Brazilian Journal of Medical and Biological Research (2008) 41: 380-388
ISSN 0100-879X

Cloning, sequence analysis, and expression of cDNA coding for the major house dust mite allergen, Der f 1, in *Escherichia coli*

Y. Cui[1,2,3], P. Zhou[2], J. Peng[3], M. Peng[2], Y. Zhou[1], Y. Lin[3] and L. Liu[1]

[1]Yancheng Health Vocational and Technical College, Yancheng, Jiangsu Province, the People's Republic of China
[2]State Key Laboratory of Tropical Crop Biotechnology, Institute of Tropical Science, China Academy of Tropical Agricultural Science, Haikou, the People's Republic of China
[3]Hainan Medical College, Haikou, the People's Republic of China

Correspondence to: P. Zhou, State Key Laboratory of Tropical Crop Biotechnology, Institute of Tropical Science, China Academy of Tropical Agricultural Science, Chengxi Road, Haikou 571101, the People's Republic of China
E-mail: 6801609@163.com

Our objective was to clone, express and characterize adult *Dermatophagoides farinae* group 1 (Der f 1) allergens to further produce recombinant allergens for future clinical applications in order to eliminate side reactions from crude extracts of mites. Based on GenBank data, we designed primers and amplified the cDNA fragment coding for Der f 1 by nested-PCR. After purification and recovery, the cDNA fragment was cloned into the pMD19-T vector. The fragment was then sequenced, subcloned into the plasmid pET28a(+), expressed in *Escherichia coli* BL21 and identified by Western blotting. The cDNA coding for Der f 1 was cloned, sequenced and expressed successfully. Sequence analysis showed the presence of an open reading frame containing 966 bp that encodes a protein of 321 amino acids. Interestingly, homology analysis showed that the Der p 1 shared more than 87% identity in amino acid sequence with Eur m 1 but only 80% with Der f 1. Furthermore, phylogenetic analyses suggested that *D. pteronyssinus* was evolutionarily closer to *Euroglyphus maynei* than to *D. farinae*, even though *D. pteronyssinus* and *D. farinae* belong to the same *Dermatophagoides* genus. A total of three cysteine peptidase active sites were found in the predicted amino acid sequence, including 127-138 (QGGCGSCWAFSG), 267-277 (NYHAVNIVGYG) and 284-303 (YWIVRNSWDTTWGDSGYGYF). Moreover, secondary structure analysis revealed that Der f 1 contained an α helix (33.96%), an extended strand (17.13%), a ß turn (5.61%), and a random coil (43.30%). A simple three-dimensional model of this protein was constructed using a Swiss-model server. The cDNA coding for Der f 1 was cloned, sequenced and expressed successfully. Alignment and phylogenetic analysis suggests that *D. pteronyssinus* is evolutionarily more similar to *E. maynei* than to *D. farinae*.

Key words: *Dermatophagoides farinae*; Der f 1; Recombinant allergens; Cloning; Expression; Bioinformatics

Research supported by the National Sciences Foundation of China (#30060166), Natural Science Fund of Hainan Province (#2005-80556), and Hainan Provincial Health Department (#2005-6).

Received September 28, 2007. Accepted April 23, 2008

Introduction

It is well known that house dust mites are associated with IgE-mediated allergies such as asthma, rhinitis, conjunctivitis, urticaria, and atopic dermatitis. Although allergens from several mite families have been reported, including Acaridae, Glycophagidae and Pyroglyphidae, proteins produced by mites of the Pyroglyphidae family, especially the species *Dermatophagoides farinae*, *D. pteronyssinus*, and *Euroglyphus maynei* are the most important

mites because they affect human beings. Indeed, these species are commonly found in human dwellings worldwide (1). As complex organisms, house dust mites produce thousands of different proteins and macromolecules, and over 30 IgE-binding components have been demonstrated in patients allergic to house dust mites by immunological analysis (2). Of these allergens, group 1 is the major one and has been identified to be homologous with cysteine proteases (3).

Although the recombinant allergen Der p 1 has been studied in clinical trials in some developed countries, currently, only the crude extracts from *D. farinae* are used in diagnosis and immunology in China. To characterize house dust mite allergens, we cloned, sequenced and analyzed the cDNA encoding the major allergen *D. farinae* group 1 (Der f 1). Interestingly, results of homology searches and phylogenetic analyses suggested that *D. farinae* seemed to be more similar to *E. maynei* than to *D. pteronyssinus*, even though *D. farinae* and *D. pteronyssinus* belong to the same *Dermatophagoides* genus. Here, we aimed to clone and sequence Der f 1 and describe its relevant molecular characteristics.

Material and Methods

D. farinae culture and isolation of adult mites

To isolate house dust mites, we first obtained house dusts from the floors of rice and flour shops in Haikou City, Hainan Island, and isolated house dust mites under a stereomicroscope. All of the mites potentially belonging to *D. farinae* were placed in a culture chamber. After about 2 months, the whole culture in the chamber was examined under a stereomicroscope to select for *D. farinae*. Subsequently, mites considered to be *D. farinae* were cultivated in small culture chambers for pure culture. Two months later, the mites were taken from the chambers to identify mite species. If the mite was confirmed to be *D. farinae*, all of the mites in the chamber were considered to be *D. farinae*. These mites were then raised on a large-scale basis at 25°C and 75% relative humidity. A culture medium consisting of yeast, wheat flour and rice was used for the large-scale cultivation. According to our previous protocols (4), the pure adults were isolated as follows: the whole culture was placed on glass plates, and after 30 min the culture media were removed manually. The adult mites, larval mites and some media were collected into a small ceramic cup using a small writing brush. Under a lamp, the adults moved rapidly to the basal part of the ceramic cup. The adults were removed with a writing brush under a microscope, and about 600 mites were chosen for total RNA isolation.

