

RAPID PUBLICATION

cDNA encoding the major mite allergen *Der f* II

M. TRUDINGER, K. Y. CHUA and W. R. THOMAS

Western Australia Research Institute for Child Health, Princess Margaret Hospital, Perth, Western Australia

Summary

cDNA encoding the major house dust mite allergen *Der f* II from *Dermatophagoides farinae* was amplified using the polymerase chain reaction and cloned into *E. coli*. It encoded a 129-residue protein with a calculated molecular weight of 14021 D and had the expected high homology (88%) with *Der p* II including the absence of *N*-glycosylation sites and conserved cysteine residues. These results are consistent with the high degree of antibody crossreactivity and may help identify the differences in T-cell epitopes revealed for these molecules so far.

Clinical and Experimental Allergy, Vol. 21, pp. 33–37. Submitted 12 July 1990; revised 23 August 1990; accepted 13 September 1990.

Introduction

Recent reports have documented the importance of responses to the group I and group II allergens in house dust mite allergy. For example, it has been documented that over 60% of patients have more than 50% of their anti-mite antibodies directed towards these proteins [1–3]. It is possible that children show a greater degree of reactivity [4]. To help understand the regulation of responses to these proteins or determine T- and B-cell epitopes, it is necessary to obtain structural information. It is also useful to be able to produce the allergens, or fragments of the allergens, in large quantities for epitope studies or for stimulating T-cell cultures to determine their role in the inflammatory response found in diseases such as asthma. To this end, cDNA clones encoding *Der p* I and *Der p* II of *Dermatophagoides pteronyssinus* have been produced and analysed [5–7]. These mites are the most common *Dermatophagoides* species in house dust in Western Europe and Australia, while the species *D. farinae* predominates in other countries, for example North America and Japan [8]. A cDNA clone encoding the major *D. farinae* allergen *Der f* II is now described.

Materials and methods

Preparation of lambda gt11 *D. farinae* cDNA ligations

D. farinae was purchased from Commonwealth Serum Laboratories, Parkville, Australia and used to prepare

mRNA (polyadenylated RNA) as described previously [9]. The mRNA was suspended at approximately 0.5 µg/µl and 5 µl used to prepare cDNA by the RNase H method [10] using a kit (Amersham International, Aylesbury, U.K.) EcoRI linkers (Amersham, GGAATTCC) were attached according to [11]. The DNA was then digested with EcoRI and recovered from an agarose gel purification by electrophoresis into a DEAE membrane (Schleicher and Schuell, Dassel, Germany; NA-45) according to protocol 6.24 of Sambrook *et al.* [12] except 0.5 M arginine base was used for elution. The cDNA was then ligated in lambda gt10 and gt11 at an arms to insert ratio of 2:1. Some was packaged for plaque libraries and an aliquot was retained for isolating sequences by polymerase chain reaction as described below.

Isolation of *Der f* II cDNA by polymerase chain reaction (PCR)

To isolate *Der f* II cDNA an oligonucleotide primer based on the N-terminal sequence of *Der p* II was made because their amino-acid residues are identical in these regions [2]. The primer GGATCCGATCAAGTCGATGT-3' was used. The first GGATCC encodes a BamHI site and the following sequence GAT... encodes the first four residues of *Der p* II. For the other primer the lambda gt11 TTGACACCAGACCAACTGGTAATG-3' reverse primer flanking the EcoRI cloning site was used (New England Biolabs, Beverly, U.S.A.). The *Der p* II primer was designed to have approximately 50–60% G–C content and to end on the second base of a codon to avoid the redundancy of third bases [13,14]. The PCR reactions

