

# Molecular Analysis of the Group 1 and 2 Allergens from the House Dust Mite, *Euroglyphus maynei*

W. Smith<sup>a</sup> K.L. Mills<sup>a</sup> L.A. Hazell<sup>a</sup> B.J. Hart<sup>b</sup> W.R. Thomas<sup>a</sup>

<sup>a</sup>TVW Telethon Institute for Child Health Research, West Perth, Australia, and

<sup>b</sup>Royal Agricultural College, Cirencester, UK

## Key Words

House dust mite · *Euroglyphus maynei* · Eur m 1 · Eur m 2 · cDNA

## Abstract

**Background:** There is increasing evidence that the house dust mite *Euroglyphus maynei* may be a significant source of allergic sensitization. The structural information for the *E. maynei* allergens is largely restricted to a single partial genomic sequence of Eur m 1. **Methods:** A cDNA library was constructed from a culture of *E. maynei*. Clones encoding the major group 1 and 2 allergens were isolated by DNA hybridization and sequenced. **Results:** The sequence of several full length clones of Eur m 1 and Eur m 2 were obtained. The full pre-proenzyme sequence of the cysteine protease Eur m 1 was determined. The translated amino acid sequence of Eur m 1 and Eur m 2 had 84–86% sequence identity with the corresponding allergens from *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae* mites. This is the same as the degree of sequence identity found between *D. pteronyssinus* and *D. farinae* despite *Euroglyphus* being a member of the Pyroglyphinae subfamily rather than the Dermatophagoidinae subfamily. **Conclusion:** The sequences of the major Eur

m 1 and Eur m 2 allergens are described. Their degree of divergence from the *Dermatophagoides* spp. is similar to that observed between *D. pteronyssinus* and *D. farinae* group 1 and group 2 allergens.

## Introduction

*Euroglyphus maynei* has a wide geographical distribution [1–4]. Although *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae* are usually the most prevalent species of house dust mite, in some houses up to 80% are *E. maynei* [3] occurring as mixed populations with either *D. pteronyssinus* [3], *D. farinae* [4, 5] or *Blomia tropicalis* [6]. Given that the levels of *E. maynei* in a mixed population of mites have often been reported to exceed the 100 mites/g dust, a level proposed to represent a risk factor for the development of allergic disease, it is important that the allergenic components of these mites be more clearly defined. In support of this, Hales et al. [7] have shown that *E. maynei* group 1 allergens are a significant source of primary T-cell sensitization and have little T-cell cross-reactivity with *D. pteronyssinus* or *D. farinae* [7]. Three of 17 subjects responding to peptides from the group 1 allergens had strong and preferential responses to Eur m 1 peptides, with Eur m peptides causing the strongest stimulation in a further 2 subjects. These results highlight the probable clinical importance of the *E. maynei* allergens and suggest that more con-

<sup>1</sup> TVW Telethon Institute for Child Health Research is an institution affiliated with The University of Western Australia.

sideration should be given to *E. maynei* in the development of immunotherapy and diagnostic tests.

A partial genomic sequence for the Eur m 1 allergen and subsequent isolation of natural Eur m 1 was first described by Kent et al. [8]. In order to further characterize the *E. maynei* allergens, this study describes several Eur m 1 and Eur m 2 clones isolated from an *E. maynei* cDNA library. The sequence identity of both groups of allergens from *E. maynei* was the same as that found between the *Dermatophagoides* spp. The study also demonstrates the presence of polymorphic residues as previously reported for the *Dermatophagoides* spp. [9, 10].

## Materials and Methods

### Construction of the *E. maynei* cDNA Lambda ( $\lambda$ ) gt10 Library

*E. maynei* (500 mg) were grown in culture at 25°C in medium containing yeast extract and beard shavings at a ratio of 10:1. Polyadenylated mRNA was isolated from the live mites [11] and used to synthesize cDNA by the RNase H method [12] using a kit (Amersham International, Aylesbury, UK). Following the addition *Eco*R1 linkers (New England Biolabs, Beverly, Mass., USA) the cDNA was ligated into *Eco*R1 -digested  $\lambda$  gt10 arms. The ligated DNA was packaged and grown in *Escherichia coli* C600Hfl to select for recombinant phage producing a library with  $2.1 \times 10^5$  recombinant phage. Phage DNA was grown in *E. coli* JP777 to produce amplified stock for library screening.