Isolation of total RNA

Total RNA was isolated from mites using an RNA isolator (Code #D312; TaKaRa Biotechnology Limited Company, Dalian, China) according to manufacturer instructions. Briefly, about 600 live adult mites were rapidly frozen in liquid nitrogen and 1 mL RNA isolator was added. The mites in RNA isolator were homogenized using a PowerGen 125 Tissue Homogenizer (Fisher Scientific, UK) starting at 5000 rpm and gradually increasing to approximately 20,000 rpm for 30-60 s at room temperature. After homogenization, the entire content of the tissue homogenizer was transferred to an Eppendorf tube and 0.2 mL chloroform was added. After precipitation and air-drying, the RNA pellet was dissolved in RNase-free water and stored at -80°C.

Polymerase chain reaction

The oligonucleotide primers F (5'GGATCCATGAAATTCGTTTTGGCCATTG3'), R0 (5'TCGCAAGAGTAGTTGTTTTTAT3') and R (5'CTCGAGTCACATGATTACAACATATGGATATT3') were designed on the basis of the sequence of the Der f 1 gene (GenBank AB034946) and synthesized. To facilitate cloning, the F and R primers contained a *Bam*HI site at the 5' end of the coding sequence (underlined) of F and an *Xho*I site at the 5' end of R, respectively. First, a reverse transcriptase-polymerase chain reaction (RT-PCR) amplification was performed using the total RNA isolated from the mites with the one-step RNA PCR kit (Code #DRR024A, TaKaRa) according to manufacturer instructions. In this reaction, the final concentrations of the individual components per reaction (50 µL final volume) were as follows: 20 pmol of the primers F and R0 each (1 µL each), total RNA (4 µL), 2X one-step RNA PCR buffer (25 µL), 2.5 mM of the dNTP mixture (2 µL), 40 U/µL of an RNase inhibitor (1 µL), 5 U/µL of the M-MLV reverse transcriptase XL (0.5 µL), and 5 U/µL Ex Taq (1 µL). RT-PCR was performed in the PCR Thermal Cycler Dice (Code #TP600, TaKaRa). The PCR conditions used in this study included an initial incubation for 2 min at 94°C followed by 30 cycles of 15 s at 98°C, 30 s at 57°C, and 1 min at 72°C. After a final incubation for 5 min at 72°C, the amplicons were analyzed by agarose gel electrophoresis (1.0%) and visualized with ImageMaster® VDS (USA). To obtain the gene fragment encoding Der f 1, the above amplicon was treated as a template and the F and R primers were used in the second PCR with the PrimeSTARTM HS DNA polymerase kit (Code #DR010A, TaKaRa). In this reaction, the final concentrations of the components per reaction (50 µL final volume) were as follows: 20 pmol each of the F and R primers (0.5 µL each), the above amplicons (1 µL), 2X PrimerSTARTM buffer (25 µL), 2.5 mM of the dNTP mixture (2 µL), and 2.5 U/µL of the PrimeSTARTM HS DNA polymerase (0.5 µL). PCR

was performed in the same PCR Thermal Cycler Dice as described above with an initial incubation of 5 min at 94°C, followed by 30 cycles of 10 s at 98°C, 10 s at 57°C, and 1 min at 72°C. The PCR product was then analyzed and visualized as described above.

Cloning and DNA sequencing

After being recovered from the gel using the agarose gel DNA purification kit ver. 2.0 (Code #DV805, TaKaRa), a poly "A" tail was added to the PCR-amplified DNA with the DNA A-tailing kit (Code #D404, TaKaRa) and ligated into the pMD19-T simple vector (Code #D104, TaKaRa). The *Escherichia coli* JM109 (Code #D9052, TaKaRa) was then transformed with the recombinant plasmids and positive clones were selected by blue/white screening on Luria-Bertani (LB) plates containing 100 µg/mL ampicillin, and confirmed by restriction enzyme analysis and automatic DNA sequencing.

Construction of expression plasmids

The recombinant plasmids constructed previously were digested with *Bam*HI and *Xho*I to release the Der f 1 fragment. After recovery from the gel using the agarose gel DNA purification kit ver. 2.0 (Code #DV805, TaKaRa), the fragment was subcloned into expression vector pET28a(+) (kit lot #N72770, Novagen, Germany) to create pET28a(+)-Der f 1 using the DNA ligation kit (Code #D6023, TaKaRa). *E. coli* competent JM109 cells (Code #D9052, TaKaRa) were transformed with pET28a(+)-Der f 1 plasmids and positive clones were selected by blue/white screening and determined by restriction enzyme analysis with *Bam*HI and *Xho*I.

Expression of rDer f 1 in *E. coli* BL21

Five microliters of pET28a(+)-Der f 1 plasmid purified with the MiniBEST plasmid purification kit ver. 2.0 was used to transform 100 µL *E. coli* BL21 (Stratagene, USA). The *E. coli* BL21 carrying pET28a(+)-Der f 1 was grown on LB plates containing 50 µg/mL kanamycin at 37°C overnight. The cells were further cultured for 2 h in 2 mL LB solution containing 1 mM isopropyl-ß-D-thiogalactopyranoside at 37°C to induce the *tac* promoter. The *E. coli* cells were harvested by centrifugation and 20 µL of the cell pellet including both lysed supernatant and the whole cells was taken and 20 µL of 2X SDS sample buffer was added to the mixture. The samples were heated for 10 min at 95°C and subjected to SDS-PAGE.

