

***Dermatophagoides farinae* Allergen Der f 9: Cloning, Expression, Purification, Characterization, and IgE-Binding in Children With Atopic Asthma**

Yubao Cui, PhD,^{1,2*} Feixiang Teng, MS,² LiLi Yu, MS,² Ying Zhou, MS,¹ Chengbo Zhang, BM,² and Li Yang, BM²

Summary. Background: The house dust mite species *Dermatophagoides farinae* releases allergens that cause allergies and asthma worldwide. This study sought to clone and express the full-length cDNA encoding the group 9 allergen of *D. farinae* (Der f 9). Methods: The published sequence of Der f 9 was used to design primers for RT-PCR and RACE to obtain the full-length cDNA encoding Der f 9. After removal of signal peptide sequence, Der f 9 was then sub-cloned into plasmid pET-28b (+), and the plasmid was transformed into *Escherichia coli* BL21 (DE3) cells for expression. The recombinant protein was purified by Nickel affinity chromatography, identified by SDS-PAGE, Western blotting, dot blotting, and MALDI-TOF, and tested by ELISA for IgE reactivity with sera from children with asthma. Bioinformatics analyses were used to identify features of Der f 9. Results: By RT-PCR, 3'-RACE, and 5'-RACE, the full-length sequence of Der f 9 was generated, which was confirmed by nucleotide sequencing. The mature Der f 9 was expressed successfully in *E. coli*, which was identified by SDS-PAGE. The recombinant allergen was purified by chromatography and confirmed by SDS-PAGE, Western blotting, dot blotting, and MALDI-TOF. Sera from 56.7% (17/30) of mite-allergic patients reacted with the purified recombinant Der f 9. Conclusions: The successful production of recombinant Der f 9 protein revealed the importance of Der f 9 in mite allergy, and provides a foundation for further study of this allergen in diagnosis and treatment of symptoms. **Pediatr Pulmonol.** © 2016 Wiley Periodicals, Inc.

Key words: *Dermatophagoides farinae*; recombinant allergen; molecular cloning; house dust mite.

Funding source: National Natural Sciences Foundation of China, Numbers: 30060166, 81001330, 31272369, 31572319.

INTRODUCTION

House dust mites produce allergies and allergy-related symptoms in populations worldwide. Current treatments for mite allergies are allergen avoidance and alleviation of symptoms by pharmacologic intervention. Recent efforts have been aimed at developing allergen-specific immunotherapies that might change the course of disease, but with fewer side effects and better standardization than existing immunotherapies.¹

At least 33 groups of house dust mite allergens have been identified (<http://www.allergen.org/>); 31 are derived from the *Dermatophagoides* spp. of house dust mites. Of these, the groups 1 and 2 allergens contribute to the majority of dust mite allergy, reacting with sera from 80% of mite-sensitive patients. Because of this high reactivity, recombinant groups 1 and 2 allergens have been generated from *Dermatophagoides pteronyssinus*. However, the recombinant Der p 1 is only weakly reactive, while the recombinant Der p 2 exhibits IgE reactivity comparable to the native allergen.^{2,3} Although

the groups 1 and 2 allergens are the major allergens from *Dermatophagoides* spp., other allergens from these species contribute to allergic disease. Thus, researchers

¹Department of Clinical Laboratory, Affiliated Yancheng Hospital, School of Medicine, Southeast University, Yancheng 224001, Jiangsu Province, P. R. China.

²Department of Laboratory Medicine, Yancheng Health Vocational and Technical College, Yancheng 224006, Jiangsu Province, P. R. China.

Conflicts of Interest: None.

*Correspondence to: Yubao Cui, PhD, Department of Clinical Laboratory, Affiliated Hospital, School of Medicine, Southeast University, No. 299 at Jiefangnan Road, Yancheng 224000, Jiangsu Province, P. R. China. E-mail: ybcui1975@hotmail.com

Received 15 March 2016; Revised 8 June 2016; Accepted 3 July 2016.

