease from *D. farinae* that they concluded was the Der f 3 allergen.

KARGER

E-Mail karger@karger.ch
Fax +41 61 306 12 34

© 1996 S. Karger AG, Basel
1018–2438/96/1092–0133 \$10.00/0

Correspondence to: Dr. Wayne Thomas
Institute for Child Health Research
PO Box 855
West Perth
Western Australia 6872 (Australia)

Received:
April 24, 1995
Accepted after revision:
August 10, 1995

Table 1. PCR and sequencing oligonucleotides used to isolate and sequence the genomic DNA for both Der f 3 and Der p 3.

Oligonucleotides	Ori-entation ^a	Source ^b
1. CAGGAATTCATACACGTGTTGG	F	N
2. TCTGAATTCGACATTCACCAGC	R	N
3. TCAGAAAAAGCTTTGGCTGGTGAATCACCATATCAAAT	F	P
4. <u>GCGGATCC</u> ATTGTTGGTGGTGA AAAAAGC	F	N
5. AAAGCTGTTGGATTACCA	F	N
6. GAATCAACACCACCATTAGCAACATCACCACCGCAAATCAT	R	P
7. TACATCCGATCCTTTTGC	R	N
8. AATCACTGTGAACGTTTGA	R	N

Oligonucleotides 4 and 8 were used to PCR amplify the genomic DNA. The nucleotide sequence underlined in oligonucleotide 4 represents a *Bam*HI restriction site. All oligonucleotides are typed 5'-3'.

^a F = Forward primers, i.e. 5'-3'; R = reverse primers, i.e. 3'-5'.

^b The sequence information, nucleotide (N) and protein (P), used to generate these oligonucleotides has been described previously [6].

Their quantitative IgE-binding assays showed that the frequency of reactivity with Der f 3 was as high as that observed for Der f1 (79%) and Der f2 (83%). There was also no significant difference in the mean IgE-binding levels amongst these three purified allergens.

The immunological cross-reactivities and structural similarities of group 1 and 2 allergens of *Dermatophagoides pteronyssinus* and *D. farinae* have been well characterised. For example, Heymann et al. [4] observed that human IgE from mite-allergic patients demonstrated extensive cross-reactivity between Der f1 and Der p1. Similarly skin tests and antigen-binding assays by the same group showed a strong correlation between IgE antibodies and the group II allergens. This is in keeping with the structural similarity observed between Der p1 and Der f1, and between Der p2 and Der f2. Immunological cross-reactivity has also been observed within the group 3 allergens [1, 4]. It appears, therefore, that to assess the immunological importance of this cross-reactivity, the primary structures of both group 3 allergens, Der f 3 and Der p 3 need to be determined.

We have previously reported the isolation of the cDNA for Der p 3 P3WS1 confirming that the group 3 allergens are trypsin-like proteins [6]. The deduced amino acid sequence was highly homologous to previously published N-terminal natural protein sequences for both Der f 3 and Der p 3 [4, 7]. Genomic DNA sequencing was undertaken (1) to determine the full sequence of Der f 3 and (2) to determine the genomic organisation for studies on sequence polymorphisms. Using the cDNA sequence information for Der p 3 P3WS1, we took two oligonucleotide primers to amplify both genomic Der f 3 and Der p 3 by the polymerase chain reaction (PCR).

Materials and Methods

Dermatophagoides Cultures

Whole mite bodies were purchased from CSL, (Australia) and the spent mite medium was a gift from the same source.

Oligonucleotides

Oligonucleotides were synthesised using an Applied Biosystem PCR mate (table 1).

Extraction of Genomic DNA from *Dermatophagoides*

Live mites (5 g) were homogenised in a Dounce homogeniser in 30 ml of 6 M guanidine hydrochloride, and the soluble extract was, centrifuged over a 4.8 M CsCl pad [8]. The DNA at the top of the pad was collected and extracted with phenol.