Western blotting

For Western blotting, proteins were transferred to a PVDF membrane (Tiangen, Biotech Co. Led, Beijing, China), the blot was incubated with the primary antibody of the His-Tag Mab at 4°C overnight, followed by incubation with the secondary anti-mouse goat IgG antibody AP conjugated for 1 h at room temperature, and proteins were then visualized using 3 mL BCIP/NBT (Roche, Switzerland).

Nucleotide sequence, inferred amino acid sequence, alignment, and phylogeny

Sequences were edited to remove the vector sequence and the extra restriction sites. One open reading frame (ORF) was obtained using the ORF finder at the National Center for Biotechnology Information (NCBI) website. The amino acid sequence of Der f 1 was deduced using the Translate Tools in the ExPaSy web server, and matched with the group 1 allergen and cysteine protease sequences published in GenBank using BLAST. Next, the published sequences of ticks and mites were chosen for alignment using VECTOR NTI 9.0 software (IBI, New Haven, CT, USA), and a polygenic tree was constructed using the PAUP 4.0b10 software.

Secondary and three-dimensional structure analysis

The secondary structure prediction was performed by SOPMA, and a simple three-dimensional model was constructed with a template (PDB ID:1XKG) using the SWISS-MODEL server.

Results

Cloning and sequence of Der f 1 cDNA

Total RNA was isolated from adult mites and Der f 1 cDNA fragments were amplified using RNA isolated from adult mites by nested-PCR (966 bp, Figure 1). The recovered PCR product was cloned into the plasmid vector pMD19-T, which was used to transform competent *E. coli* cells (JM109). The positive clones were selected by blue/white screening, determined by restriction enzyme digestion and sequenced to confirm their identity by automatic DNA sequencing (Figure 2, the vector sequence and the added restriction sites were removed from the sequencing results). The sequencing results demonstrated that our Der f 1 cDNA is identical to the published sequence (GenBank AB034946), except for a point mutation at position 23 (C>T). Using the ORF finder in the NCBI web server, a complete ORF was found within the Der f 1 cDNA, whose length was 966 bp from the start codon ATG to the stop codon TGA (Figure 2).

Inferred amino acid sequence and its physicochemical property

In order to forecast the physicochemical property of the

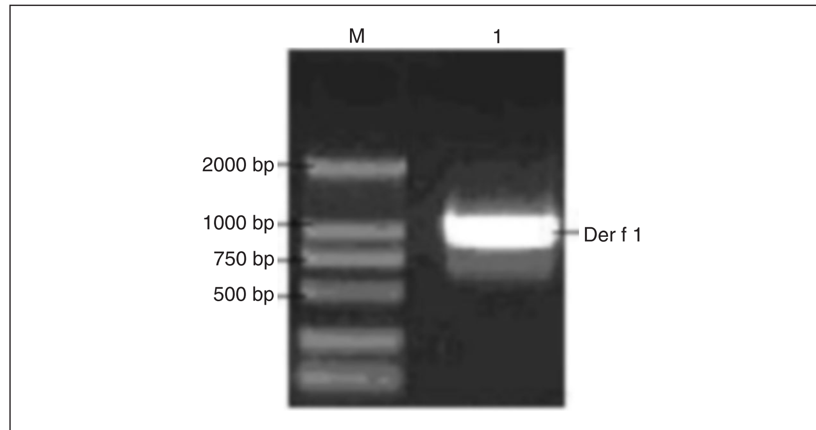

Figure 1. Amplification of Der f 1 by nested-PCR. Total RNA was isolated from mites using an RNA isolator and subjected to RT-PCR as described in Material and Methods. The PCR products were separated on 1% agarose gel containing ethidium bromide. *Lane 1*, DNA marker DL2,000; *lane 2*, PCR product.