DOI 10.1002/ppul.23526
Published online in Wiley Online Library
(wileyonlinelibrary.com).

are working toward the development of recombinant allergens that mimic the native proteins and can be used for safer and more effective allergy treatments.⁴⁻⁶

As in much of the world, house dust mite hypersensitivity, particularly to *D. farinae*, is common in China.⁷ To better understand house dust mite hypersensitivity, our laboratory has cloned and expressed several allergens from *D. farinae*.⁸⁻¹⁶ The current study sought to clone the group 9 allergen from *D. farinae* (Der f 9), which has not yet been listed in the website of allergen nomenclature (<http://www.allergen.org/>). These results will provide a foundation for improved diagnosis and treatment of house dust mite allergy in China and worldwide.

MATERIALS AND METHODS

Serum Sources

Serum samples from children with allergic asthma were provided by Department of Laboratory, the first affiliated hospital of Shanghai Jiao Tong University. Patients had not received any treatment for allergy. Asthma diagnosis was made following the WHO criteria¹⁷; all children had cough, asthma, obvious wheeze, rhonchus, and moist rale in the lung. X-ray examination showed increased lung markings or lung hyperinflation and infection. The diagnosis of allergy to dust mite allergens was established on the basis of a suggestive clinical history and positive serum IgE to *D. farinae* extracts as determined by the Allergy Screening test panel for atopy (Mediawiss Analytic, Moers, Germany). The inclusion criteria were (1) allergy to *D. farinae* with more than 1 year of follow-up, and (2) levels of specific IgE to *D. farinae* extracts higher than 3 kU/L at the time of inclusion into the study. A total of 30 pediatric patients, mean age (4.83 ± 2.04) years, were enrolled as serum-positive for *D. farinae*, and then used for detection of specific IgE-binding to recombinant fusion allergen. Additionally, the sera of five non-atopic healthy children, mean age (5.17 ± 1.17) years, were treated as negative controls. The study was approved by the Ethics Committee for Clinical Investigation of Affiliated Yancheng Hospital, School of Medicine, Southeast University, and written informed consent was obtained from the parents of all participants.

Obtaining the Full-Length Der f 9 Sequence

D. farinae were cultured and isolated as we previously reported.⁸⁻¹⁵ Total RNA was isolated using RNA isolator (TaKaRa Biotech, Dalian, China. No. D312) and stored at -80°C . DNA primers were designed and synthesized based on the published partial sequence of Der f 9 (GenBank Accession No. AY283282). The sequences of the forward primers were (F₁) 5'-TATTTTCATTGACCACAGCCGAACC-3' and (F₂)

5'-ACTTCAACTTACTGGCTGGGGTAG-3' and of the reverse primer was (R₁) 5'-GTGTTTGACAGACAGGTG-GACGGG-3'. Reverse transcription (RT) was performed with High Fidelity PrimeScript™ RT-PCR Kit (TaKaRa Code No. DR027A). The nested polymerase chain reaction (Nested PCR) was conducted with primers F₁ and R₁ in the first PCR reaction using the RT product as template, and with primers F₂ and R₁ in the second PCR reaction using the first PCR reaction products as template. Products of the second PCR, recovered with MiniBEST Agarose Gel DNA Purification Kit Ver.2.0 (TaKaRa Code No. D823A), were submitted for nucleotide sequencing.

Based on this obtained Der f 9 partial sequence, the primers R₂ (5'-CATTCGATTAGTGATGGCATTTCAC-3') and R₃ (5'-TGAGCAGCAGTCAATACGGTTC-3') were designed and synthesized, and 5'-RACE was conducted with a 5'-Full RACE Kit (TaKaRa Code No. D315) as follows: First, *D. farinae* total RNA was treated with CIAP (Calf-intestinal alkaline phosphatase) and TAP (tobacco acid pyrophosphatase), linked with a 5'-RACE Adaptor (5'-Full RACE Kit, TaKaRa Code No. D315), and subjected to reverse transcription to obtain the cDNA fragment. Second, the nested PCR was conducted with primer R₁ and the 5'-RACE Outer Primer in the first PCR reaction using the RT product as template, and with primers R₂ and 5'-RACE Inner Primer in the second PCR reaction using the first PCR reaction product as template. Products of the second PCR were recovered with an Agarose Gel DNA Purification Kit Ver.2.0 (TaKaRa Code No. DV805A) and submitted for nucleotide sequencing with primers R₂ and R₃.

