

K.Y. Chua
P.K. Kehal
W.R. Thomas

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

Sequence Polymorphisms of cDNA Clones Encoding the Mite Allergen *Der p I*

Key Words

Sequence polymorphisms
Der p I
House dust mite
Dermatophagoides
cDNA

Abstract

Sequence information has been obtained from 5 cDNA clones encoding the major house dust mite allergen *Der p I*, including one which codes for the full length preproenzyme form of the molecule. All translated sequences were unique with 1–3 differences between each clone. 4/5 of the substitutions found were to residues found in *Der f I* and two positions had a substitution in more than one clone. The polymorphisms included residues already described to be in T cell epitopes, so as well as being necessary to consider them in the construction of synthetic allergens, the substitutions will be important for the interpretation of experimental and genetic studies.

Introduction

The analysis of cDNA-encoding allergens has already made a significant contribution to defining allergens and providing material and information for epitope and T cell studies. It is anticipated that its usefulness could extend into developing reagents to improve diagnosis and immunotherapy. A cDNA sequence encoding the major allergen *Der p I* from *Dermatophagoides pteronyssinus* was reported in 1988 [1], and later in a revised form [2], which identified it as a cysteine protease. The sequences of the related allergen *Der f I* from *Dermatophagoides farinae* was then determined and shown to be 81% homologous to *Der p I*, which was significantly less than that found for the comparison between *Der p II* and *Der f II* [3–6]. The existence of these sequences has highlighted the importance of determining the polymorphism in the allergen not only for further applications of recombinant and related technologies but also for our understanding of mite

allergy. Sequence polymorphism could for example have a profound influence on the genetics of allergic sensitization and T cell cross-reactivity. Here we describe sequences from 5 cDNA clones encoding *Der p I*, which although they support the veracity of our previous data also show a degree of polymorphism which could be of considerable importance. In addition we also report the cloning of a cDNA encoding the full preproenzyme form of the allergen.

Materials and Methods

cDNA Library

Poly A⁺ RNA was isolated from live mites (Commonwealth Serum Laboratories, Melbourne, Australia) [7] and cDNA prepared using the S1 nuclease method described by Huynh et al. [8]. It was ligated using *EcoRI* linkers into lambda gt10 and cloned into *Escherichia coli* Y 1073 to select against parental phage. The c600 *E. coli* JP 777 was used for hybridization and growing stocks. Details are described in Bowtell [9].

Table 1. Nucleotide changes in *Der p* I cDNA clones

Clone	Residue with polymorphic codon					
	50	81	124	136	149	215
a	CAC (H)	GAA (E)	GCA (A)	AGC (S)	GCA (A)	GAA (E)
b	NA	NA	NA	NA	NA	GAA (E)
c	TAC (Y)	GAA (E)	GTA (V)	AGC (S)	GCA (A)	GAA (E)
d	TAC (Y)	AAA (K)	GTA (V)	AGC (S)	GCA (A)	GAA (E)
e	CAC (H)	GAA (E)	GCA (A)	ACC (T)	GCT (A)	CAA (Q)

The table shows the codon and amino acid residue in parentheses which differed in the cDNA clones. NA = Not available because clone was truncated.

Screening of Libraries

Phages were plated in JP 777 at 20,000 PFU/17.5 cm petri dish and phage lifts made on nitrocellulose filters (Schleicher & Schull, Dassel, FRG). The DNA was denatured with 0.1 M NaOH and the filters baked at 80 °C in vacuo for 2 h and prehybridized for 2 h at 42 °C. Hybridization was performed with 1× Denhardt, 50% formamide at 42 °C overnight and the filters were washed sequentially in 0.1% SDS containing 2, 0.5, and 0.1 × SSC at room temperature for 15 min and a final wash of 0.1 × SSC in 1% SDS at 50 °C for 30 min. The DNA probe used was the 857 bp insert from the *Der p* I cDNA clone 13T [1] labelled by nick translation by ³²P-dATP (Amersham Australia, North Ryde). General procedures were according to Sambrook et al. [10].

DNA Preparation and Sequencing

Plaques showing positive hybridization were purified using two rounds of selection and the DNA prepared from a plate stock by a PEG/NaCl precipitation [1]. The DNA insert released by *Eco*RI digestion was cloned into the M13-derived vector mp 18 which was used to transform *E. coli* TG-1. Single-stranded dideoxynucleotide sequencing chain termination sequencing was performed using Sequenase 2.0 (UBS, Cleveland, Ohio, USA). Sequence analysis was performed with the assistance of Mac Vector software (IBI, New Haven, Conn., USA).