Figure 2. The amino acid sequence of Der f 1 deduced from its nucleotide sequence using Translate Tools (http://www.expasy.org). The bold numbers 1-18 represent the signal peptide sequence, the underlined amino acids 19-98 represent the sequence coding for pro-peptides.

```
1    ATG AAA TTC GTT TTG GCC ATT GTC TCT TTG TTG GTA TTG AGC ACT   45
1    Met Lys Phe Val Leu Ala Ile Val Ser Leu Leu Val Leu Ser Thr   15

46   GTT TAT GCT CGT CCA GCT TCA ATC AAA ACT TTT GAA GAA TTC AAA   90
16   Val Tyr Ala Arg Pro Ala Ser Ile Lys Thr Phe Glu Glu Phe Lys   30

91   AAA GCC TTC AAC AAA AAC TAT GCC ACC GTT GAA GAG GAA GAA GTT   135
31   Lys Ala Phe Asn Lys Asn Tyr Ala Thr Val Glu Glu Glu Glu Val   45

136  GCC CGT AAA AAC TTT TTG GAA TCA TTG AAA TAT GTT GAA GCT AAC   180
46   Ala Arg Lys Asn Phe Leu Glu Ser Leu Lys Tyr Val Glu Ala Asn   60

181  AAA GGT GCC ATC AAC CAT TTG TCC GAT TTG TCA TTG GAT GAA TTC   225
61   Lys Gly Ala Ile Asn His Leu Ser Asp Leu Ser Leu Asp Glu Phe   75

226  AAA AAC CGT TAT TTG ATG AGT GCT GAA GCT TTT GAA CAA CTC AAA   270
76   Lys Asn Arg Tyr Leu Met Ser Ala Glu Ala Phe Glu Gln Leu Lys   90

271  ACT CAA TTC GAT TTG AAT GCC GAA ACA AGC GCT TGC CGT ATC AAT   315
91   Thr Gln Phe Asp Leu Asn Ala Glu Thr Ser Ala Cys Arg Ile Asn   105

316  TCG GTT AAC GTT CCA TCG GAA TTG GAT TTA CGA TCA CTG CGA ACT   360
106  Ser Val Asn Val Pro Ser Glu Leu Asp Leu Arg Ser Leu Arg Thr   120

361  GTC ACT CCA ATC CGT ATG CAA GGA GGC TGT GGT TCA TGT TGG GCT   405
121  Val Thr Pro Ile Arg Met Gln Gly Gly Cys Gly Ser Cys Trp Ala   135

406  TTC TCT GGT GTC GCC GCA ACT GAA TCA GCT TAT TTG GCC TAC CGT   450
136  Phe Ser Gly Val Ala Ala Thr Glu Ser Ala Tyr Leu Ala Tyr Arg   150

451  AAC ACG TCT TTG GAT CTT TCT GAA CAG GAA CTC GTC GAT TGC GCA   495
151  Asn Thr Ser Leu Asp Leu Ser Glu Gln Glu Leu Val Asp Cys Ala   165

496  TCT CAA CAC GGA TGT CAC GGC GAT ACA ATA CCA AGA GGC ATC GAA   540
166  Ser Gln His Gly Cys His Gly Asp Thr Ile Pro Arg Gly Ile Glu   180

541  TAC ATC CAA CAA AAT GGT GTC GTT GAA GAA AGA AGC TAT CCA TAC   585
181  Tyr Ile Gln Gln Asn Gly Val Val Glu Glu Arg Ser Tyr Pro Tyr   195

586  GTT GCA CGA GAA CAA CAA TGC CGA CGA CCA AAT TCG CAA CAT TAC   630
196  Val Ala Arg Glu Gln Gln Cys Arg Arg Pro Asn Ser Gln His Tyr   210

631  GGT ATC TCA AAC TAC TGC CAA ATT TAT CCA CCA GAT GTG AAA CAA   675
211  Gly Ile Ser Asn Tyr Cys Gln Ile Tyr Pro Pro Asp Val Lys Gln   225

676  ATC CGT GAA GCT TTG ACT CAA ACA CAC ACA GCT ATT GCC GTC ATT   720
226  Ile Arg Glu Ala Leu Thr Gln Thr His Thr Ala Ile Ala Val Ile   240

721  ATT GGC ATT AAA GAT TTG AGA GCT TTT CAA CAT TAT GAT GGA CGA   765
241  Ile Gly Ile Lys Asp Leu Arg Ala Phe Gln His Tyr Asp Gly Arg   255

766  ACA ATC ATT CAA CAT GAC AAT GGT TAT CAA CCA AAC TAT CAT GCC   810
256  Thr Ile Ile Gln His Asp Asn Gly Tyr Gln Pro Asn Tyr His Ala   270

811  GTC AAC ATT GTC GGT TAC GGA AGT ACA CAA GGC GTC GAT TAT TGG   855
271  Val Asn Ile Val Gly Tyr Gly Ser Thr Gln Gly Val Asp Tyr Trp   285

856  ATC GTA CGA AAC AGT TGG GAT ACT ACC TGG GGT GAT AGC GGA TAC   900
286  Ile Val Arg Asn Ser Trp Asp Thr Thr Trp Gly Asp Ser Gly Tyr   300

901  GGA TAT TTC CAA GCC GGA AAC AAC CTC ATG ATG ATC GAA CAA TAT   945
301  Gly Tyr Phe Gln Ala Gly Asn Asn Leu Met Met Ile Glu Gln Tyr   315

946  CCA TAT GTT GTA ATC ATG TGA   966
316  Pro Tyr Val Val Ile Met End
```

Der f 1 protein, the amino acid sequence of Der f 1 was deduced from its nucleotide sequence using the Translate Tools software in the ExPaSy web server (Figure 2). Based on ProtParam Tools, the putative protein had a molecular mass of 36,419 kDa and a theoretical pI of 5.66, the most likely cleavage site of which is between amino acid positions 18 and 19 (Figure 3). Using ProtParam, the instability index of the complete preproform protein was computed to be 38.17, which classifies the protein as stable. The deduced grand average of hydropathicity was -0.354, demonstrating that this protein was hydrophobic. This result is the same as that predicted by the Hydrophobicityplot 1.0 software.

Subcloning, expression and Western blotting

After the recombinant plasmid pMD19-T-Der f 1 was digested by both *Bam*HI and *Xho*I, the recovered cDNA encoding Der f 1 was subcloned into the expression vector of pET28a(+) and identified by restriction analysis. The plasmid pET28a(+)-Der f 1 was then transformed into *E. coli* BL21 and the protein was expressed in the presence of the IPTG inducer. A single band of about 34 kDa from SDS-PAGE and Western blotting was observed, in accordance with prediction results for the sequence encoding Der f 1 without the sequence for 18 residues of the single peptide (Figure 4A and B).