Primers F₃ (5'-GGGGTTCAGTAAATGCCATCACTA-3') and F₄ (5'-TGCCATC ACTAATCGAATGTTGTG-3') were designed and synthesized for 3'-RACE. First, the cDNA was obtained by RT with 3'-Full RACE Core Set Ver.2.0 (TaKaRa Code No. D314). Second, the nested PCR was conducted with primer F₃ and the 3'-RACE Outer Primer in the first PCR reaction using the RT product as template, and with primer F₄ and the 3'-RACE Inner Primer in the second PCR reaction using the first PCR reaction products as template. The second PCR products were recovered and sequenced with primer F₄.

To verify the sequence by RACE, the following primers were synthesized: YF1, 5'-CTATTCGTCTTCCATTGCC-TAA-3', YF2, 5'-AGAATCGTTGGCGGTGAAAATG-3', and YR, 5'-ATGATCCAATTACGTAGAG-3'. The first PCR reaction was conducted using the cDNA products of the 3'-RACE reaction as template and with primers YF1 and YR, and the second PCR reaction was conducted with primers YF2 and YR using products of the first PCR as template.

Primers were designed to amplify full-length Der f 9 from *D. farinae*: forward (F₅), 5'-TACCCTC-GAGGGATCCATGAAATTCGCCACCATTTTTGTA-3'; reverse (R₄), 5'-TAGACTGCAGGTCGACTTAACTG-TATTCGAAATGATCC-3'. These included *Bam* H I and

Sal I sites, respectively, at their 5' ends (underlined). The PCR was performed with PrimeSTAR[®] HS DNA Polymerase (TaKaRa Code No. DR010A) and products were identified by using the Agarose Gel DNA Purification Kit Ver 2.0 (TaKaRa Code No. DV805A). After the PCR-amplified DNA was recovered with a MiniBEST Agarose Gel DNA Purification Kit Ver 2.0 (TaKaRa Code No. D823A), it was then cloned into pCold-TF with the DNA Ligation Kit (TaKaRa Code No. D6020A). *Escherichia coli* JM109 (TaKaRa Code No. D9052) was then transformed with the recombinant plasmid pCold-TF-Der f 9, positive clones were selected by blue/white screening on Luria-Bertani (LB) plates containing 100 µg/ml ampicillin, and samples were submitted for automatic DNA sequencing on ABI PRISM[™] 377XL DNA Sequencer (Perkin Elmer) with primers for pCold-TF: VP1 (5'-TGGCGAAAGCGAAAGTGACT-3') and VP2 (5'-GACGAAAGGGCCTGCGC ATT-3').

Construction of Expression Plasmid pET-28b(+)-Der f 9

Bioinformatics analyses of the full-length Der f 9 allergen identified a 19-residue signal peptide sequence (<http://www.cbs.dtu.dk/services/SignalP/>) and an *Nco* I site. Primers were designed and synthesized to amplify the mature Der f 9 allergen with a mutated *Nco* I site: F₆, 5'-CTCCATGGATCATAGTGAAGAACAAGCTATTC-3' (*Nco* I site underlined); R₅, 5'-GTTTCATTGATCATCAAGGAAAACG-3'; F₇, 5'-CGTTTTCTTGGATGATCAATGAAC-3'; R₆, 5'-TCCTCGAGAACTGTATTCGAAATGATCCAATTACG-3' (*Xho* I site underlined). Two PCRs were performed using the pCold TF-Der f 9 as template, one with primers F₆ and R₅, and one with primers F₇ and R₆. Using these two PCR products as template, another PCR was conducted with primers F₆ and R₇ to amplify the mature Der f 9 allergen with mutated *Nco* I site.

Recovered PCR products were digested by *Nco* I and *Xho* I, and then sub-cloned into the expression vector pET-28b(+) (Novagen, Kit Lot No. 69865) to create pET-28b(+)-Der f 9 using the DNA Ligation Kit (TaKaRa Biotech Co. Ltd, Code No. D6023). *E. coli* JM109 competent cells (TaKaRa Biotech Co. Ltd, Code No. D9052) were transformed with pET-28b (+)-Der f 9 plasmids, positive clones were selected by blue/white screening, subjected to restriction digestion with *Nco* I and *Xho* I, and confirmed by nucleotide sequencing.