### PCR Generation of Eur m 1 and Der p 2 Probes

The oligonucleotides (Bresatec Ltd., Australia) used for PCR amplification of Eur m 1 corresponded to nucleotide residues 321–339 (5'-GCGGATCCCTCAATTTCGATTGAATGC-3') and 1012–991 (5'-GGAATTCTTAGAGAATGACAACATATGG-3') of the Der p 1 cDNA sequence [13]. *Tth* polymerase was used according to the manufacturer's specifications (Fisher Biotech International, Australia) with the addition of 20 pmol of each of the primers, 250  $\mu$ M dNTP and an aliquot of the Eur m library phage stock equivalent to  $2 \times 10^8$  recombinant phage. The thermal cycling was performed in a Perkin-Elmer 2400 machine. The 40-cycle programme was as follows: (5 min, 94°C; 30 s, 94°C; 1 min, 55°C; 2 min, 72°C)  $\times$  40 cycles, 8 min, 72°C. For Der p 2 oligonucleotides which corresponded to nucleotide residues 105–114 (5'-CGGGATCCTTGTTGGTCCGACGCCGTTG-3') and 443–428 (5'-GCGAATTCAACAGTGACGACAAC-3') were used to amplify a fragment corresponding to amino acid residues 1–108 [14]. The PCR was performed as outlined above with the exception that 40 cycles were performed with an annealing temperature of 60°C.

### Cloning and DNA Sequencing

PCR amplified DNA were labelled with a Giga prime random prime kit (Bresatec Ltd.) using  $\alpha$ -<sup>32</sup>P dATP and the reaction was stopped by heating at 65°C for 10 min. The labelled probes were then purified using a Sephadex G50 column. Screening of the library and isolation of both Eur m 1 and Eur m 2 clones was performed as previously outlined [15, 16]. Aliquots of the purified  $\lambda$ gt10 clones were amplified by PCR with *Pfu* polymerase according to the manufactur-

er's instructions (Stratagene, La Jolla, Calif., USA) with the addition of 20 pmol of both  $\lambda$ gt10 forward primer (5'-AGCAAGTTCAGCCTGGTTAAG-3') and  $\lambda$ gt10 reverse primer (5'-CTTATGAGTATTTCTTCCAGGGTA-3'). The Eur m 2 PCR products were digested with *Eco*R1 and subcloned into *Eco*R1-digested pUC18 plasmid vector. The Eur m 1 PCR clones were blunt-end cloned into *Sma*I-digested pUC18. The recombinant plasmids were transformed into *E. coli* XL-1 blue cells for sequencing. The pUC18 clones were purified by small scale alkaline lysis. Sequencing was performed in both directions and where necessary clones from 2 separate PCRs were sequenced. Automated cycle sequencing was performed on an ABI 373 sequencer at CMII (Perth, Australia) using both dye-primer and dye-terminator reactions. Nucleotide and protein sequence analysis were performed using Mac Vector Software version 6.0.1 (Oxford Molecular Group, Oxford, UK).

## Results

### Characterization of Eur m 1 and Its Variants

The Eur m 1 PCR probe hybridized to 4 clones following screening of approximately 150,000 recombinant phage. Three clones encoded full length Eur m 1 sequence including 5' and 3' untranslated regions (UTR). These have been called Eur m 1.0101, 0102 and 0103. The fourth clone was a partial 600-bp sequence identical to Eur m 1.0101. Eur m 1.0101 was the largest with 1,139 bp, then Eur m 1.0103 with 1,106 bp and finally Eur m 1.0102 with 1,092 bp. The nucleotide sequences of 0101 and 0102 differed only at 5 nucleotide residues (147, 571, 577, 652 and 1052; fig. 1), whereas Eur m 1.0103 differed from 0101 and 0102 by 17 and 13 nucleotide residues, respectively. The 5' and 3' UTR showed an extremely high degree of similarity with only 2–3 residues difference between all three clones. The 3' UTR include a polyadenylation signal (underlined in fig. 1) and a poly A tail. The nucleotide sequence of the published Eur m 1 genomic sequence had only a single-base transversion at residue 635, with a G in the Eur m 1.0101 sequence substituted for a T in the genomic sequence (fig. 1) resulting in a silent mutation.

The translated sequence of Eur m 1.0101 encoded an open reading frame (ORF) of 954 bp with a classical hydrophobic 18-amino-acid signal peptide and an 80-amino-acid proenzyme sequence upstream of the mature N-termini. The theonine at the +1 position is the previously identified N-terminus of the 223-amino-acid mature protein. The 0101 and 0102 translated sequences differed at a single amino acid as a result of the substitution at nucleotide residue 148. This resulted in a conservative change from a threonine (0101) to a serine (0102) at position –63 (fig. 2). Eur m 1.0103 had 2 conservative substitutions at positions –63 (identical to 0102), 222 (Met to Ile) and a single non-

**Fig. 1.** Eur m 1 nucleotide sequence alignment. The nucleotide sequence for Eur m 1.0101 is aligned with the sequences for the polymorphic variants 0102 and 0103. The sequences are also aligned to the previously published Eur m 1 genomic sequence. Identical nucleotide residues are indicated by (.) and substitutions by the relevant nucleotides.