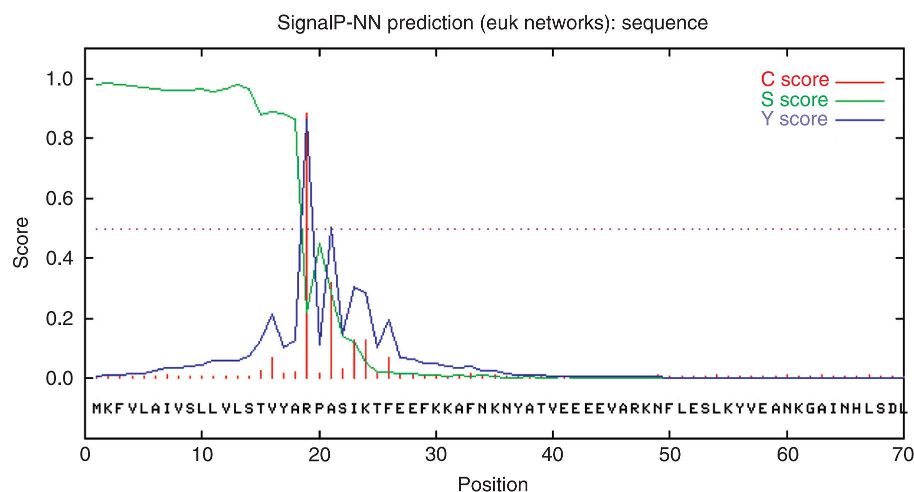

Figure 3. Prediction of signalP-NN using the euk networks of the SignalP 3.0 software (http://www.cbs.dtu.dk/services/SignalP/). The most likely cleavage site between the positions 18 and 19 is VYA-RP.

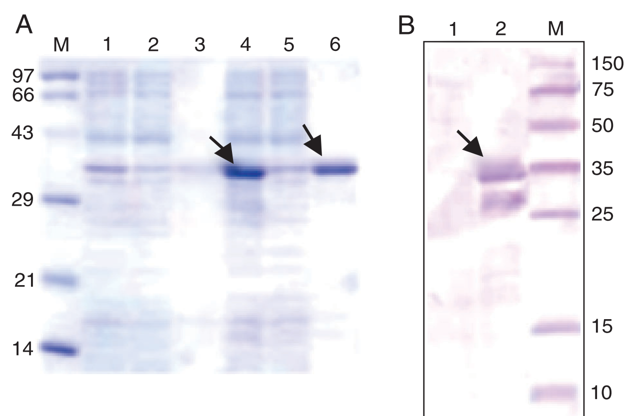

Figure 4. Expression rDer f 1 in *Escherichia coli* BL21 cells. *E. coli* BL21 cells were transformed with either pET28a(+)-Der f 1 or empty vector pET28a(+) as control. SDS-PAGE (stained with CBB-R250) and Western blotting of rDer f 1 protein were performed as described in Material and Methods. *A*, SDS-PAGE analysis of rDer f 1 protein; *lane 1*, the whole cell lysate of *E. coli* BL21 cells containing pET28a; *lane 2*, the supernatant of cells containing pET28a; *lane 3*, the pellet of cells containing pET28a; *lane 4*, the whole cell lysate of *E. coli* BL21 cells containing pET28a(+)-Der f 1; *lane 5*, the supernatant of cells containing pET28a(+)-Der f 1; *lane 6*, the pellet of cells containing pET28a(+)-Der f 1. *B*, Western blotting of rDer f 1 protein; *lane 1*, the whole cell lysate of *E. coli* BL21 cells containing pET28a; *lane 2*, the whole cell lysate of *E. coli* BL21 cells containing pET28a(+)-Der f 1. *M* represents protein markers. The arrows point to rDer f 1 whose molecular mass is about 34 kDa.