Production and Purification of Recombinant Fusion Protein

According to our previous reports,¹⁶ the verified pET-28b-Der f 9 plasmids were transformed and expressed in *E. coli* BL21 (DE3, Stratagene, La Jolla, CA), purified by Nickel affinity chromatography, identified by SDS-PAGE,

Western blotting, and analyzed by MALDI-TOF/TOF mass spectrometer (model 4800, Applied Biosystems).

Dot Blotting

The recombinant allergen was spotted on to a piece of membrane, let air dry for 30 min, blocked with bovine serum albumin (BSA, 3%) for one hour at room temperature, the serum (1:5 dilution with PBS) was incubated and placed overnight at 4°C. Secondary antibody (50 µl HRP-mouse anti-human IgE, 1:5,000, Sigma Immunochemicals, St. Louis, MO) was added and samples were incubated at 37°C for 1 h. A 100 µl DAB (3,3N-Diaminobenzidine Tetrahydrochloride, Sigma) substrate solution were added, and samples were incubated at room temperature for 20 min. The four sera samples used in dot blotting were the first four sera in the subsequent IgE-binding ability test by enzyme-linked immunosorbent assay (ELISA).

Specific IgE-Binding to Recombinant Allergen

The IgE-binding ability of the recombinant allergen was detected by ELISA. For antigen coating, 100 µl of rDer f 9 (1 µg/ml) diluted in coating buffer (0.1 M phosphate-buffer saline) were added to a 96-well microtiter plate and incubated at 8°C overnight. After blocking with 300 µl of PBST buffer (8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, 0.24 g KH₂PO₄, 2 ml tween-20, 800 ml distilled H₂O), 100 µl of patient serum individually (diluted 1:4 with blocking buffer) were added, and samples were incubated overnight at 4°C. Secondary antibody (50 µl HRP-mouse anti-human IgE, 1:5,000, Sigma Immunochemicals) was added and samples were incubated at 37°C for 1 h. A 100 µl tetramethylbenzidine (TMB) substrate solution were added, and samples were incubated at room temperature for 15 min. Reaction was terminated upon addition of 50 µl NH₂SO₄. Optical density was measured at the 405 nm wavelength (OD₄₀₅) for three replicates using an iMark Microplate Absorbance Reader (Bio-Rad). The reactivity was considered to be positive if the optical density (OD) values of the detected sera were higher than the cut-off ELISA value (mean ELISA value of healthy donors ± 3 standard deviations).

Bioinformatic Characterization of the Cloned Der f 9

The open reading frame (ORF) was determined by using the ORF finder from NCBI (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>). Using the SIB Bioinformatics Resource Portal (ExPaSy, <http://www.expasy.org/>), the amino acid sequence of Der f 9 was deduced by Translate Tools, its physico-chemical properties by ProtParam, its hydrophilicity by ProtScale, its phosphorylation sites by

NetPhos2.0 Server, its secondary structure by GOR4.0, and its functional sites by InterProScan, ScanProsite, and MotifScan. Homology search and sequence alignment were performed with ClustalW omega through the European Bioinformatics Institute (<http://www.ebi.ac.uk/>) and VECTOR NTI 11.5.1 (IBI, New Haven, CT).

RESULTS

Full-Length Sequence of Der f 9

Using the GenBank-available partial sequence encoding for Der f 9, RT-PCR was used to generate a template for 3'-RACE and 5'-RACE; these techniques were used to obtain the full-length sequence of Der f 9. The full-length sequence encoding Der f 9 was amplified successfully from total RNA of *D. farinae* and was confirmed by agarose gel electrophoresis. This sequence was inserted into a pCold plasmid for sequencing. Its sequence was 99.85% (674/675) identical with the reference sequence (GenBank No. AY283282), but with a dissimilarity in the 5'-end; it was and 98.43% (818/831) identical with GenBank Sequence No. KM010013, with consistent sequence in the 5'-end (Fig. 1). The sequencing results showed that one ORF of Der f 9 was 831 bp long, from the start codon ATG to the stop codon TAA.

Generation of Recombinant Der f 9

Bioinformatic analyses identified a 19-residue signal peptide and an *Nco* I site in the full-length Der f 9. Using the plasmid pCold TF-Der f 9 as a template, nested PCR was used to remove the signal peptide and mutate the *Nco* I site that will affect the Der f 9 gene subcloned into pET-28b(+) via *Nco* I and *Xho* I sites. The nested-PCR product and the plasmid pET-28b(+) were digested by *Nco* I and *Xho* I, respectively, and used to generate the fusion construct. The recombinant pET-28b(+)-Der f 9 was then transformed into BL21(DE3) and identified by PCR.