|         |                                                                                  |      |      |      |      |      |      |      |
|---------|----------------------------------------------------------------------------------|------|------|------|------|------|------|------|
|         | 10                                                                               | 20   | 30   | 40   | 50   | 60   | 70   | 80   |
| 0101    | AATTCGCTTTTAAAAATCTAAACATCTATCCACAATCATGAAATCATTTTGCCATCGCCTCATGTTGGTATTTGA      |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....G.....                                                                      |      |      |      |      |      |      |      |
|         | 90                                                                               | 100  | 110  | 120  | 130  | 140  | 150  | 160  |
| 0101    | GCGCTGTTTATGCTCGTCCAGCTTCGATCAAACTTTTGAAGAATTCAAAAAGCTTCAACAAAACCTATGCTACCCCT    |      |      |      |      |      |      |      |
| 0102    | .....G.....                                                                      |      |      |      |      |      |      |      |
| 0103    | .....G.....                                                                      |      |      |      |      |      |      |      |
|         | 170                                                                              | 180  | 190  | 200  | 210  | 220  | 230  | 240  |
| 0101    | GAAGAAGAAAGTTGCCGTAAAACTTTTGAATCATTGAAATATGTGGAATCAACAAAGGTGCCATCAACCATTTT       |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....C.....                                                                      |      |      |      |      |      |      |      |
|         | 250                                                                              | 260  | 270  | 280  | 290  | 300  | 310  | 320  |
| 0101    | GTCCGATTGTCTATTGGATGAATCAAAACCAATTTTGTATGAATGCTAATGCTTTTGAACAACCAAACTCAATTCG     |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |
|         | 330                                                                              | 340  | 350  | 360  | 370  | 380  | 390  | 400  |
| 0101    | ATTTGAATGCTGAACCTTATCCCTGCAGTATTAACCTAGTGAGTTTACCATCGGAACCTGATTTAAGTTCACCTGCGA   |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....C.....G.....                                                                |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 410                                                                              | 420  | 430  | 440  | 450  | 460  | 470  | 480  |
| 0101    | GTAATCCATTCGATATGCAAGGAGCTGTGTTTCATGCTGGCTTTCTCTGGTGTGCCCAACTGAGTCAGCTTATTT      |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....AT.....C.....A.....                                                         |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 490                                                                              | 500  | 510  | 520  | 530  | 540  | 550  | 560  |
| 0101    | GGCTACCGCAACATGCTCTTTGATCTTGTCTGAACAAGAAATTAGTCGATTGTCTTCACAAACGGTTGCCATGGTGATA  |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....T.....T.....C.....                                                          |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 570                                                                              | 580  | 590  | 600  | 610  | 620  | 630  | 640  |
| 0101    | CAATTCACCGGGAATTGATACATCCAAACAAATGGTGTCTCCAGAACACTACTATCCATACGTTGCACGAGAACAA     |      |      |      |      |      |      |      |
| 0102    | .....A.....C.....                                                                |      |      |      |      |      |      |      |
| 0103    | .....A.A.....C.....                                                              |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 650                                                                              | 660  | 670  | 680  | 690  | 700  | 710  | 720  |
| 0101    | TCATGCCATCGACCAATGCAACGTTACGGTCTCAAGAACTATTGCCAAATTTGGCACCAGACTCGAACAAATTCG      |      |      |      |      |      |      |      |
| 0102    | .....G.....                                                                      |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 730                                                                              | 740  | 750  | 760  | 770  | 780  | 790  | 800  |
| 0101    | TCAAGCTTTGACTCAAAACACATACAGCCGTTGCCGTCATTATTGGCATCAAGATTTGAACGCTTTTCGACATTATGATG |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 810                                                                              | 820  | 830  | 840  | 850  | 860  | 870  | 880  |
| 0101    | GACGAACATCATGCAACACGACAAATGGTTATCAACCAACTATCATGCGCTCAACATTGTTGGTTACGGTAATACACAA  |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 890                                                                              | 900  | 910  | 920  | 930  | 940  | 950  | 960  |
| 0101    | GCTGTTGATTATTGGATCGTAAGAAACGTTGGGATACCACTTGGGGTGATAATGCTATGGTTATTTCGCTGCCAAT     |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 970                                                                              | 980  | 990  | 1000 | 1010 | 1020 | 1030 | 1040 |
| 0101    | CAATTTGATGATGATCGAACAAATCCATATGTATGTCATGCTTTAAACATGACGTTAATTTATTATTCGAATCTAAAT   |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....T..C.....A.T.....                                                           |      |      |      |      |      |      |      |
| genomic | .....                                                                            |      |      |      |      |      |      |      |
|         | 1050                                                                             | 1060 | 1070 | 1080 | 1090 | 1100 | 1110 | 1120 |
| 0101    | TTTTAAATAAATAAATTTTAAAAA.....                                                    |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |
|         | 1130                                                                             |      |      |      |      |      |      |      |
| 0101    | AAAAAAAAAAAAAAG                                                                  |      |      |      |      |      |      |      |
| 0102    | .....                                                                            |      |      |      |      |      |      |      |
| 0103    | .....                                                                            |      |      |      |      |      |      |      |

conservative substitution of a Met to an Asn residue at 28. The 223 residue mature protein has a calculated MW of 25.1 kD and a pI of 5.94. The full-length protein has 85 and 86% identity with Der p 1 and Der f 1, respectively. All the amino acid residues which make up the active site of cys-

teine proteases namely, Glu29, Gly33, Ser34, Cys35, His171, Asp191, Ser192 and Try193, are conserved across all 3 species (fig. 2). A potential N-glycosylation site is also present at positions 53–55 (NMS, underlined in fig. 2).