Results

Five cDNA clones including the 13T reported previously were sequenced. One clone was 1267 bp and contained the complete coding sequence for the preproenzyme form of *Der p* I (fig. 1). It closely resembled the *Der f* I clone from *D. farinae* reported previously [2] and contained nucleotides encoding a putative leader peptide from residues -98 to -79 and a proenzyme sequence from -80 to -1. The assignment of these regions is based on the known N-terminal sequence of *Der p* I isolated from house dust mite and the known preproenzyme structure of other cysteine proteases compared by Dilworth et al. [2]. The translated sequence from -98 to -80 conforms with a typical leader sequence although the exact beginning of the

proenzyme can only be estimated. Three of the other clones began at an internal *Eco*RI site at the nucleotides encoding residue -25 shown for clone in figure 1. The fifth clone was truncated and began at nucleotides encoding residue 151. In all there were only 6 nucleotide transitions between the clones (table 1). Five of these led to changes in the translation of the mature form of the allergen. No differences were found in the 3' untranslated regions except for differences in the size of the poly A tails. All of the translated sequences of the clones were unique and are aligned in figure 2 relative to the previously published 13T clone (a) and the sequence of *Der f* I. 4/5 of the substitutions were to residues found in *Der f* I. The full length preproenzyme clone shown in figure 1 is (c) in this figure.

Discussion

The sequences obtained for the most part confirm that reported previously and show the presence of a small but significant degree of polymorphism. Each of the five sequences was unique although they only differed from the others by 1-3 residues. Of the 5 substitutions reported 2 His:Tyr 51 and Glu:Lys 81 were nonconservative with respect to the predicted influence they would have on the tertiary structure [11] but studies with mutagenized protein and peptides have shown that any change can profoundly affect T cell epitopes including alterations in residues outside the T cell epitopes [12-14]. Evavold and Allen [13] have for example reported the loss of epitypic activity by the substitution of Thr and Ser or Gln and Asn for an antihemoglobin T cell clone. Of the 7 epitopes reported to date for human T-cell responses to *Der p* I (45-73, 89-117, 85-109, 111-139 [15] and 45-67, 94-104, 117-143 [16]) three include sequences with a polymorphic residue reported here.

Fig. 2. Alignment of translated sequences of *Der p* I cDNA clones. The sequences are shown in comparison with clone (a) (13T [reported in 1, 2] and *Der f* I [2]). ● = Identity; * = deletion. Clone (b) was truncated.

	10	20	30	40	50	60
<i>Der p</i> I(a)	TNAC	SING	*NAPAE	IDLQ	MRTV	TPIRM
<i>Der p</i> I(b)	TNAC	SING	*NAPAE	IDLQ	MRTV	TPIRM
<i>Der p</i> I(c)		*				
<i>Der p</i> I(d)		*				
<i>Der p</i> I(e)		*				
<i>Der f</i> I	.S..	R..	SV.V	S.L..	SL.....	
	70	80	90	100	110	120
<i>Der p</i> I(a)	LVDC	ASQH	GC	HD	TI	PRG
<i>Der p</i> I(b)	LVDC	ASQH	GC	HD	TI	PRG
<i>Der p</i> I(c)			E.			
<i>Der p</i> I(d)			K.			
<i>Der p</i> I(e)			E.			
<i>Der f</i> I			E...	Q...	E.RS.	P....
	130	140	150	160	170	180
<i>Der p</i> I(a)	ANKI	REAL	AQ	TH	SA	IAV
<i>Der p</i> I(b)	ANKI	REAL	AQ	TH	SA	IAV
<i>Der p</i> I(c)	V.....	S.....				
<i>Der p</i> I(d)	V.....	S.....				
<i>Der p</i> I(e)	A.....	T.....				
<i>Der f</i> I	VKQ...	T...	T.....	R..Q.....	H.....	GST..D..
	190	200	210	220		
<i>Der p</i> I(a)	WIVR	NSWD	TNWG	DNGY	GYFA	ANID
<i>Der p</i> I(b)	WIVR	NSWD	TNWG	DNGY	GYFA	ANID
<i>Der p</i> I(c)				E.....		
<i>Der p</i> I(d)				E.....		
<i>Der p</i> I(e)				E.....		
<i>Der f</i> I		T...	S.....	Q.GNN.....	Q.....	M

The frequency of the polymorphism is greater than the 1/10,000 error expected from replication error during the construction of the cDNA and their authenticity is strongly supported by the fact that $\frac{1}{5}$ of the substitutions were to residues found in the allergen *Der f* I from the other pyroglyphid mite *D. farinae*. This may confound the measurement of T cell cross-reactivity between these allergens or indeed genetic studies on immune responsiveness. Because all of the sequences obtained were different it is unlikely that the full range of polymorphisms have been found. Some constraints probably exist, however, since the changes found were biased to residues which varied between *Der p* I and *Der f* I. There are admittedly 43 of these but another indication of constraint is that even though all the clones were unique, residues 51 and 81 had two representatives of each amino acid. Interestingly, the polymorphisms were in regions of the molecule where most residues are conserved between *Der f* I and *Der p* I and not in the N and C terminal region or the central outside loop which have a high degree of interspecies variation [2, 17]. The pIs of the polypeptides were calculated to be between 5.58 and 6.06, which is still a considerably

smaller range than the 4.4–7.1 found in *Der p* I isolated from mite extracts [18].