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

																10								
GAT	CAA	GTC	GAT	GTT	AAA	GAT	TGT	GCC	AAC	AAT	GAA	ATC	AAA	AAA	GTA	ATG	51							
Asp	Gln	Val	Asp	Val	Lys	Asp	Cys	Ala	Asn	Asn	Glu	Ile	Lys	Lys	Val	Met								
																20		30						
GTC	GAT	GGT	TGC	CAT	GGT	TCT	GAT	CCA	TGC	ATC	ATC	CAT	CGT	GGT	AAA	CCA	102							
Val	Asp	Gly	Cys	His	Gly	Ser	Asp	Pro	Cys	Ile	Ile	His	Arg	Gly	Lys	Pro								
																40		50						
TTC	ACT	TTG	GAA	GCC	TTA	TTC	GAT	GCC	AAC	CAA	AAC	ACT	AAA	ACC	GCT	AAA	153							
Phe	Thr	Leu	Glu	Ala	Leu	Phe	Asp	Ala	Asn	Gln	Asn	Thr	Lys	Thr	Ala	Lys								
																60								
ACT	GAA	ATC	AAA	GCC	AGC	CTC	GAT	GGT	CTT	GAA	ATT	GAT	GTT	CCC	GGT	ATT	204							
Thr	Glu	Ile	Lys	Ala	Ser	Leu	Asp	Gly	Leu	Glu	Ile	Asp	Val	Pro	Gly	Ile								
																70		80						
GAT	ACC	AAT	GCT	TGC	CAT	TTT	ATG	AAA	TGT	CCA	TTG	GTT	AAA	GGT	CAA	CAA	255							
Asp	Thr	Asn	Ala	Cys	His	Phe	Met	Lys	Cys	Pro	Leu	Val	Lys	Gly	Gln	Gln								
																90		100						
TAT	GAT	GCC	AAA	TAT	ACA	TGG	AAT	GTG	CCG	AAA	ATT	GCA	CCA	AAA	TCT	GAA	306							
Tyr	Asp	Ala	Lys	Tyr	Thr	Trp	Asn	Val	Pro	Lys	Ile	Ala	Pro	Lys	Ser	Glu								
																110								
AAC	GTT	GTC	GTT	ACA	GTC	AAA	CTT	GTT	GGT	GAT	AAT	GGT	GTT	TTG	GCT	TGC	357							
Asn	Val	Val	Val	Thr	Val	Lys	Leu	Val	Gly	Asp	Asn	Gly	Val	Leu	Ala	Cys								
																120		129						
GCT	ATT	GCT	ACC	CAC	GCT	AAA	ATC	CGT	GAT	<u>TAAAAA</u>	<u>AAAAA</u>	<u>ATAAT</u>	<u>TGAAAA</u>	<u>ATT</u>	<u>GAAAA</u>	<u>ATT</u>	414							
Ala	Ile	Ala	Thr	His	Ala	Lys	Ile	Arg	Asp															
																TTACCAACATCGAACAAATTCAATAACCAAAATTTGAATCAAAACGGAATTC	CAAGCTGAGCGC							481
																CGGTCGCTAC							491	

Fig. 1. Nucleotide and inferred amino-acid sequence of *Der f* II cDNA. Numbers to the right are nucleotide positions and numbers above are amino-acid residues. The stop (TAA) signal is underlined. The first eight nucleotides are from the oligonucleotide primer used to generate the cDNA, based on the *Der p* II sequence.

were carried out in a final reaction volume of 25 μ l containing 67 mM Tris-HCl (pH 8.8 at 25°C), 16.6 mM $(\text{NH}_4)_2\text{SO}_4$, 40 μ M dNTPs, 5 mM 2-mercaptoethanol, 6 μ M EDTA, 0.2 mg/ml gelatin, 2 mM MgCl_2 , 10 pmol of each primer and 2 units of Taq polymerase. This was obtained as a kit from Biotech International Ltd (Perth, Australia). Target DNA (0.001 μ g) was added and the contents of the tube were mixed and overlayed with paraffin oil. The tubes were initially denatured at 95°C for 6 min, then annealed at 55°C for 1 min and extended at 72°C for 2 min. Thereafter for 38 cycles, denaturing was carried out for 30 sec and annealing and extension as before. In the final (40th) cycle the extension reaction was increased to 10 min to ensure that all amplified products were full length. Five microlitres of the reaction were then checked for amplified bands on a 1% agarose gel. The remainder of the reaction mixture was extracted with chloroform to remove the paraffin oil and ethanol precipitated prior to purification of the amplified product on a low melting point agarose gel (Bio-Rad, Richmond, U.S.A.).