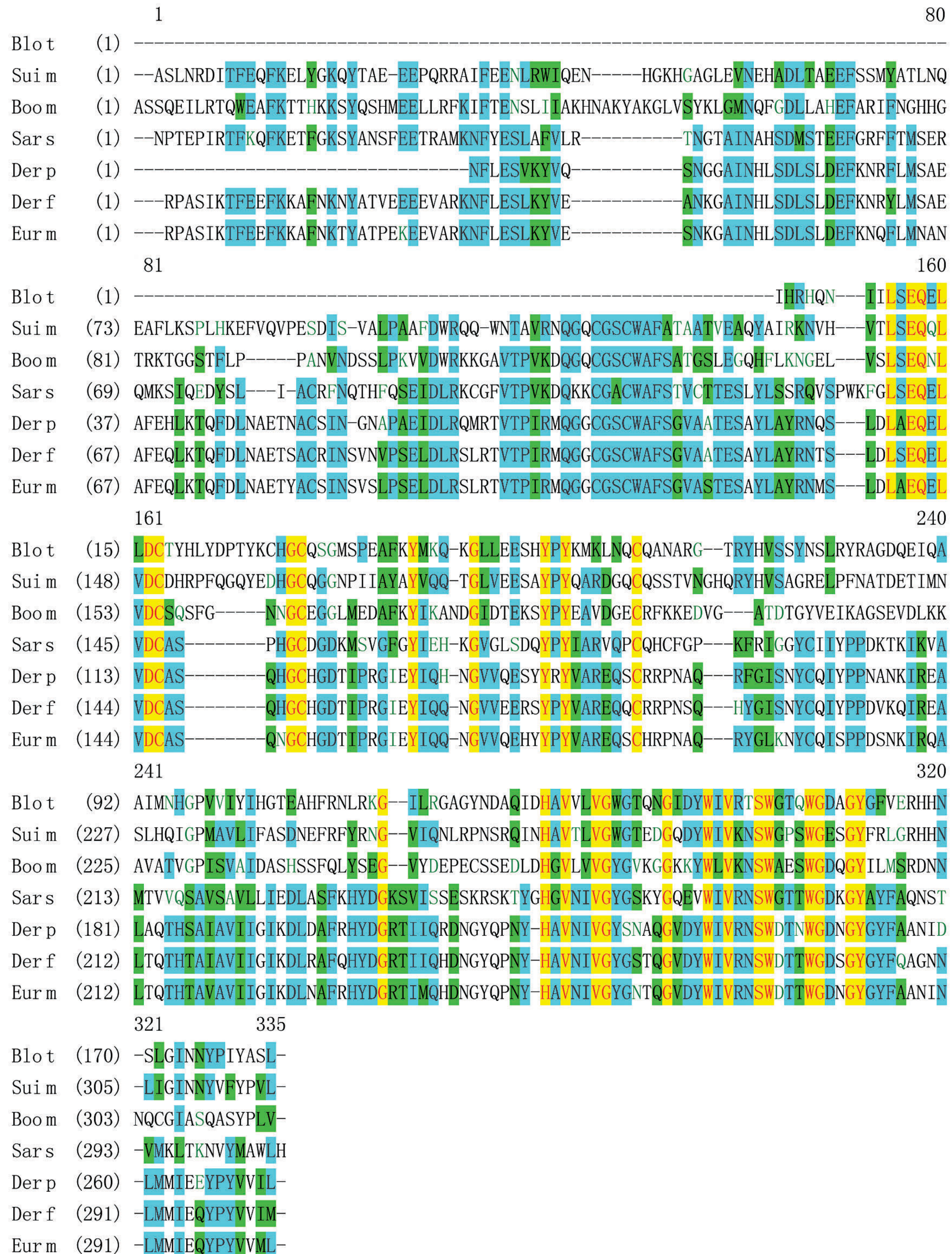

Figure 5. Alignment of Der f 1 with other group 1 allergens and cysteine proteases from ticks and mites determined using the VECTOR NTI 9.0 software. Group 1 allergens for *Dermatophagoides farinae*, *Blomia tropicalis*, *D. pteronyssinus*, *Euroglyphus maynei*, *Sarcoptes scabiei*, and *Suidasia medanensis* are abbreviated as Der f, Blo t, Der p, Eur m, Sar s, and Sui m, respectively. The tick *Boophilus microplus* is abbreviated as Boo m.

Table 1. Analysis of the similarity (%) of Der f 1 and other group 1 allergens or cysteine proteases from ticks and mites using the VECTOR NTI 9.0 software.

	Blo t	Sui m	Boo m	Sar s	Der p	Der f	Eur m
Blo t	–	29	18	20	32	24	23
Sui m	11	–	32	27	29	33	31
Boo m	12	20	–	27	24	28	28
Sar s	10	19	14	–	44	48	47
Der p	11	16	14	14	–	77	78
Der f	10	15	14	15	6	–	88
Eur m	10	17	13	15	7	6	–

Der f, Blo t, Der p, Eur m, Sar s, and Sui m are group 1 allergens of *Dermatophagoides farinae*, *Blomia tropicalis, D. pteronyssinus, Euroglyphus maynei, Sarcoptes scabiei*, and *Suidasia medanensis*, respectively, and Boo m represents the tick *Boophilus microplus*. For sequence source information, see the section "Amino acid sequence homology analysis, alignment, and phylogeny".

Table 2. Prediction results for protein subcellular localization by CELLO version 2.5 (http://cello.life.nctu.edu.tw/).

SVM	Localization	Reliability
Amino acid composition	Extracellular	0.528
N-peptide composition	Extracellular	0.517
Partitioned sequence composition	Extracellular	0.765
Physicochemical composition	Cytoplasmic	0.292
Neighboring sequence composition	Extracellular	0.741
CELLO prediction		
	Extracellular	2.609*
	Cytoplasmic	0.743
	Nuclear	0.421
	Lysosomal	0.415
	Mitochondrial	0.173
	Plasma membrane	0.140
	Vacuole	0.131
	Peroxisomal	0.124
	Chloroplast	0.109
	Endoplasmic reticulum	0.093
	Golgi	0.023
	Cytoskeletal	0.021

SVD = support vector machine.

Amino acid sequence homology analysis, alignment and phylogeny

The homology between the putative amino acid sequence of Der f 1 and other proteins was determined by comparing their sequences in all non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF, excluding environmental samples, using BLASTp in the NCBI website. Based on our BLASTp search results, the Der f 1 amino sequences of ticks and mites, i.e., *Blomia tropicalis* (GenBank AAK58415), *Boophilus microplus* (GenBank AAF61565), *D. pteronyssinus* (GenBank ABA39435), *E. maynei* (GenBank P25780), *Sarcoptes scabiei* (GenBank AAS93667), and *Suidasia medanensis* (GenBank AAX34043) were chosen and, after sequences coding for signal peptides were deleted, they were aligned using the Vector NTI 9.0 software. A phylogenetic tree was then constructed using the Treeview 5.0 1.81 software. Interestingly, Der f 1 had 88% identity to Eur m 1 in the amino acid sequence, while it had only 77% sequence identity to Der p 1 (Figure 5 and Table 1). And *D. farinae* was clustered with *E. maynei* but not with *D. pteronyssinus* in the polygenic tree for their group 1 allergens (Figure 6).