The *E. coli* BL21 (DE3) containing the plasmid pET28a(+)-Der f 9 was induced for protein expression by IPTG. A single, specific band matching the predicted molecular weight of Der f 9 was identified at ~28 kDa by SDS-PAGE (Fig. 2A). After the induced bacteria were collected and centrifuged, the recombinant allergen was purified by chromatography, and a specific band was detected by SDS-PAGE (Fig. 2B). The purified product was analyzed by MALDI-TOF/TOF. The peptide mass fingerprint (PMF) is shown in Figure 2C and D.

Dot Blotting

Sera samples from four pediatric patients with allergic asthma and from one non-allergic child were mixed with

the purified Der f 9 allergen, separately. As shown in Figure 3, the recombinant fusion protein can bind with sera from patients with allergic asthma, but do not react with sera from healthy control.

IgE-Binding Reactivity of the Recombinant Allergen

Sera samples from 30 pediatric patients with allergic asthma and from five non-allergic children were exposed to purified Der f 9 to determine IgE reactivity. As shown in Figure 4, the mean OD value of the five healthy volunteers was 0.222 ± 0.017 , and the postulated cut-off value was 0.273. The average rDer f 9-specific IgE OD value from 30 asthmatic patients was 0.294 ± 0.102 . Seventeen of them (56.7%) were considered Der f 9-positive.

Structural and Functional Prediction for Der f 9

The nucleotide sequence of Der f 9 was used to predict the amino acid sequence. Its most likely cleavage site was deduced between amino acid (aa) positions 19 and 20 by SignalP 3.0 Server (Fig. 5A). After removal of the signal peptide sequence, Der f 9 comprised 257 amino acids, with an estimated molecular weight of 27.8 kDa and a theoretical isoelectric point (pI) of 8.23. The instability index (II) was predicted as 35.64, which classifies the protein as stable. The grand average of hydropathicity (GRAVY) was computed as -0.147 . Combined with analysis using the ProScale software scale Hphob./Kyte & Doolittle, these results demonstrate the hydrophilicity of the protein (Fig. 5B). The average flexibility was calculated with a maximum score of 0.497 at aa position 40 and minimum score of 0.366 at aa position 73 (Fig. 5C). The relative mutability was calculated with a maximum score of 103.111 at aa position 136 and a minimum score of 54.556 at aa position 156 (Fig. 5D). Using the scale Polarity/Zimmerman, individual values were scored at a maximum of 28.714 at residue 5 and at a minimum of 0.652 at residue 226 (Fig. 5E). The accessible residues (%) were predicted with min score of 3.567 at residue 20 and max score of 7.278 at residue 206 (Fig. 5F). NetPhos2.0 identified 14 phosphorylation sites: 10 serines (aa 3, 41, 56, 82, 84, 110, 113, 179, 202, and 206), one threonine at aa 92, and three tyrosine sites at aa 85, 112, 242, and 230 residues. The secondary structure of Der f 9, predicted using GOR4.0, revealed that 10.89% (28 amino acids) of the protein is in alpha helix, 32.30% (83 amino acids) is extended strand, and 56.81% (146 amino acids) is random coils (Fig. 6). Subcellular localization of Der f 9 was predicted to be plasma membrane and extracellular (Table 1). Further, Der f 9 was deduced as a peptidase S1A, chymotrypsin-type (IPR001314) protein family member with trypsin-like cysteine/serine peptidase domain (Table 2). BLAST