PCR Conditions and Thermal Cycling Parameters

All the PCRs performed included a negative control. The thermal cycling was conducted in a PTC-100 programmable thermal cycler (Bresatec, Australia). The 40-cycle programme was as follows: 1 cycle of 5 min at 94°C, 2 min at 60°C, 2 min at 72°C (elongation step); 38 cycles of 1 min at 94°C, 2 min at 60°C, 2 min at 72°C, and 1 cycle at 94°C, 2 min at 60°C, 10 min at 72°C.

PCR with *Pfu* Polymerase

The PCR was performed according to the manufacturer's instructions with the addition of 20 pmol oligonucleotide primers, 250 µM dNTPs and up to 250 ng of target DNA.

Subcloning and DNA Sequencing

Purified *Pfu* PCR products (40–60 ng) were ligated to M13mp19 or mp18 sequencing vectors digested with *Sma*I [9]. 3 U T4 DNA ligase (Promega Biotec, USA) were used in each of the ligation reactions performed according to the manufacturer's specifications. The recombinant DNA was transformed using *Escherichia coli* TG-1, and sequencing was performed using the dideoxynucleotide chain termination method [10] and Sequenase (Amersham International, UK). The

Table 2. Predicted MW's and pI's of genomic Der p 3 and Der f 3

	pI	MW, kD
Der p 3 P3WS1	8.5	24.99
Der p 3g.6	6.11	25.12
Der f 3g	5.47	25.27

oligonucleotide primers used for sequencing included universal primers and those listed in table 1. The primers were used to sequence the M13 clones in both directions.

DNA and Protein Sequence Analysis

Sequence analysis was performed using MAC VECTOR software (IBI, 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 (December 15, 1994).

Results

Nucleotide sequence

Der p 3 genomic DNA was amplified by PCR using oligonucleotides 4 and 8, which span residues 1–20 and 769–789, respectively (table 1, fig. 1). The PCR product was a single band of approximately 790 bp as determined by agarose gel electrophoresis. Der f 3 genomic DNA was amplified using the same set of primers, as a single band of approximately 770 bp in size. DNA from the Der p 3 and Der f 3 bands was purified and ligated into M13 sequencing vectors. Three Der p 3g M13mp19 clones were sequenced. Clone 3 (Der p 3g.3) and 5 (Der p 3g.5) had 100% identity with the original Der p 3 P3WS1 cDNA nucleotide sequence. Clone 6 (Der p 3g.6) differed at 6 positions, as illustrated in figure 1. The two Der f 3g M13mp19 clones were sequenced and shown to have 86% identity at the nucleotide level with the P3WS1 cDNA sequence. The two sequences were identical in all but two nucleotides. Der f 3g.2 had a primer-derived sequence, for nucleotide residues 1–20, whereas Der f 3g.1 had a two-base substitution at residues 14–15 with AA being substituted for TG.

All 5 genomic clones contained a single intron inserted between the amino acid residues 132 and 133. The size of the intron varied between the clones, but all contained the conserved consensus sequences at the splice junction sites, i.e. 5' GT...AG 3' (as underlined in fig. 1) Both Der f 3g clones had identical 60 bp introns whereas Der p 3g.3 had an 85, Der p 3g.5 an 86 and Der p 3g.6 an 87 bp intron.

Protein Sequence Alignments

The deduced amino acid sequence for the Der f 3.1g clone differs from the previously derived N-terminal protein sequence of the natural Der f 3 protein at one out of thirty-one residues where there is a conservative Ser to Cys (unpaired) substitution at residue 12 [6]. Residue 25 in the natural protein sequence was indeterminable, as is often the case for Cys residues (fig. 2). The Der f 3 genomic sequence does in fact contain a Cys at residue 25. The Der f 3 genomic DNA sequence also has a high degree of identity with two cyanogen bromide (CNBr)-generated peptides of the natural Der f 3 protein [6]. Both peptides differ at only one amino acid out of eighteen (peptide 1) and sixteen (peptide 2), with the substitutions occurring at residues 183 and 112, respectively.