The ends of the purified PCR product were filled in with Klenow enzyme [12] and the resulting blunt-ended DNA was ligated into M13mp18 digested with Sma I and then transferred into *E. coli* strain TGI competent cells. Isolated white plaques were picked and used to prepare phage stocks and single-stranded DNA by PEG precipitation [12].

DNA and sequence analysis

DNA sequencing was performed with the dideoxynucleotide chain termination [15] using DNA produced from M13-derived vectors mp18 and 19 in *E. coli* strain TGI and Sequenase version 2.0 (USB Corp., Cleveland, U.S.A.). Restriction endonucleases were from Toyobo (Osaka, Japan). All procedures were by standard techniques [12] and the sequence analysis performed using the Mac VECTOR software (IBI, New Haven, U.S.A.). Hydrophobicity plots of *Der f* II and *Der p* II were determined by the Hopp and Woods algorithm [16] computed with the Mac VECTOR sequence analysis software using a six-residue window.

Results

D. farinae cDNA ligated into lambda gt11 vector was used to amplify a sequence using an oligonucleotide primer with homology to nucleotides coding for the four N-terminal residues of *Der p* II and a reverse primer for the lambda gt11 sequence flanking the coding site. Two bands of about 500 b.p. and 350 b.p. were obtained when the product was electrophoresed. These were ligated into m13 mp18 and a number of clones containing the 500 b.p. fragment were analysed by DNA sequencing. Three clones produced sequence data from the N-terminal primer and one clone produced sequence data from the

Dp II:	CACAAATTCTTCTTTCTTCTTACTACTGATCATTAACTGAAAACAAAACCAAAACCAT	63
	-16 -10	
Dp II:	TCAAAATGATG TAC AAA ATT TTG TGT CTT TCA TTG TTG GTC GCA GCC GTT	113
	Met Tyr Lys Ile Leu Cys Leu Ser Leu Leu Val Ala Ala Val	
	-1 1 10	
Dp II:	GCT CGT GAT CAA GTC GAT GTC AAA GAT TGT GCC AAT CAT GAA ATC AAA	161
	Ala Arg Asp Gln Val Asp Val Lys Asp Cys Ala Asn His Glu Ile Lys	
Df II:	T C A..	42
	Asn	
	20 30	
Dp II:	AAA GTT TTG GTA CCA GGA TGC CAT GGT TCA GAA CCA TGT ATC ATT CAT	209
	Lys Val Leu Val Pro Gly Cys His Gly Ser Glu Pro Cys Ile Ile His	
Df II:	A A.. ..C GAT ..T T ..T C C ...	90
	Met Asp Asp	
	40	
Dp II:	CGT GGT AAA CCA TTC CAA TTG GAA GCC GTT TTC GAA GCC AAC CAA AAC	257
	Arg Gly Lys Pro Phe Gln Leu Glu Ala Val Phe Glu Ala Asn Gln Asn	
Df II:	 ACT T.A T	138
	Thr Leu Asp	
	50 60	
Dp II:	ACA AAA ACG GCT AAA ATT GAA ATC AAA GCC TCA ATC GAT GGT TTA GAA	305
	Thr Lys Thr Ala Lys Ile Glu Ile Lys Ala Ser Ile Asp Gly Leu Glu	
Df II:	..T C C AGC C.. C.T ...	186
	Thr Leu	
	70	
Dp II:	GTT GAT GTT CCC GGT ATC GAT CCA AAT GCA TGC CAT TAC ATG AAA TGC	353
	Val Asp Val Pro Gly Ile Asp Pro Asn Ala Cys His Tyr Met Lys Cys	
Df II:	A.. T A.C T TT T	234
	Ile Thr Phe	
	80 90	
Dp II:	CCA TTG GTT AAA GGA CAA CAA TAT GAT ATT AAA TAT ACA TGG AAT GTT	401
	Pro Leu Val Lys Gly Gln Gln Tyr Asp Ile Lys Tyr Thr Trp Asn Val	
Df II:	T GCC G	282
	Ala	
	100 110	
Dp II:	CCG AAA ATT GCA CCA AAA TCT GAA AAT GTT GTC GTC ACT GTT AAA GTT	449
	Pro Lys Ile Ala Pro Lys Ser Glu Asn Val Val Val Thr Val Lys Val	
Df II:	C T ..A ..C C..	330
	Leu	
	120	
Dp II:	ATG GGT GAT GAT GGT GTT TTG GCC TGT GCT ATT GCT ACT CAT GCT AAA	497
	Met Gly Asp Asp Gly Val Leu Ala Cys Ala Ile Ala Thr His Ala Lys	
Df II:	G.T A.. T ..C C ..C	378
	Val Asn	
	129	
Dp II:	ATC CGC GAT TAA ATCAACAAAATTATTGATTTGTAATCACAATGATTGATTTCTT	557
	Ile Arg Asp END	
Df II:	T AA...A...TAAATA...AAA.T.TCA.CA.C...CGAAC.AAA.TCA	438
Dp II:	TCCAAAAAATAAATAAAATTTTGGGAATTC	591
Df II:	ATA.CC...TTTG...TC...AC__GGAATTC	469