AY283282	-----CC-----
KM010013	ATGAAATTCGCCACCATTTTGTATTTATTGCTATTGGAACATCCGTAGTGATCGGTGA-
Sequencing	ATGAAATTCGCCACCATTTTGTATTTATTGCTATTGGAACATCCGTAGTGATCGGTGAT
AY283282	-----GCACATGGACAG-----CCGAGATCTGGGTACCGACGACGACGA
KM010013	-----ACAAGCTATTGCTCTCCATTGCCTAAAGCTATAAATGAACG-----
Sequencing	CATAGTGAAGAACAAGCTATTGCTCTCCATTGCCTAAAGCTATAAATGAACG-----
AY283282	CAAGGCCATGGCTGATATCGGATCC-----GAATTCATCGTTGGCGGTGAAA
KM010013	-TTTTCATGGATGATCAATGAACCAATCACAAATGGTGAAAGAAATCGTTGGCGGTGAAA
Sequencing	-TTTTCATGGATGATCAATGAACCAATCACAAATGGTGAAAGAAATCGTTGGCGGTGAAA
AY283282	ATGCATCACCCGGTGATGCTATTATCAAATTGCATTGTTACGAAAAGATTCAATTACCT
KM010013	ATGCATCACCCGGTGATGCTATTATCAAATTGCATTGTTACGAAAAGATTCAATTACCT
Sequencing	ATGCATCACCCGGTGATGCTATTATCAAATTGCATTGTTACGAAAAGATTCAATTACCT
AY283282	GTGGTGGTTCACTTATATCATCAAGAACCGATTGACTGCTGCTCATTGTGTTTTGGTG
KM010013	GTGGTGGTTCACTTATATCATCAAGAACCGATTGACTGCTGCTCATTGTGTTTTGGGG
Sequencing	GTGGTGGTTCACTTATATCATCAAGAACCGATTGACTGCTGCTCATTGTGTTTTGGTG
AY283282	ATGAAGCATCACCATCATATTTCAAGATCCGTTACAATACATTAGATCGAATAATGGTC
KM010013	ATGAAGCATCACCATCATATTTCAAGATCCGTTACAATACATTAGATCGAATAATGGTC
Sequencing	ATGAAGCATCACCATCATATTTCAAGATCCGTTACAATACATTAGATCGAATAATGGTC
AY283282	CAACGATTGGTGTAGTAAATTTATCGCCATAGTTTATATTCATCAACGACCATTGATT
KM010013	CAACGATTGGTGTAGTAAATTTATCGCCATAGTTTATATTCATCAACGACCATTGATT
Sequencing	CAACGATTGGTGTAGTAAATTTATCGCCATAGTTTATATTCATCAACGACCATTGATT
AY283282	ATGATGTTGCTACATTGATTTTATCTGAACCATTTACACCATCGGCAAAATGCCGATATTA
KM010013	ATGATGTTGCTACATTGATTTTATCTGAACCATTTACACCATCGGCAAAATGCCGATATTA
Sequencing	ATGATGTTGCTACATTGATTTTATCTGAACCATTTACACCATCGGCAAAATGCCGATATTA
AY283282	TTTCATTGACCACAGCCGACCAAGCTGATAATACACAACCTCAACTTACTGGCTGGGTA
KM010013	TTTCATTGACCACAGCCGACCAAGCTGATAATACACAACCTCAACTTACTGGCTGGGTA
Sequencing	TTTCATTGACCACAGCCGACCAAGCTGATAATACACAACCTCAACTTACTGGCTGGGTA
AY283282	GACTTAAATCCGGTGGTACAATACCTACCCCTTTACAGATTGCAACTATATCGAAAAATGT
KM010013	GACTTAAATCCGGTGGTACAATACCTACCCCTTTACAGATTGCAACTATATCGAAAAATGT
Sequencing	GACTTAAATCCGGTGGTACAATACCTACCCCTTTACAGATTGCAACTATATCGAAAAATGT
AY283282	CTAGAACAAAATGTTTCAATATTTGGGGTTCAGTAAATGCCATCACTAATCGAATGTTGT
KM010013	CTAGAACAAAATGTTTCAATATTTGGGGTTCAGTAAATGCCATCACTAATCGAATGTTGT
Sequencing	CTAGAACAAAATGTTTCAATATTTGGGGTTCAGTAAATGCCATCACTAATCGAATGTTGT
AY283282	GTGCACATAGTAAAAACAATCAGCATGCAATGGAGATTGTTGGTGGTCCATTAGTTTACA
KM010013	GTGCACATAGTAAAAACAATCAGCATGCAATGGAGATTGTTGGTGGTCCATTAGTTTACA
Sequencing	GTGCACATAGTAAAAACAATCAGCATGCAATGGAGATTGTTGGTGGTCCATTAGTTTACA
AY283282	ATGGACATCTGGTGGGAGTTGTTCTTGGGGCCCGTCCACCTGTCTGTCAAAACCTTATC
KM010013	ATGGACATCTGGTGGGAGTTGTTCTTGGGGCCCGTCCACCTGTCTGTCAAAACCTTATC
Sequencing	ATGGACATCTGGTGGGAGTTGTTCTTGGGGCCCGTCCACCTGTCTGTCAAAACCTTATC
AY283282	CAACCATTTATAGTAATGTTGCAACTCTACGTAATTGGATCATTTTGAATACAGTTTAAA
KM010013	CAACCATTTATAGTAATGTTGCAACTCTACGTAATTGGATCATTTTGAATACAGTTTAAA
Sequencing	CAACCATTTATAGTAATGTTGCAACTCTACGTAATTGGATCATTTTGAATACAGTTTAAA
AY283282	TGAAATATTATTAATTAATTATTGTCATTGTCATTTCATCCATTAAAGTCGCTTATTTTTT
KM010013	TGAAATATTATTAATTAATTATTGTCATTGTCATTTCATCCATTAAAGTCGCTTATTTTTT
Sequencing	TGAAATATTATTAATTAATTATTGTCATTGTCATTTCATCCATTAAAGTCGCTTATTTTTT
AY283282	TTATTCATTATGACACTTGTTTTGTGTGTGCACATTTTTTCAAAAATATCCGGTTTAC
KM010013	TTATTCATTATGACACTTGTTTTGTGTGTGCACATTTTTTCAAAAATATCCGGTTTAC
Sequencing	T-----
AY283282	ACTGTGCAAAAATAAAATACTTGCCCTTTAATAACAAAAA
KM010013	ACTGTGCAAAAATAAAATACTTGCCCTTTAATAACAAAAA
Sequencing	-----