A search of the GenBank protein sequence database confirmed that the Der f 3g clones were most homologous to the Der p 3 cDNA P3WS1 clone and, like P3WS1, have almost 50% homology with both vertebrate and invertebrate trypsins. Figure 3 displays the alignment of Der f 3g sequences with Der p 3g sequences and both mosquito and bovine trypsins.

All seven Cys residues are conserved between Der f 3 and the Der p 3 genomic sequences. The residues involved in the catalytic activity and in the substrate specificity of trypsin proteins are clearly conserved in all five protein sequences represented in figure 3, as are the sequences on either side of these residues. The calculated molecular weights (MWs) and pIs of Der f 3.1g and Der p 3g.6 were different from each other and from the original Der p 3 P3WS1 clone (table 2).

Discussion

The isolation and sequencing of Der f 3 genomic DNA represents the first primary-sequence determination for the group 3 allergens from the *D. farinae* mite. The deduced amino acid sequence for Der f 3g was almost identical to the N-terminal protein sequences determined from the natural Der f 3. There are, however, minor differences in the N-terminal sequences published by different laboratories. The N-terminal sequence of the protein isolated from *D. farinae* SMM by affinity purification using the Der f 3 5A12 monoclonal antibody column [6] was identical in twenty-nine of thirty determined residues, the difference being a Cys/Ser substitution at residue 12. As illustrated in Figure 2, N-terminal protein sequences of Der f 3 isolated by size exclusion chromatography [4] or by substrate affinity chromatography from *D. farinae* SMM [4, 11] are also identical

Der f 3g.1 **ATTGTTGGTGGTGTGAAAGCAAAAGCAGGTGATGTCCATATCAAATTTCCCTTGCAATCAAGCAGCCAT** 69
Der f 3g.2**AA**.....
Der p 3g.3**AA**.....TT...T...A.....G.....A..A.....T..T...
Der p 3g.5**AA**.....TT...T...A.....G.....A..A.....T..T...
Der p 3g.6**AA**.....TT...T...A.....G.....A..A.....T..T...

Der f 3g.1 **TTTGTGGTGGTAGTATCCTGGATGAATATTGGATCTTGACCGCTGCACATTGTGTCAATGGGTAAGT** 141
Der f 3g.2
Der p 3g.3 C...T..T.....T..A..A.....TGCC.....G
Der p 3g.5 C...T..T.....T..A..A.....TGCC.....G
Der p 3g.6 C...T..T.....T..A..A.....TGCC.....G

Der f 3g.1 **TT-TTG-----TAAATTTTGGACTTGA--TAAATCATCTCG-----CA--CAAAAAAAC** 213
Der f 3g.2
Der p 3g.3 ..G...TTGAC.....GAAT..A..CA..AA...TAAT..T..TTTT-GTTGTAATTTTTT...-TTG...TTT
Der p 3g.5 ..G...TTGAC.....GAAT..A..CA..AA...TAAT..T..TTTTGTTGTAATTTTTT...-TTG...TT
Der p 3g.6 ..G...TTGAC.....GAAT..A..CA..AA...TAAT..T..TTTTGTTGTAATTTTTTGT..ATTG...TT

Der f 3g.1 **CATAGACAATCAGCAAAAAAACTTTCAATTTCGTTACAATACTCTTAAACATGCATCTGGTGGTGAAAAGATT** 285
Der f 3g.2
Der p 3g.3T.....A.....GT.....G..T..A.....T..TA.....A...
Der p 3g.5T.....A.....GT.....G..T..A.....T..TA.....A...
Der p 3g.6T.....A.....GT.....G..T..A.....T..TA.....A...