Fig. 2. Alignment of *Der f* II and *Der p* II cDNA sequences. Numbers to the right are nucleotide position and numbers above are amino-acid residues. The top line gives the *Der p* II nucleotide sequence and the second the *Der p* II amino-acid residues. The next two lines show differences of *Der f* II to these sequences.

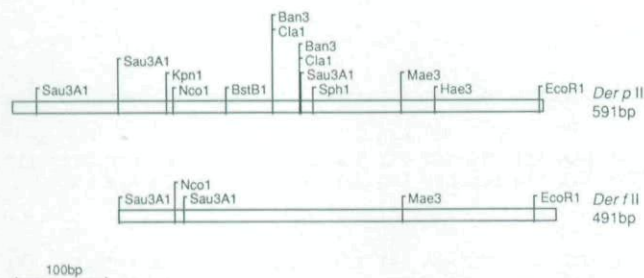


Fig. 3. Restriction map of *Der f* II cDNA. This was generated by computer from the sequence data as an aid to subcloning. A map of *Der p* II similarly generated is shown for comparison (there are few common restriction enzyme sites conserved). Sites marked with an asterisk were introduced by cloning procedures.

other orientation. Where the sequence data from the two directions overlapped a complete match was found. One of the clones, read from the N-terminal primer, contained a one-base deletion which shifted the reading frame. It was deduced to be a copying error (either in the reverse transcription of the mRNA to cDNA or in the amplification of the sequence by Taq polymerase) as the translated sequence from the other two clones matched the protein sequence available for the first 20 amino-acid residues of the allergen.

The sequence of the clones showing consensus and producing a correct reading frame is shown in Fig. 1 along with the inferred amino-acid sequence. It coded for a 129 residue protein with no *N*-glycosylation site and a calculated molecular weight of 14021 D. No homology was found when compared to other proteins on the GenBank database (61.0 release). It did, however, show 88% amino-acid residue homology with *Der p* II shown in

the alignment in Fig. 2. Seven out of the 16 changes were conservative. The conserved residues also included all the cysteines, that is at positions 8, 21, 27, 73, 78 and 119. There was also considerable nucleotide homology, although the restriction enzyme map generated from the sequence data for commonly used enzymes is different from *Der p* II (Fig. 3). The hydrophobicity plots of the translated sequence of *Der f* II and *Der p* II shown in Fig. 4 are almost identical.