Prediction of protein subcellular localization, specific motifs, secondary structure, and three-dimensional structures

Using CELLO ver. 2.5, it was concluded that the subcellular localization of Der f 1 was extracellular (Table 2). Four different softwares (InterProScan, ScanProsite, PPSearch, and PROSITE) were used in the present study to predict the specific motifs of the Der f 1 protein. Interestingly, all analyses showed the same results. There are three cysteine

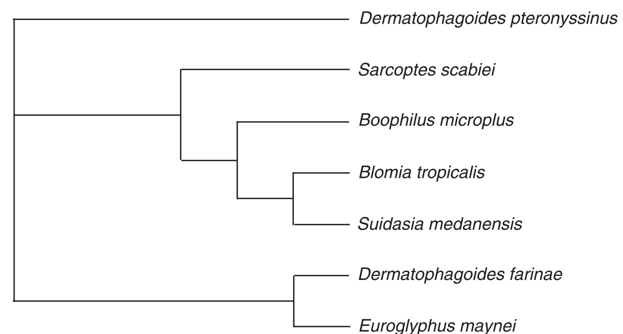

Figure 6. Phylogenetic tree constructed for Der f 1 and other group 1 allergens or cysteine proteases from ticks or mites using ClustalX 1.81.

peptidase active sites in this protein, respectively positioned at 127-138 (QGGCGSCWAFSG), 267-277 (NYHAVNIV GYG) and 284-303 (YWIVRNSWDTTWGDSGYGYF). The secondary structure of Der f 1 was predicted using the SOPMA software and the results showed that 33.96% of the protein (109 amino acids) was the α helix, 17.13% (55 amino acids) was the extended strand, 5.61% (18 amino acids) the ß turn, and 43.30% (139 amino acids) were random coils (Figure 7). In order to further understand the spatial arrangement of the three cysteine peptidase active sites in the three-dimensional structure of this protein, we then imported the amino acid sequence into the Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do) and found a template (ID: 1XKG). Subsequently, we submitted both our amino acid sequence and the template to the Swiss-model server

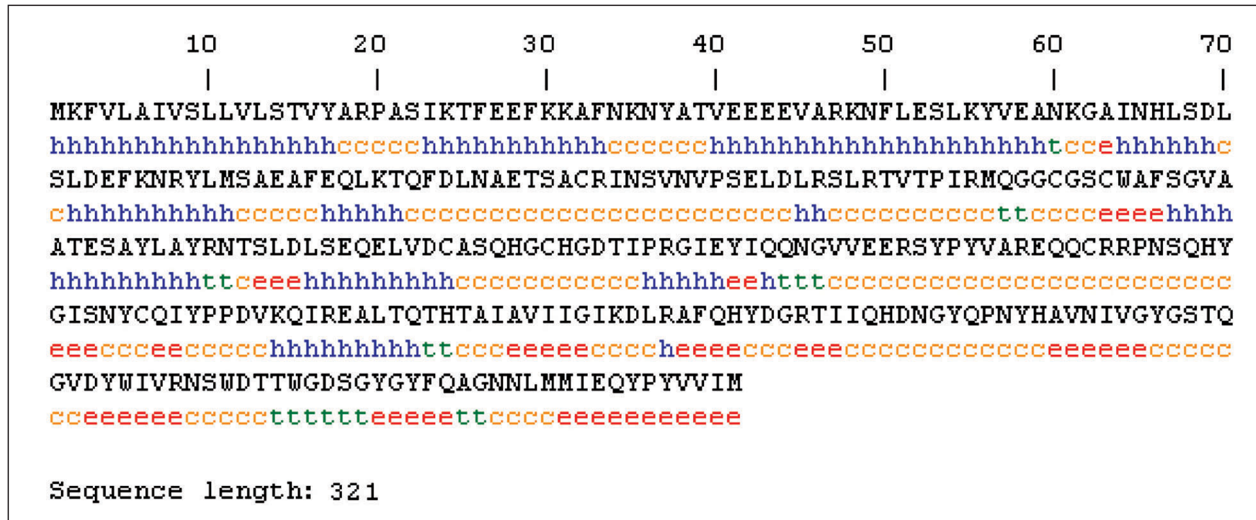

Figure 7. Secondary structure prediction of Der f 1 protein using the Sopma method (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). α helix (Hh); extended strand (Ee); ß turn (Tt); random coil (Cc).

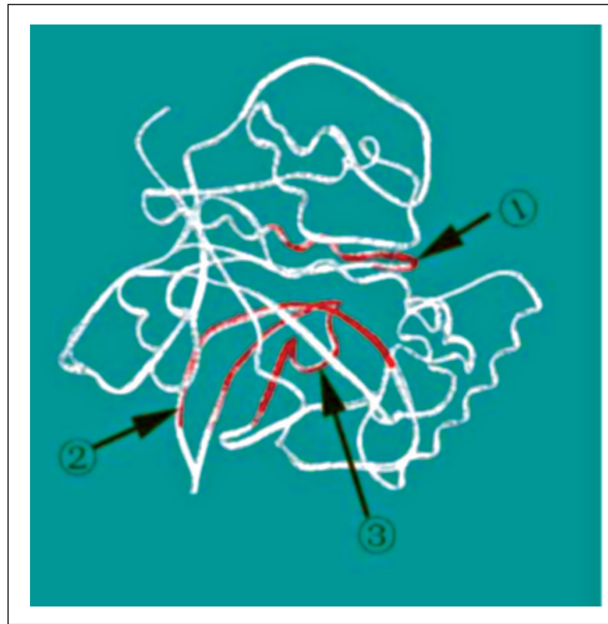

Figure 8. Three-dimensional model of the Der f 1 protein constructed by Swiss-PdbViewer version 3.7. The three cysteine peptide active regions are colored red and indicated by the arrows: 1) residues 127-138 (QGGCGSCWAFSG), 2) residues 267-277 (NYHAVNIVGYG), and 3) residues 284-303 (YWIVRNSWDTTWGDSGYGYF).

(http://swissmodel.expasy.org), and obtained a simple three-dimensional model of this protein in which we highlighted the three cysteine peptidase active sites using the Swiss-Pdb Viewer version 3.7 (Figure 8).