The CSL colony of mites which was used here has been maintained for 20 years from mites isolated in Melbourne so it is important to investigate the degree of polymorphism in the wider environment. An indication of the relatedness of the clones was that their 3' untranslated regions, which usually show the highest amount of polymorphism in cDNA [19, 20], were identical. The degree of sequence variation described here is not usually found in DNA coding other eukaryotic proteins regions [21] and Kreitman [22] has for example reported that 11 different isolates of *Drosophila melanogaster* had only one amino acid change in their alcohol dehydrogenase genes. On this basis the frequency of polymorphism described here is unusual.

Clone (c) described here had a 1172 bp insert which encoded all the preproenzyme of *Der p* I which was similar to that reported and analyzed for *Der f* I [2]. The predicted 18-residue leader sequence and the 80-residue proregion had 83% homology with the *Der f* I [2] preproregion which is similar to the rest of the molecule. All of the residues

identified for *Der f* I with a high degree of conservation in other cysteine protease proregions were also conserved in *Der p* I. We have already shown that the expression of recombinant *Der f* I from a proenzyme cDNA has advantages over expression of the mature form [Dilworth et al., unpubl. data] and it is anticipated that similar results can be obtained for *Der p* I.

Acknowledgement

This work was supported by the National Health and Medical Research Council of Australia, the Asthma Foundation of Western Australia and ImmuLogic Pharmaceutical Corporation, Massachusetts.

References

- Chua KY, Stewart GA, Thomas WR, Simpson RJ, Dilworth RJ, Plozza TM, Turner KJ: Sequence analysis of cDNA coding for a major house dust mite allergen, *Der p* I. Homology with cysteine proteases. *J Exp Med* 1988;167:175-182.
- 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.
- 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;91:118-123.
- Trudinger M, Chua KY, Thomas WR: cDNA encoding the major mite allergen *Der f* II. *Clin Exp Allergy* 1991;21:33-37.
- Yuuki T, Okumura Y, Ando T, Yamakawa H, Suko M, Haida M, Okudaira H: Cloning and sequencing of cDNAs corresponding to mite major allergen *Der f* II. *Arerugi* 1990;39:557-561.
- Thomas WR, Chua KY, Smith WA: Molecular polymorphisms of house dust mite allergens. *Exp Appl Acarol* 1992;16:153-166.
- 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.
- Huynh TV, Young RA, Davis RW: Constructing and screening cDNA libraries in gt10 and gt11; in Glover DM (ed): *DNA Cloning*. Oxford, IRL Press, 1985, vol 2: A Practical Approach, pp 47-78.
- Bowtell DD, Saint RB, Rickard MD, Mitchell GF: Expression of *Taenia taeniaeformis* antigens in *Escherichia coli*. *Mol Biochem Parasitol* 1984;13:173-185.
- Sambrook J, Fritsch EF, Maniatis T: *Book Molecular Cloning. A Laboratory Manual*, ed 2. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 1989.
- Taylor WR: Identification of protein sequence homology by consensus template alignment. *J Mol Biol* 1986;188:233-258.
- Liu ZR, Williams KP, Chang YH, Smith JA: Single amino acid substitution alters T cell determinant selection during antigen processing of *Staphylococcus aureus* nuclease. *J Immunol* 1991;146:438-443.
- Evavold BD, Allen PM: Separation of IL-4 production from Th cell proliferation by an altered T cell receptor ligand. *Science* 1991;252:1308-1310.
- Brown LE, Jackson DC, Tribbick G, White DO, Geysen HM: Extension of a minimal T cell determinant allows relaxation of the requirement for particular residues within the determinant. *Int Immunol* 1991;3:1307-1313.
- O'Hehir RE, Garman RD, Greenstein JL, Lamb JR: The specificity and regulation of T-cell responsiveness to allergens. *Annu Rev Immunol* 1991;9:67-95.
- Yssel H, Johnson KE, Schneider PV, Wideman J, Terr A, Kastelein R, De Vries JE: T cell activation-inducing epitopes of the house dust mite allergen *Der p* I. Proliferation and lymphokine production patterns by *Der p* I-specific CD4⁺ T cell clones. *J Immunol* 1992;148:738-745.
- Thomas WR, Chua KY: Recombinant arthropod allergens; in Kraft D (ed): *Molecular Biology and Immunology of Allergens*. Boca Raton, CRC Press, 1992, in press.
- Chapman MD, Platts-Mills TAE: Purification and characterization of the major allergen from *Dermatophagoides pteronyssinus*-antigen P1. *J Immunol* 1980;125:587-592.
- Levitt RC: Polymorphisms in the transcribed 3' untranslated region of eukaryotic genes. *Genomics* 1991;11:484-489.
- Van Deursen J, Schepens J, Peters W, Meijer D, Grosveld G, Hendriks W, Wieringa B: Genetic variability of the murine creatine kinase B gene locus and related pseudogenes in different inbred strains of mice. *Genomics* 1992;12:340-349.
- Bowcock A, Cavalli-Sforza L: The study of variation in the human genome. *Genomics* 1991;11:491-498.
- Kreitman M: Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 1983;304:412-417.