Discussion

Although considerable information about group I and group II allergens from *D. pteronyssinus* has been obtained, similar data on *D. farinae* is only just becoming available. Because *D. farinae* is widespread, *Der f* II is an important allergen and may have interesting crossreactivity at the T- and/or B-cell level with *D. pteronyssinus*, particularly in areas where both species coexist. Although considerable serological crossreactivity between *Der f* II and *Der p* II has been reported [1,2] the only T-cell clones studied to date show no crossreactivity [17,18]. These, however, have been from the one donor, so further studies are indicated. The sequence differences responsible for these non-crossreacting epitopes may, however, be easy to locate because the data presented here show a high degree of homology between these group II allergens.

By using the PCR strategy described it has been possible to deduce the complete amino-acid sequence of mature *Der f* II and compare it to *Der p* II. The homology was 88%, which is higher than the 78% homology found between *Der p* I and *Der f* II [19] ($P < 0.01$; chi-square test). The reason for this may simply be that the group I

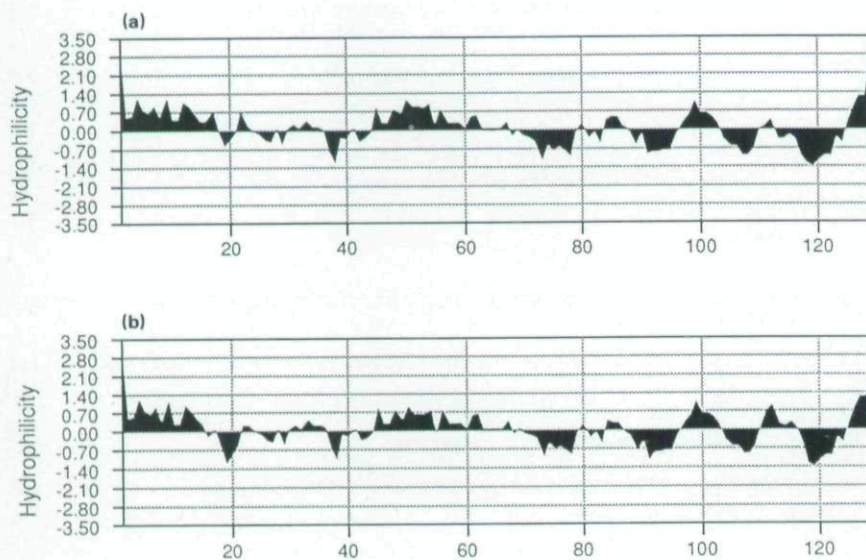


Fig. 4. Hydrophilicity plots of *Der f* II (a) and *Der p* II (b) using the Hopp and Woods algorithm [20] computed with the Mac VECTOR sequence analysis software (IBI, New Haven, U.S.A.) using a six-residue window.

allergens are larger and each residue may be less critical for the structure and function of the molecule. It is known, for example, that assuming they adopt a similar conformation to other cysteine proteases [20] many of the amino-acid differences in *Der p I* and *Der f I* lie in residues linking the two domain structures of the molecule [19]. The six cysteine molecules are also conserved between the group II allergens, suggesting a similar disulphide bonding, although this may be expected given the high overall homology. The homology of the amino-acid sequence can also be seen by the near identity of the hydrophilicity plots.