**Fig. 2.** Comparison of the amino acid sequences of Eur m 1 with other group 1 allergen sequences. The deduced amino acid sequences for Eur m 1.0101, the polymorphic variants 0102 and 0103 and the previously published genomic Eur m 1 sequence [8] are compared with the other group 1 allergens, Der p 1 and Der f 1. The positions of identical residues are marked with a single full stop (.), conservative changes with the single letter amino acid code in lower case and non-conservative changes with the single letter code in uppercase. A dash (–) indicates a space included to optimize the alignment. The following Der p 1 polymorphic residues have been described, 51H, 82E, 96P, 125A, 131D, 137T, 155Y, 162H, 179G and 216Q (Eur m 1 numbering) [9, 20].

|              |  |                       |               |                 |                 |           |           |          |           |
|--------------|--|-----------------------|---------------|-----------------|-----------------|-----------|-----------|----------|-----------|
|              |  | signal peptide        |               |                 |                 |           |           |          |           |
|              |  | -90                   | -80           | -70             | -60             | -50       |           |          |           |
| Eur m 1 0101 |  | MKII                  | LAIASLLVLSAVY | AR              | PASIKTFE        | EFKKAFNK  | TYATPEKEE | VARKN    |           |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Der p 1      |  | ...v.....             | A.....        | s.....          | y.....          | s...F.D.. | A....     |          |           |
| Der f 1      |  | ..fv.....             | -t.....       | -.....          | N.....          | V.E.....  |           |          |           |
|              |  | -40                   | -30           | -20             | -10             |           |           |          |           |
| Eur m 1 0101 |  | FLESLKYVES            | NKGAINHLS     | DLSDLDEF        | KNQFLMNA        | FAFEQLKT  | QFDLNAE   |          |           |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     |          |           |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | .....     | .....     |          |           |
| Der p 1      |  | ...v...Q..            | G.....        | .....           | R...S.E..       | H.....    |           |          |           |
| Der f 1      |  | .....a.....           | .....         | Ry...S.E.....   |                 |           |           |          |           |
|              |  | +1                    | 10            | 20              | 30              | 40        |           |          |           |
| Eur m 1 0101 |  | TYAC                  | SINSVSLP      | SELDLRS         | LRVTPTIR        | MQGGCGS   | WAFSGV    | ASTESAYL |           |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | N.....    | .....     | .....    |           |
| Eur m 1      |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Der p 1      |  | .N.....               | gN-A.a.i..    | Qm.....         | .....           | a.....    |           |          |           |
| Der f 1      |  | .S..R....             | Nv.....       | .....           | .....           | a.....    |           |          |           |
|              |  | 50                    | 60            | 70              | 80              | 90        |           |          |           |
| Eur m 1 0101 |  | AYR                   | NMSLDLAE      | QELVDCAS        | QNGCHGDT        | IPRGIEYI  | QNGV      | VQEHYY   | PPYV      |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1      |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Der p 1      |  | ...Q.....             | .....         | H.....          | .....           | H.....    | S..R..    |          |           |
| Der f 1      |  | ...T....s.....        | .....         | H.....          | .....           | E.rS....  |           |          |           |
|              |  | 100                   | 110           | 120             | 130             | 140       |           |          |           |
| Eur m 1 0101 |  | AREQ                  | SCHRPNAQ      | RYGLKNYC        | QISPPDSN        | KIRQALTQ  | THTAVAVI  | IGIKD    |           |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1      |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Der p 1      |  | .....r.....           | f.iS....      | Y..NV....       | E..A...s.i..... |           |           |          |           |
| Der f 1      |  | ...R.r....s.h..iS.... | .....         | y...VKQ..E..... | i.....          |           |           |          |           |
|              |  | 150                   | 160           | 170             | 180             | 190       |           |          |           |
| Eur m 1 0101 |  | LNA                   | FRHYDGR       | TIMQH           | DNGYQP          | NYHAVN    | IVGYGNT   | QGV      | DIWVRNSWD |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    | TT        |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1      |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Der p 1      |  | .D.....i.r.....       | .....         | s.a.....        | N               |           |           |          |           |
| Der f 1      |  | .R..Q.....i.....      | .....         | S...D.....      |                 |           |           |          |           |
|              |  | 200                   | 210           | 220             |                 |           |           |          |           |
| Eur m 1 0101 |  | WGD                   | NGYGYFA       | ANINL           | MMIEQ           | YPYV      | VV        | ML       |           |
| Eur m 1 0102 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1 0103 |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Eur m 1      |  | .....                 | .....         | .....           | .....           | .....     | .....     | .....    |           |
| Der p 1      |  | .....D.....E.....     | i..           |                 |                 |           |           |          |           |
| Der f 1      |  | ...S....Q.GN.....     | im            |                 |                 |           |           |          |           |