Der f 3g.1 **CAAGTGGCGGAAATTTATCAACACGAAATTTATGATAGCATGACTATCGATAATGATGTTGCATTGATAAAA** 357
Der f 3g.2
Der p 3g.3TCT..T..TA.....T..GC...T...A.....TTATCAA..T.....A.....T..G
Der p 3g.5TCT..T..TA.....T..GC...T...A.....TTATCAA..T.....A.....T..G
Der p 3g.6TCT..T..TA.....T..GC...T...A.....TATCAA..T.....A.....T..G

Der f 3g.1 **CTCAAAACACCAATGACATTGGATCAAACAAATGCTAAACCGTACCATTGCCACCACAAGGATCAGATGTA** 429
Der f 3g.2
Der p 3g.3 ..T...T...T...A...AA...G..A.....C...G..T..TGG...A...G..A.....G.....
Der p 3g.5 ..T...T...T...A...AA...G..A.....C...G..T..TGG...A...G..A.....G.....
Der p 3g.6 ..T...T...G...A...A...G..A.....C...G..T..CTGG...A...G..A.....G.....

Der f 3g.1 **AAAGTTGGTGATAAAATTCGTGTTTCTGGTTGGGGTTATCTTCAGGAAGGAAGTTATTCATTACCATCGGAA** 501
Der f 3g.2
Der p 3g.3 C..G.....G..A.....T...
Der p 3g.5 C..G.....G..A.....T...
Der p 3g.6 C..G.....G..A.....T...

Der f 3g.1 **TTACAACGTGTTGATATTGATGTTGTATCACGTGAACAAATGTGACCAATTATATTCAAAGCAGGCGCCGAT** 573
Der f 3g.2
Der p 3g.3 ..AG.....C.....A..G...A..TG.....TAAT..T..A
Der p 3g.5 ..AG.....C.....A..G...A..TG.....TAAT..T..A
Der p 3g.6 ..AG.....G...CA..TG.....TAAT..T..A

Der f 3g.1 **GTTAGTGAAAAATATGATTTGCGCGGTGATGTCGCTAATGGTGGTGTGATTTCATGTCAAGGTGATCTGCG** 645
Der f 3g.2
Der p 3g.3 ..C...T.....T..T.....T..A.....AAA...T.....T
Der p 3g.5 ..C...T.....T..T.....T..A.....AAA...T.....T
Der p 3g.6 ..C...T.....T..T.....T..A.....AAA...T.....T

Der f 3g.1 **GGACCAGTTGTTGATATTGCCACTAAACAAATGTTGGTATTGTTTCATGGGGTTATGGTTGTGCACGTAAA** 717
Der f 3g.2
Der p 3g.3 ..G.....G...AAA..A...T...G.....
Der p 3g.5 ..G.....G...AAA..A...T...G.....
Der p 3g.6 ..G.....G...AAA..A...T...G.....

Der f 3g.1 **GGTTATCCAGGTGTCTATACACGTGTTGGTAATTTTGTGCGATTGGATTGAATCAAAACGGTTACAGTGA** 786
Der f 3g.2
Der p 3g.3 T.....A.....
Der p 3g.5 T.....A.....
Der p 3g.6 T.....A.....

Der f 3g.1 **TTG**
Der f 3g.2 ...
Der p 3g.3 ...
Der p 3g.5 ...
Der p 3g.6 ...

789

Fig. 1. Nucleotide sequences of the Der f 3 and Der p 3 genes. The sequences for the 2 Der f 3g clones and the 3 Der p 3g clones – Der p 3g.3, Der p 3g.5 and Der p 3g.6 – are aligned to show sequence identity. The underlined residues demonstrate the single intron found in all 5 sequences. The dots and the bold letters represent identical and primer-derived residues respectively (see text for differences in Der f 3g.1).