Another indication of the conservation of these proteins is that 34/55 of the nucleotide changes of the coding sequence are in the third base of a codon, which usually does not change the amino acid. Residues that may be of importance in the function of the molecule are Ser 57 where all three bases are changed but the amino acid is conserved and residue 88 where a complete codon change has conserved a small aliphatic residue. Like *Der p II* the *Der f II* cDNA clone does not have a poly A tail although the 3' non-coding region is rich in adenosine and has a polyadenylation signal. The nucleotides encoding the first four residues are from the PCR primer which was designed from the known identity of the *Der p II* and *Der f II* N-terminal amino acids [2]. A primer based on the C-terminal sequence can now be used to determine these bases coding for these conserved residues as well as the signal sequence using the same strategy reported here. The amino-acid sequence deduced here is, however, the complete sequence of mature *Der f II* and can be used for studying immune responses to peptides and expressing *Der f II* in bacteria and other hosts.

Acknowledgments

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

References

- Lind P, Weeke B, Lowenstein H. A reference allergen preparation of the house dust mite *D. pteronyssinus* produced from whole mite culture—a part of the DAS 76 study. *Allergy* 1984; 39:259–74.
- 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 mite *Dermatophagoides spp.* *J Allergy Clin Immunol* 1989; 83:1055–87.
- van der Zee JS, van Swieten P, Jansen HM, Aalberse RC. Skin tests and histamine release with P1-depleted. *Dermatophagoides pteronyssinus* body extracts and purified P1. *J Allergy Clin Immunol* 1988; 81:884–96.
- Thompson PJ, Whitely HJ, Naysmith JD, Carswell F. IgE antibodies to *D. pteronyssinus* in atopic patients. *Immunology* 1988; 64:311–14.
- Thomas WR, Stewart GA, Simpson RJ, et al. Cloning and expression of DNA coding for the major house dust mite allergen *Der p I* in *Escherichia coli*. *Int Arch Allergy Appl Immunol* 1988; 85:127–9.
- Chua KY, 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.
- Chua KY, Doyle CR, Simpson RJ, Turner KJ, Stewart GA, Thomas WR. 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.
- Wharton GW. House Dust Mites. *J Med Entomol* 1976; 12:577–621.
- Stewart GA, Thomas WR. *In vitro* translation of messenger RNA from the house dust mite *Dermatophagoides pteronyssinus*. *Int Arch Allergy Appl Immunol* 1987; 83:384–9.
- Gubler U, Hoffman BJ. A simple and very efficient method for generating cDNA libraries. *Gene* 1983; 25:263–9.
- Huynh TV, Young RA, Davis RW. Constructing and screening cDNA libraries in gt10 and gt11. In Glover ed. *DNA cloning*, Vol. 2. A practical approach. Oxford: IRL Press, 1985: 47–78.
- Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning. A Laboratory Manual*. 2nd Edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989.
- Gould SJ, Subamani S, Scheffler IE. Use of the DNA polymerase chain reaction for homology probing: isolation of partial cDNA or genomic clones encoding the iron-sulphur protein of succinate hydrogenase from several species. *Proc Natl Acad Sci U.S.A.* 1989; 86:1934–8.
- Summer R, Tautz D. Minimal homology requirements for PCR primers. *Nucleic Acid Res* 1989; 17:6749–53.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci U.S.A.* 1977; 74:5463–7.
- Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U.S.A.* 1981; 78:3824–8.
- O'Hehir RE, Young DB, Kay AB, Lamb JR. Cloned human T lymphocytes reactive with *Dermatophagoides farinae* (house dust mite): a comparison of T and B cell antigen recognition. *Immunology* 1987; 62:635–40.
- Lamb JR, Kay AB, O'Hehir RE. HLA class II restriction specificity of *Dermatophagoides spp* reactive T lymphocyte clones that support IgE synthesis. *Clin Exp Allergy* 1989; 19:389–93.
- Dilworth RJ, Chua KY, Thomas WR. Sequence analysis of cDNA coding for the major house dust mite allergen *Der f I*. *Clin Exp Allergy* 1990; 21:25–32.
- Baker E, Drenth J. The thiol proteases structure and mechanism. In: Jurnak FA, McPherson A, eds. *Biological macromolecules and assemblies*, Vol. 3. London: John Wiley and Sons, 1983: 313–68.

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.