### Characterization of Eur m 2.0101 and Its Variant

Four Eur m 2 clones were isolated following screening of approximately 150,000 recombinant phage. The largest was designated Eur m 2.0101 with the remaining 3 being

designated Eur m 2.0102 (fig. 3). Two of the 3 Eur m 2.0102 clones were identical at the nucleotide level and the third differed from the other 2 at nucleotide 87 but with no amino acid substitution. The 3 Eur m 2.0102 clones also differed

```

      10      20      30      40
CAA TCA TCC AAC TCT AAC AAA AAC CAA AAC ACC GAA AAA ATG ATG TAC
                               M M Y

50      60      70      80      90
AAA ATC TTA TGC CTT TCA TTG TTG GTG GCA GCC GTT GCT GCT GAT CAA
K I L C L S L L V A A V A A D Q

100     110     120     130     140
GTC GAT ATC AAA GAC TGT GCC AAC CAT GAA ATC AAA AAA GTT ATG GTT
V D I K D C A N H E I K K V M V

150     160     170     180     190
CCC GGA TGC AAA GGT TCT GAA CCA TGC GTC ATT CAC CGT GGT ACA GCA
P G C K G S E P C V I H R G T A

200     210     220     230     240
TTC CAA TTG GAA GCC GTA TTC GAT GCC AAC CAA AAC TCA AAC GCC GCT
F Q L E A V F D A N Q N S N A A

250     260     270     280
AAA ATT GAA ATC AAA GCT ACG ATC GAT GGC GTT GAA ATC GAT GTT CCC
K I E I K A T I D G V E I D V P

290     300     310     320     330
GGT ATT GAT AAC AAT CTC TGC CAT TTC ATG AAA TGT CCA TTG GTT AAA
G I D N N L C H F M K C P L V K

340     350     360     370     380
GGA CAA GAA TAC GAT ATC AAA TAT ACA TGG AAT GTC CCA AGA ATT GCA
G Q E Y D I K Y T W N V P R I A

390     400     410     420     430
CCA AAA TCA GAA AAT GTT GTC GTT ACC GTC AAA CTT TTG GGT GAT AAT
P K S E N V V V T V K L G G D N

440     450     460     470     480
GGT GTA TTA GCT TGC GCT ATC GCA ACC CAT GCC AAA ATC CGT GAT TAA
G V L A C A I A T H A K I R D *

490     500     510     520
ATA CAA TGA TTG AAT TGT TGA AAT CGT TTT AAA TTG AAT CAA AAT AAA

530     540
AAT TTC TTT AAT TTG CG

```

**Fig. 3.** Eur m 2 annotated nucleotide sequence. The nucleotide sequence for Eur m 2.0101 is shown with the deduced amino acid sequence corresponding to the open reading frame. The isoform Eur m 2.0102 clone 1 had a substitution of a CG dinucleotide base for the TA dinucleotide at residues 102–103 in Eur m 2.0101. The second clone of Eur m 2.0102 had the same 102–103 CG substitution in addition to a substitution at residue 87 from a T (0101) to a G nucleotide. The location of these substitutions is underlined in the figure.

from the Eur m 2.0101 clone at nucleotide residues 102 and 103, resulting in a single conservative amino acid substitution.

The translated sequences of the variants differed by a single conservative amino acid substitution at residue 5 of the mature protein with an isoleucine (0101)/valine (0102) substitution (fig. 4). A complete ORF of 440 bp (fig. 3) is evident in Eur m 2.0101 with 2 possible translation initiation sites at positions –16 and –17. Residues –17 to –1 contain predominantly hydrophobic residues characteristic of a

signal peptide. The translated 129-amino-acid mature protein had a calculated MW of 14.1 kD and pI of 6.07. All the clones have a polyadenylation signal (AATAAA), with the 3' UTR being identical in sequence but not in length (fig. 3). Eur m 2 has approximately 84% identity with both Der p 2 and Der f 2, and 39% identity (59% overall homology) with Lep d 2.0101 (fig. 4).

**Fig. 4.** Comparison of amino acid sequences of Eur m 2 with other group 2 allergens. The deduced amino acid sequences for Eur m 2.0101 and 0102 are compared with the corresponding sequences for Der p 2 [14], Der f 2 [17] and Lep d 2.0101 [18]. The positions of identical residues are marked with a full stop (.), conservative changes with the single letter amino acid code in lowercase and non-conservative changes with the single letter code in uppercase. Der p 2 polymorphic residues include, -1A, 40L, 47S, 111L, 114N and 127L [10]. The polymorphic residues described for Der f 2 are 63I, 70T, 76V, 88A, 110L, 111V/I, 114N, 122G, 125G [17].