Der f3.1(g) IVVGVKAK A GDCPYQISLQSSSHFCG G SILD
Der f3 (p¹)[6] (.) X. (.) . . .
Der f3 (p²)[4] L . . . X . E . . .
Der f3 (p³)[11] Q . . . X
Der f3 (p⁴)[5] Q . . . X . . . - . -

		169		186
Der f3	(peptide 1)[6]		MI (C) GGDVANGGVDS (E) QGD	
Der f3	(g)		 C . . .	
		96		113
Der f3	(peptide 1)[6]		(M) TLDQTNA (K) PVPL (P) -- (T) (S)	
Der f3	(g)		 -- G . .	

Der f3g.1 IVGGVVKAKAGDCPYQISLQSSS-HFCGGSILDEYWILTAACHCVN-G-QSAKKLSIRYN TL 57
Der f3g.2 ::::E:::::-::::-:::::
Der p3g.5 ::E:L:E::::T::::A--:T:S::::S:
Der p3g.6 ::E:L:E::::-::T::::A--:T:S::::S:
Mosquito ::FEIDVS:A::V::YNKR:N::V:SSK:V::::TA--A:TSS:TV:LG:S
Bovine ::YTCG:NTV::V:N:GY-::::LINSQ:VVS::::YKS:I:VR--:QQD:-IN

*

Der f3g.1 KHASGGEK-IQVAE-IYQHENVDSMTIDNDVALMKLKTPM TLDQTNAKPVLPPQGSDVK 115
Der f3g.2 :::::::::::-:::::
Der p3g.5 ::SL::-:S:K-FA:K::YQ::I:I:S:K:N:K::A:G:AK::::
Der p3g.6 ::SL::-:S:K-FA:K::YQ::I:I:S:K::K::AA G:AK::::
Mosquito R::::TV-VR::R-VV:PK::SS:F:YS:LE:EDEL:FSDAV-Q:G:K:DET::
Bovine VVE:-NQOF:SASKS-V-PS:N:N:LN::IM:I::SAAS:NSRV:-SIS::-T:CAS

* *

Der f3g.1 VGDKIRVSGWG YLQE GSYSLPSELQRVDIDVVSREQCDQLYSKAGADVSENMICGGDVAN 175
Der f3g.2 :::::::::::-:::::
Der p3g.5 ::QV::::E::::R::A::KE:NE::::N:E:TD::::
Der p3g.6 ::QV::::E::::R::::E:NE::::N:E:TD::::
Mosquito D:TMTT:::NT:SAAE:NAV--RAANYPT:NOKE:NKA::DF:G--TDR:L:A:YQQ-
Bovine A:TQCLI:::NTKSSGT:Y:DV:KCLKAPIL:DSS:KSA:P--:QIT:-N:F:A:YLE-

* * *

Der f3g.1 GGVDSCQGD SGGPVV DIATK-QIVGI VSWGYGCARKGYPGVYTRVGNFVDWIESKR SQ* 233
Der f3g.2 :::::::::::-:::::
Der p3g.5 ::K::::::::::VNKN:-V::::I::::
Der p3g.6 ::K::::::::::VNKN:-V::::I::::
Mosquito :K:A::::L:-DG-KL:V::::QA::::S:AVVR:VRENSGV:
Bovine :K::::::::::CSG-L:-::::S::ONK::::K:C:Y:S:KOTIASN*

137

when compared over the twenty-residue N-terminal Der f 3.1g protein sequence. Both trypsin-like protease sequences have a non-conservative change at residue 8 from a positively charged Lys (Der f 3g) to a polar uncharged Glu (trypsin-like protease). The only other non-conservative difference between all the protein sequences is a change from a polar uncharged Glu at residue 15 in Der f 3g.1 to a negatively charged Glu in Der f 3 (p²) [4]. These changes may be a consequence of the different sources of *D. farinae* SMM used and may represent different allelic isoforms of the Der f 3 allergen. As will be discussed below, the Der f 3 sequence presented here has the structural characteristics of a trypsin-like protein.