Discussion

A large body of evidence indicates that recombinant allergens are helpful to identify and produce the important IgE-binding determinants present on proteins which cause allergic reactions to house dust mites (5,6). In the present study, a cDNA coding for the major house dust mite allergen Der f 1 was cloned and its DNA sequence was determined. A complete ORF was found in the cloned fragment with a full length of 966 bp. Sequence comparisons showed only 1 bp difference between our fragment and that of the reference sequence (GenBank AB034946). Using the ProtParam Tools, the protein with 321 amino acids was found to have a molecular mass of 36,4189 kDa. However, this is only a theoretical prediction, since the sequence contained characteristic features of signal peptides with 18 amino acid residues at the 5' proximal end. In addition, the pro-peptides with 80 amino acid residues that have been reported previously (3) should be included. Once the presumptive pre- (54 bp) and pro-peptides (240 bp) were deleted, the remaining nucleotide sequence (672 bp) was predicted by the ProtParam Tools and the results showed a molecular mass of 25,148 kDa, the same as that reported earlier for Der f 1 sequence (3).

The active Der f 1 protein with 223 residues has received a lot of attention and was cloned and expressed. In contrast, subsequent experiments demonstrated that Der f 1 directly expressed in *E. coli* without the pro-sequence had very low IgE binding activity. For example, Best et al. (7) insisted that the pre- or pro-sequence ought to result in

secretion of the mature form of the protein and that both the α factor signal peptide and the pro-enzyme region were efficiently processed during secretion, and the recombinant forms of Der f 1 (rDer f 1) expressed with pro-sequence in *Pichia pastoris* had similar IgE binding activity to native forms of Der f 1 (nDer f 1) obtained from mites. In the present study, we successfully expressed the allergen in *E. coli* BL21 with the complete preliminary sequence. Despite the absence of exact molecular weight data, we observed a single band of about 34 kDa using SDS-PAGE and Western blotting, in accordance with predicted results of 34,469 kDa with the sequence of the signal peptide (18 residues) being deleted.

Interestingly, sequence analyses suggest that *D. farinae* was evolutionarily more similar to *E. maynei* than to *D. pteronyssinus*, even though *D. farinae* and *D. pteronyssinus* are classified as belonging to the same *Dermatophagoides* genus. Firstly, Der f 1 had 88% identity to Eur m 1 in the amino acid sequence, while it had only 77% sequence identity to Der p 1. Secondly, *D. farinae was* clustered with *E. maynei* but not with *D. pteronyssinus* in the polygenic tree for their group 1 allergens. The results suggest either one of two hypotheses. The first is that the taxonomy for mites and ticks on the basis of morphology may not correctly reflect their true evolutionary history. Alternatively, the sequences of cysteine proteases may not correctly reflect the evolution of ticks and mites. More genes need to be analyzed to test these two hypotheses.

It is widely believed that the major dust mite allergen Der f 1 may be a digestive protease secreted by the gastrointestinal tract of mites (1,2,8). Based on its biochemical and biophysical properties, the predicted subcellular localization of Der f 1 was also extracellular, in agreement with previous observations. Our bioinformatic analyses identified three cysteine peptidase active sites and these site data were used to construct a simple three-dimensional model of the molecule.

In conclusion, we successfully cloned, sequenced and expressed Der f 1 in *E. coli* BL21, and identified some of its molecular characteristics using bioinformatic software. In particular, we drew the preliminary conclusion that *D. pteronyssinus* is evolutionarily more similar to *E. maynei* than to *D. farinae*, which will be confirmed using some molecular markers in the future.

Acknowledgments

We are grateful to Dr. Wang Bo, Faculty of Health Sciences, University of Cape Town, South Africa, to Dr. Xu Jianping, Department of Biology, McMaster University, Canada, and to Dr. M. Medeiros Jr., Serviço de Imunologia, Hospital Universitário Prof. Edgar Santos, Salvador, BA, Brazil, for their patient revision of our English writing.

References

1. Thomas WR, Smith WA, Hales BJ. The allergenic specificities of the house dust mite. *Chang Gung Med J* 2004; 27: 563-569.
2. Thomas WR, Smith WA, Hales BJ, Mills KL, O'Brien RM. Characterization and immunobiology of house dust mite allergens. *Int Arch Allergy Immunol* 2002; 129: 1-18.
3. 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.
4. Cui YB, Ling YZ, Zhou Y, Feng ZW, Xing YR, Zhang SW. An effective indirect fluorescent antibody test for diagnosis of intestinal acariasis. *Southeast Asian J Trop Med Public Health* 2006; 37: 452-455.
5. Smith W, Mills K, Hazell L, Hart B, Thomas W. Molecular analysis of the group 1 and 2 allergens from the house dust mite, *Euroglyphus maynei*. *Int Arch Allergy Immunol* 1999; 118: 15-22.
6. Mora C, Flores I, Montealegre F, Diaz A. Cloning and expression of Blo t 1, a novel allergen from the dust mite *Blomia tropicalis*, homologous to cysteine proteases. *Clin Exp Allergy* 2003; 33: 28-34.
7. Best EA, Stedman KE, Bozic CM, Hunter SW, Vailes L, Chapman MD, et al. A recombinant group 1 house dust mite allergen, rDer f 1, with biological activities similar to those of the native allergen. *Protein Expr Purif* 2000; 20: 462-471.
8. Tovey ER, Chapman MD, Platts-Mills TA. Mite faeces are a major source of house dust allergens. *Nature* 1981; 289: 592-593.