Fig. 1. Full-length Der f 9. Alignment of the full-length Der f 9 nucleotide sequence was verified following insertion of the amplified sequence into a plasmid. The alignment shows near identity with the partial sequence listed in GenBank. Notes: “-” means that the residues or nucleotides in that column are identical in all sequences in the alignment; “:” indicates conserved substitutions; “.” indicates semi-conserved substitutions. The underlined nucleotide sequences correspond to residues of the signal peptide sequence.

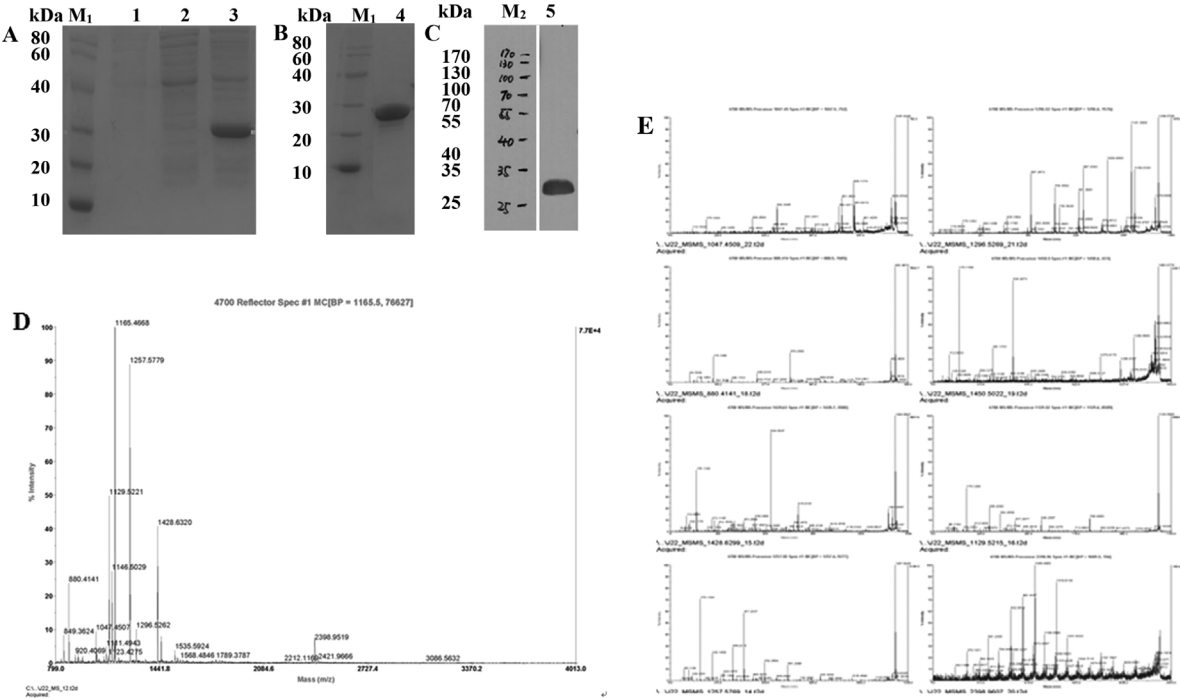


Fig. 2. Expression and purification of rDer f 9 in *E. coli* BL21 (DE3). (A) SDS-PAGE analysis of the protein expressed from the pET-28b(+)-Der f 9 recombinant plasmid in *E. coli* BL21 (DE3) cells. Lane M, Protein Marker (Beijing TransGen Biotech Co., Ltd.). Lane 1, Whole *E. coli* BL21 (DE3) containing pET-28b(+)-Der f 9 before induction. Lane 2, the supernatant from cells containing pET-28b(+)-Der f 9 at 3 hr after induction. Lane 3, the pellet of cells containing pET-28b(+)-Der f 9 at 3 hr after induction. (B) SDS-PAGE analysis of purified recombinant protein from *E. coli* BL21 (DE3) cells. Lane M, Protein Marker (Beijing TransGen Biotech Co., Ltd.). Lane 4, the purified protein. (C) Western-blotting analysis of the rDer f 4 protein. Lane 3, The purified fusion protein rDer f 9; Lane M₂, Thermo Scientific Pierce Prestained Protein Molecular Weight Marker. (D) The first-order structure of the purified recombinant fusion protein from *E. coli* BL21 cells by MALDI-TOF-MS. (E) The second-order structure of the purified recombinant fusion protein from *E. coli* BL21 cells by MALDI-TOF-MS.