The open reading frame (ORF) for both the Der p 3 and Der f 3 genes includes a single, small intron which interrupts the coding sequence between nucleotide residues 132–133. The small size of the introns correlates well with the size of those found in the group 1 allergens of the house dust mite [12]. Like the introns in the group 1 genes, they do contain the expected consensus sequences at the 5' and 3' intron splice sites. The minor differences in both the sequences and the sizes of the introns observed among the three Der p 3 genomic clones is possibly due to the existence of an allelic variation expected between individual mites.

The predicted Der f 3 mature protein sequence with a MW of 25.27 kD is larger than the predicted size for the P3WS1 protein (table 2) [6]. It is interesting to note that, although the predicted size difference is only 0.29 kD, our unpublished observations show that natural Der f 3 is normally 1–2 kD larger when compared to natural Der p 3 by SDS-PAGE analysis. The disparity between the apparent MW determined by SDS-PAGE and that calculated from the deduced amino acid is commonly observed. Most often, the calculated MW is smaller, possibly because the conformational characteristics of the protein such as disulphide bonds are not taken into consideration when calculating the MW. Der f 3g also has a more acidic pI value (5.47) than the Der p 3 P3WS1 clone (table 2). Like the Der p 3 allergen, it has been demonstrated by Stewart et al. [13] that Der f 3 exists in different isoforms. The Der f 3g pI value is close to the values predicted by Stewart et al. [14]. A comparative study of house dust mite of trypsin proteases performed by this group showed that trypsin-like protease from *D. pteronyssinus* was very heterogeneous with several major isoforms with pIs ranging from 4.8 to 8.6, whereas trypsin-like protease from *D. farinae* was less heterogeneous, with all isoforms having acidic pIs of approximately 4.8.

The identity between the amino acid sequences for the Der f 3g clone and P3WS1 is 81%, in keeping with the 20% variation in homology observed between members of the

group 1 [15] and group 2 house dust mite allergens. There are twenty-eight non-conservative and nineteen conservative substitutions between the protein sequences of the Der f 3 and Der p 3 P3WS1 clones. Our unpublished observations show that, although cross-reactivity exists, considerable antigenic differences between Der p 3 and Der f 3 can sometimes be demonstrated. While polyclonal mouse anti-Der f 3 and Der p 3 have readily been shown to be cross-reactive, polyclonal rabbit anti-Der f 3 reacted strongly with Der f 3, but not with Der p 3. Accordingly, the protein sequence does indeed contain regions which are highly conserved between the Der f 3 and Der p 3 proteins, for example, the C-terminal portion. Most of the sequence variation occurs in the central portion of the protein. The degree of cross-reactivity of different antisera of IgE in allergic sera could therefore vary depending on the epitopes recognised. Like the P3WS1 sequence, a search of the GenBank protein database confirmed that the Der f 3g protein sequence has a high degree of homology to both vertebrate and invertebrate trypsin proteins, i.e. 47% identity with *Anopheles gambiae* trypsin [16] and 42% identity with bovine trypsin [17]. As observed for most trypsins, these differences between Der f 3 and Der p 3 and the bovine and mosquito trypsins occur in regions of the protein not involved in the catalytic triad and substrate specificity. Most of the differences, therefore, occur in the central region of the protein sequence.

Another important feature of the Der f 3g protein sequence is the conservation of the seven Cys residues also found in the Der p 3 P3WS1 sequence with the role of the seventh unpaired Cys residue at residue 12 still undetermined [6]. Most of the N-terminal protein sequences listed in figure 2 showed an indeterminable residue at this position that was most often found to represent a Cys. Smith et al. [6] reported a Ser in this position, which implies that this Cys does not play an important structural role in the protein through disulphide bond formation as it is not conserved in all the transcribed isoforms of the protein. The remaining six residues are predicted to form the following disulphide bridges: 25–42, 154–171, and 183–209.