|              | -10                                                | 1               | 10                 | 20  | 30 |
|--------------|----------------------------------------------------|-----------------|--------------------|-----|----|
| Der p 2.0101 | MMYKILCLSLVAAVAR                                   | DQVDVKDCANHEIKK | VLVPGCHGSEPCIIHRGK |     |    |
| Der f 2.0101 | .iS.....VA.....N.....m.D.....d.....                |                 |                    |     |    |
| Eur m 2.0101 | .....A.....i.....m.....k.....v.....t               |                 |                    |     |    |
| Eur m 2.0102 | .....A.....m.....k.....v.....t                     |                 |                    |     |    |
| Lep d 1.0101 | .M.Fia.FA...VAsAGkmTF...GhG.vTElDiT..S.-dT.v....E  |                 |                    |     |    |
|              | 40                                                 | 50              | 60                 | 70  | 80 |
| Der p 2.0101 | PFQLEAVFEANQNTKTAKIEIKASIDGLEVDVPGIDPNACHYMKCPLVKG |                 |                    |     |    |
| Der f 2.0101 | ..T...l.d.....l.....i.....t....f.....              |                 |                    |     |    |
| Eur m 2.0101 | a.....d....sNa.....t...v.i.....N.L..f.....         |                 |                    |     |    |
| Eur m 2.0102 | a.....d....sNa.....t...v.i.....N.L..f.....         |                 |                    |     |    |
| Lep d 1.0101 | KMT...K.A...D.AKVT.KvL.KvA.TT.Q...leTDG.kfi...vK.. |                 |                    |     |    |
|              | 90                                                 | 100             | 110                | 120 |    |
| Der p 2.0101 | QQYDIKYTWNVPKIAPKSENVVTVKVMGDDGVLACAIATHAKIRD      |                 |                    |     |    |
| Der f 2.0101 | .....li..N.....g....                               |                 |                    |     |    |
| Eur m 2.0101 | .E.....r.....ll..N.....                            |                 |                    |     |    |
| Eur m 2.0102 | .E.....r.....ll..N.....                            |                 |                    |     |    |
| Lep d 1.0101 | eAL.FI.sGti.A.t..VK-AD..AEli..H..m..G-tV.GqvE      |                 |                    |     |    |

### Discussion

The mite species *E. maynei* coinhabits with *Dermatophagoides* spp. in temperate climates such as Perth, Western Australia, where it was found to be present at levels adequate to cause sensitization in 75% of homes studied and was the most abundant mite in 2/20 of homes studied [3]. IgE binding to *E. maynei* and *D. pteronyssinus* has shown extensive cross-reactivity [19], whereas more recently T-cell responses have demonstrated very little cross-reactivity [7]. There is accordingly more information required about the allergens. The cDNA sequence data reported here is the first full-length information reported for both Eur m 1 and Eur m 2.

The group 1 *Dermatophagoides* allergens are cysteine proteases synthesized as pre-proenzymes with an 18-amino-acid signal peptide sequence, an 80-residue proenzyme sequence and a mature protein of 222 amino acids for Der p 1 and 223 for Der f 1. The Eur m 1 cDNA clones exhibited the same pre-proenzyme structure with a 223-residue mature sequence. The sequence reported by this study has extended the known sequence (residues 1–212 of the mature protein translated from a PCR amplification of genomic DNA) to include the complete ORF and both 5' and 3' UTR. Like the Der p 1 sequence, the UTRs are largely conserved. The translated Eur m 1.0101 sequence reported here was identical to the genomic amino acid sequence (in the residues available for comparison) but differed from Eur m 1.0102 and Eur m 1.0103 by 1 and 3 amino acids, respectively. In fact, the Eur m 1.0101 cDNA only differed by

a single silent nucleotide substitution from the 635 residues of the genomic sequence reported by Kent et al. [8]. Although only a small number of clones have been analysed, the sequence divergence appears less than that found for Der p 1 where all the sequences have been different. It is tempting to speculate that since *E. maynei* are difficult to grow in culture that some genetic selection or founder effect may have occurred.

Although the gross morphology of *E. maynei* is very different from the *Dermatophagoides* spp., the amino acid sequence of Eur m 1 was 85–86% identical to Der p 1 and Der f 1 which is the same as the difference between the latter two groups. Of the 223 residues, about 170 were identical between the species (depending on the variants) and 15 were unique to Eur m 1. The remaining differences were equally Der p- or Der f-like. Only 5 residues were different in all 3 species. The regions of interspecies variation occurred in the same region as the variation found between Der p 1 and Der f 1, namely the N-terminal 20 residues and the residues between positions 90 and 140. These regions are almost devoid of solvent-inaccessible core hydrophobic residues and most of the region 90–130 forms a flexible loop connecting the domains [20]. The C-terminal 23 amino acids which had 35% sequence variation between Der f 1 and Der p 1 did not show as high a disparity, i.e. 13–26% between the *E. maynei* and each of the *D. pteronyssinus* and *D. farinae* sequences. The N-glycosylation site at amino acid residues 53–55 (52–54 for Der p 1) was found in all species. They also all contained an aspartate at residue –1 of the proenzyme sequence. This residue is required for the

substrate specificity of the cysteine protease activity [21] and could be conserved so that self-cleavage could be used to release the mature protein.