There is only one single base substitution between the two Der f 3 sequences with a Val at residue 5 in 3.1 being substituted for a Glu. The Val represents a non-conservative substitution between Der f 3 and Der p 3, whereas the Glu is identical to the residue found in the Der p 3 gene. The primer used to amplify the Der f 3 gene was designed using the original Der p 3 P3WS1 sequence (table 1) and encoded amino acids IVGGEK; so, while the Glu 5 in clone Der f 3.2 was expected, the Val 5 in Der f 3.1 was not. It is possible that, either because the oligonucleotides used were not purified by SDS-PAGE or HPLC or, alternatively, due to partial

degradation, an incomplete primer template annealed to the Der f 3 gene during the PCR process resulting in a residue at position 5 which was not primer derived, but was in fact coded for by the gene. This idea is supported by the fact that the N-terminal protein sequence of the natural Der f 3 protein also has a Val at this position (fig. 2). It should be noted here that the six C-terminal amino acid residues of both Der p 3 and Der f 3 sequences (fig. 3) are derived from the reverse Der p primer No. 8 (table 1).

A comparison of the coding sequences for the 3 Der p 3 genomic sequences shows that Der p 3.3 and Der p 3.5 are identical and are also identical to the original Der p 3 P3WS1 clone. The small degree of variation observed in the introns suggests that 3.3 and 3.5 simply represent the genomic variation one would expect to see between individual mites [18]. The Der p 3.6 clone, however, differs from these sequences with four substitutions, all of which are non-conservative, in the coding sequence, as well as some degree of variation in the intron sequence. The substitutions involve (1) a change from an Asn (P3WS1) to an Asp (3.6) at residue 98 (2), a change from a Val (P3WS1) to an Ala (3.6) at residue 105 (3), a change from an Ala (P3WS1) to an Asp (3.6) at residue 145, and (4) a change from a Lys (P3WS1) to a Glu (3.6) at residue 150. The mature protein encoded by Der p 3.6 has a MW of 25.12 kD and a pI value of 6.11 (table 2). It would appear, therefore, that Der p 3.6 represents an isoform, which would be classified as a variant of Der p 3 P3WS1. Several groups have also reported the isolation of polymorphic allergens, these include, for example, the group 1 [19] and group 2 [1, 20] house dust mite allergens. While the 3.6 clone maintains a very high degree of identity with the original clone Der p 3 P3WS1, the significance of these polymorphic changes needs to be assessed. Most eukaryotic genes have very minimal amino acid sequence variation in the coding region [18], and in fact, it has been demonstrated by Kreitman [21] that within eleven separate

isolates of the *Drosophila melanogaster* alcohol dehydrogenase gene, there was only a single amino acid substitution. Given that Pfu polymerase has 3'-5' exonuclease-editing function and therefore an error rate of 1.6×10^{-6} errors/base [22], the number of substitutions far exceeds that predicted by polymerase error. In addition, three of these four substitutions were to residues found in Der f 3. They may be important if they impact in any way on the 3-dimensional structure or on T cell recognition of the molecule. Some of the variations found in the Der p1 cDNA sequences do reside in the T cell epitopes [19], but the effect of these variations is as yet undetermined. The 2 Der f 2 isoforms identified by Nishiyama et al. [20] were equally recognised by IgE antibodies from a panel of fourteen different mite-allergic patients although differences in monoclonal antibody binding could be detected.

In conclusion, we have isolated and sequenced the DNA encoding the Der f 3 allergen to show that, like its homologue Der p 3, it is a trypsin-like protease. Similar analysis of *D. pteronyssinus* genomic DNA resulted in the isolation of a variant sequence for Der p 3 and of two sequences identical to the original cDNA clone [6]. From our observations, we conclude that the trypsin-like protein of the *Dermatophagoides* house dust mite are well conserved. While a number of isoforms of Der p 3 have been described by isoelectric focusing [13, 14], our data show that considerable variation in the calculated pI value can be produced by only a few amino acid changes. For measuring T cell responses to purified proteins and peptide, it will be important to consider the intra- and inter-species sequence differences described.

Acknowledgments

This work was supported by the National Health and Medical Research Council of Australia.

References

- 1 Thomas WR, Chua KY, Smith WA: Molecular polymorphisms of house dust mite allergens. *Exp Appl Acarol* 1992;16:153-164.
- 2 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.
- 3 Yasueda H, Mita H, Shida T, Ando T, Sugiyama S, Yamakawa H: Allergens from *Dermatophagoides* mites with chymotryptic activity. *Clin Exp Allergy* 1993;23:384-390.
- 4 Heymann PW, Chapman MD, Aalberse RC, Fox JW, Platts-Mills T: 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;83:1055-1067.
- 5 Ando T, Homma R, Ino Y, Ito G, Miyahara A, Yanagihara T, Kimura H, Ikeda S, Yamakawa H, Iwaki M, Okumura Y, Suko M, Haida M, Okudaira H: Trypsin-like protease of mites - purification and characterisation of trypsin-like protease from mite faecal extract *Dermatophagoides farinae*. Relationship between trypsin-like protease and *Der f* III. *Clin Exp Allergy* 1993;23:777-784.

- 6 Smith WA, Chuna KY, Kuo MC, Rogers BL, Thomas WR: Cloning and sequencing of the *Dermatophagoides pteronyssinus* group III allergen, *Der p* III. Clin Exp Allergy 1994;24: 220–228.
- 7 Stewart GA, Thompson PJ, Simpson RJ: Protease antigens from house dust mite. Lancet 1989;II:154–155 (plus corrections 462).
- 8 Stewart GA, Thomas WR: In vitro translation of messenger RNA from the house dust mite *Dermatophagoides pteronyssinus*. Int Arch Allergy Appl Immunol 1987;83:384–389.
- 9 Messing J: New M13 vectors for cloning. Methods Enzymol 1983;101:20–78.
- 10 Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 1977;74:5463–5467.
- 11 Kohmoto S, Kodera Y, Takahashi K, Nishimura H, Matsushima A, Inada Y: Activation of the kallikrein-kinin system in human plasma by a serine protease from mites. J Clin Biochem Nutr 1991;10:15–20.
- 12 Kent NA, Hill MR, Keen JN, Holland PWH, Hart BJ: Molecular characterisation of the group I allergen *Eur m* I from house dust mite *Euroglyphus maynei*. Int Arch Allergy Immunol 1992;99:150–152.
- 13 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.
- 14 Stewart GA, Kollinger MR, King CM, Thompson PJ: A comparative study of three serine proteases from *Dermatophagoides pteronyssinus* and *D. farinae*. Allergy 1994;49:553–560.
- 15 Dilworth RJ, Chua KY, Thomas WR: Sequence analysis of cDNA coding for a major house dust mite allergen *Der f* I. Clin Exp Allergy 1991;21: 25–32.
- 16 Muller HM, Vizioli I, della Torre A, Crisanti A: Temporal and spatial expression of serine protease genes in *Anopheles gambiae*. Parasitologia 1993;35:73–76.
- 17 Hartley BS: Homologies in serine proteinases. Philos Trans R Soc Lond B Biol Sci 1970;257: 77–87.
- 18 Bowcock A, Cavalli-Sforza L: The study of variation in the human genome. Genomics 1991; 11:491–498.
- 19 Chua KY, Kehal PK, Thomas WR: Sequence polymorphisms of cDNA clones encoding the mite allergen *Der p* I. Int Arch Allergy Immunol 1993;101:364–368.
- 20 Nishiyama C, Yuuki T, Usui Y, Iwamoto N, Okumura Y, Okudaira H: Effects of amino acid variations in recombinant *Der f* II on its human IgE and mouse IgG recognition. Int Arch Allergy Immunol 1994;105:62–69.
- 21 Kreitman M: Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature 1983;304:412–417.
- 22 Lundberg KS, Shoemaker DD, Adams MW, Short JM, Sorge JA, Mathur EJ: High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. Gene 1991;108:1–6.