Several polymorphisms have been described for the group 1 allergens [9, 22]. Der p 1 polymorphisms occur at a limited number of locations within the mature protein sequence. Several of the Eur m 1 residues are identical to the polymorphic residues described for Der p 1 (fig. 2). Three of these (161, 178 and 215) occur only rarely in Der p 1 sequences, but are present in both the Der f 1 and Eur m 1 sequences, supporting the notion that they are real polymorphisms and not artefacts of the cloning process. With a dearth of information on the sequence variation found in environmental mite populations, it is not possible to predict whether the rare substitutions may be more frequent in some populations of *D. pteronyssinus* mites.

This study reports the first sequence encoding Eur m 2. As for the Eur m 1, there was a high degree of identity with Der p 2 and Der f 2 of approximately 84%. This is a little less than the 88% identity described for Der p 2 and Der f 2 but the substitutions are nearly all conservative as are the substitutions between Der p 2 and Der f 2. Of the 29 differences found between the 129-residue Eur m 2 and either Der p 2 or Der f 2, 15 substitutions were unique to Eur m 2 with about an equal split between Der p-like and Der f-like for the remaining 14 substitutions. Importantly, this confirms the analysis with the group 1 allergens which shows that the *E. maynei* allergens have the same degree of divergence from *D. pteronyssinus* and *D. farinae* as they have from each other and there is no tendency to be either more Der p- or Der f-like. It is of interest to note that there is a similar moderate degree of identity (approximately 40%) between the group 2 allergens of the house dust mite and other group 2 allergens such as Lep d 2 from the mite, *Lepidoglyphus destructor*.

Polymorphic amino acid residues have been found in Der p 2 at residues -1, 40, 47, 111, 114 and 127. The residues found in Eur m 2 at positions -1, 47, 111 and 114 were the same as the reported Der p 2 polymorphisms. Residue 111 was a Met in 6/7 Der p 2 clones, with a genomic clone containing a Leu at this position. The Leu 111 was present in both Eur m 2 clones, suggesting that further studies may reveal a wider distribution of the Der p 2 variant. The four Eur m 2 clones described in this paper show only a single amino acid polymorphism, a conservative substitution at residue 5. In contrast, previous reports with Der p 2 sequences from a commercial stock of mites (CSL Ltd., Melbourne) and environmental isolates have shown an unusually high frequency of polymorphism as found for the group 1 allergens [10].

NMR studies have demonstrated that the tertiary structure of the group 2 allergens consists of a single-domain immunoglobulin fold with structural similarities to the third and fourth domains of the transglutaminase coagulation factor [23]. The non-conservative interspecies substitutions of the group 2 allergens, which are found at residues 11, 19, 36, 48, 70, 73, 84 and 114, all correspond to residues in the loops connecting the  $\beta$ -sheets. The loop from 65 to 85 includes 3 of these substitutions, each unique to Eur m 2. Along with the other group 2 allergens, Eur m 2 has approximately 45% homology to a family of proteins first represented in the GenBank database by HE1, a protein which is produced by the epithelial cells and secreted in very large quantities into the epididymal fluid [24]. Other members of this family now include proteins isolated from the dog [25], cow [26] and the tobacco hornworm moth [27].

Several studies have shown that in many temperate regions the number of *E. maynei* found in houses can reach levels which, as indicated by studies with other mite species, are associated with allergic sensitization. This species has a higher abundance in beds relative to carpets than *Dermatophagoides*, which may increase human exposure [1]. It has been directly shown by van Hage-Hamsten et al. [28] that *E. maynei* produces IgE-binding specificities distinct from *Dermatophagoides* and can be a common cause of sensitization. Our study is the first to report the complete sequence for the major allergens, Eur m 1 and Eur m 2. It also shows that the degree of divergence from the *Dermatophagoides* ssp. is similar to that observed between *D. pteronyssinus* and *D. farinae* and highlights the need to develop reagents for species-specific diagnosis and immunotherapy. It is clearly acknowledged that recombinant house dust mite allergens are invaluable tools for the development of improved diagnosis and therapy, and difficulties in culturing *E. maynei* on a commercial scale would make the future development of recombinant *E. maynei* allergens a practical solution.

## Acknowledgements

The work was conducted with grants from the National Health and Medical Research Council of Australia and the Asthma Foundation of Western Australia. All of the above sequences have been submitted to GenBank and have been allocated the following accession numbers: AF047610 (Eur m 1.0101), AF047611 (Eur m 1.0102), AF047612 (Eur m 1.0103), AF047613 (Eur m 2.0101) and AF047614 (Eur m 2.0102).

## References

- Hart BJ, Whitehead L: Ecology of house dust mites in Oxfordshire. *Clin Exp Allergy* 1990; 20:203–209.
- Colloff MJ: A review of the biology and the allergenicity of the house dust mite *Euroglyphus maynei* (Acari: Pyroglyphidae). *Exp Appl Acarol* 1991;11:177–198.
- Colloff MJ, Stewart GA, Thompson PJ: House dust mite acarofauna and *Der p* I equivalent in Australia: the relative importance of *Dermatophagoides pteronyssinus* and *Euroglyphus maynei*. *Clin Exp Allergy* 1991;21:225–230.
- Wickman M, Nordvall SL, Pershagen G, Korsgaard J, Johansen N: Sensitization to domestic mites in a cold temperature region. *Am Rev Respir Dis* 1993;148:58–62.
- Arlian LG, Bernstein D, Bernstein IL, Friedman S, Grant A, Lieberman P, Lopez M, Metzger J, Platts-Mills TAE, Schatz M, Spector S, Wasserman S: Prevalence of dust mites in the homes of people with asthma living in eight different geographic areas of the United States. *J Allergy Clin Immunol* 1992;90:292–300.
- Morgan MS, Arlian LG, Fernandez-Caldas E: Cross-allergenicity of the house dust mites *Euroglyphus maynei* and *Blomia tropicalis*. *Ann Allergy Asthma Immunol* 1996;77:868–875.
- Hales BJ, Thomas WR: T-cell sensitization to epitopes from the house dust mites *Dermatophagoides pteronyssinus* and *Euroglyphus maynei*. *Clin Exp Allergy* 1997;77:386–392.
- Kent NA, Hill MR, Keen JN, Holland PWH, Hart BJ: Molecular characterization of the group I allergen *Eur m* I from house dust mite *Euroglyphus maynei*. *Int Arch Allergy Immunol* 1992;99:150–152.
- Chua KY, Kehal PK, Thomas WR: Sequence polymorphisms of cDNA clones encoding the mite allergen *Der p* I. *Int Arch Allergy Immunol* 1993;101:363–368.
- Chua KY, Huang CH, Shen HD, Thomas WR: Analysis of sequence polymorphism of a major mite allergen, *Der p* 2. *Clin Exp Allergy* 1996;26:829–837.
- 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.
- Gubler U, Hoffman BJ: A simple and very efficient method for generating cDNA libraries. *Gene* 1983;25:263–269.
- Thomas WR, Stewart GA, Simpson RJ, Chua KY, Plozza TM, Dilworth RJ, Nisbet A, Turner KJ: Cloning and expression of DNA coding for the major house dust mite allergen *Der p* I in *Escherichia coli*. *Int Arch Allergy Immunol* 1988;85:127–129.
- 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.
- Smith WA, Chua 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.
- Smith W, Thomas WR: Comparative analysis of the genes encoding the group 3 allergens from *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*. *Int Arch Allergy Immunol* 1996;109:133–140.
- 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.
- Varela J, Ventas P, Carreira J, Barbas JA, Gimenez-Gallego G, Polo F: Primary structure of *Lep d* 1, the main *Lepidoglyphus destructor* allergen. *Eur J Biochem* 1994;225:93–98.
- Hill MR, Newton MR, Hart BJ: Comparative IgE responses to extracts of five species of house dust mite, using Western blotting. *Clin Exp Allergy* 1993;23:110–116.
- Robinson C, Kalsheker NA, Srinivasan N, King CM, Garrod DR, Thompson PJ, Stewart GA: On the potential significance of the enzymatic activity of mite allergens to immunogenicity. Clues to structure and function revealed by molecular characterization. *Clin Exp Allergy* 1997;27:10–21.
- Hewitt CRA, Horton H, Jones RM, Pritchard DI: Heterogeneous proteolytic specificity and activity of the house dust mite proteinase allergen *Der p* 1. *Clin Exp Allergy* 1997;27:201–207.
- Thomas WR, Smith W, Hales BJ, Carter MD: Functional effects of polymorphisms of house dust mite allergens. *Int Arch Allergy Immunol* 1997;113:96–98.
- Ichikawa S, Hatanaka H, Yuuki T, Iwamoto N, Kojima S, Nishiyama C, Ogura K, Okumura Y, Inagaki F: Solution structure of *Der f* 2, the major mite allergen for atopic diseases. *J Biol Chem* 1998;273:356–360.
- Kirchhoff C, Osterhoff C, Young L: Molecular cloning and characterization of HE1, a major secretory protein of the human epididymis. *Biol Reprod* 1996;54:847–856.
- Ellerbrock K, Pera I, Hartung S, Ivell R: Gene expression in the dog epididymis: A model for human epididymal function. *Int J Androl* 1994; 17:314–323.
- Larsen LB, Ravn P, Boisen A, Berglund L, Petersen TE: Primary structure of EPV20, a secretory glycoprotein containing a previously uncharacterized type of domain. *Eur J Biochem* 1997;243:437–441.
- Meszaros M, Morton DB: Identification of a developmentally regulated gene, *esr16*, in the tracheal epithelium of *Manduca sexta*, with homology to a protein from the human epididymis. *Insec Biochem Mol Biol* 1996;26:7–11.
- van Hage-Hamsten M, Johansson SG: Clinical significance and allergenic cross-reactivity of *Euroglyphus maynei* and other non-pyroglyphid and pyroglyphid mites. *J Allergy Clin Immunol* 1989;83:581–589.

Copyright: S. Karger AG, Basel 1999. Reproduced with the permission of S. Karger AG, Basel. Further reproduction or distribution (electronic or otherwise) is prohibited without permission from the copyright holder.