search and alignment identified proteins similar to Der f 9, including Der p 9 (*Dermatophagoides pteronyssinus*, GenBank Accession No. AAP57077), Sui m 9 (*Suidasia medanensis*, GenBank Accession No. AAX34056), Blo t 9 (*Blomia tropicalis*, AAQ24546), and Ale o 9 allergen (*Aleuroglyphus ovatus*, ABU50819). Of these sequences, 118 residues were identical, 41 residues were conserved substitutions, and 26 were semi-conserved

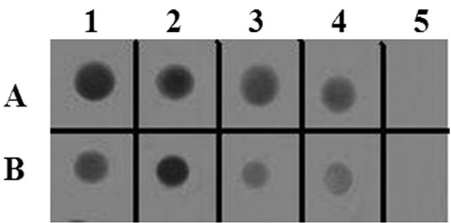


Fig. 3. Dot blotting analysis of recombinant Der f 9 allergen. Note: A. 2 µg of rDer f 9; B. 1 µg of rDer f 9; 1, 2, 3, 4, Sera of four patients allergic to *D. farinae*; 4, Serum of healthy control.

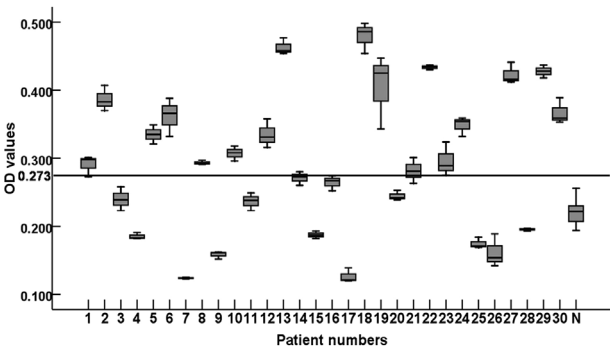


Fig. 4. The IgE-binding reactivity of the recombinant Der f 9 allergen. ELISA was used to detect the IgE-binding reactivity of the recombinant allergen to sera collected from mite-sensitive asthmatic patients (n=30). As a negative control, mite-insensitive healthy donors (n=5) were also exposed to Der f 9. The reactivity was considered to be positive if the optical density (OD) values of the exposed sera were higher than the cut-off ELISA value (mean ELISA value of healthy donors ± 3 S.D.) that is indicated as a horizontal line.

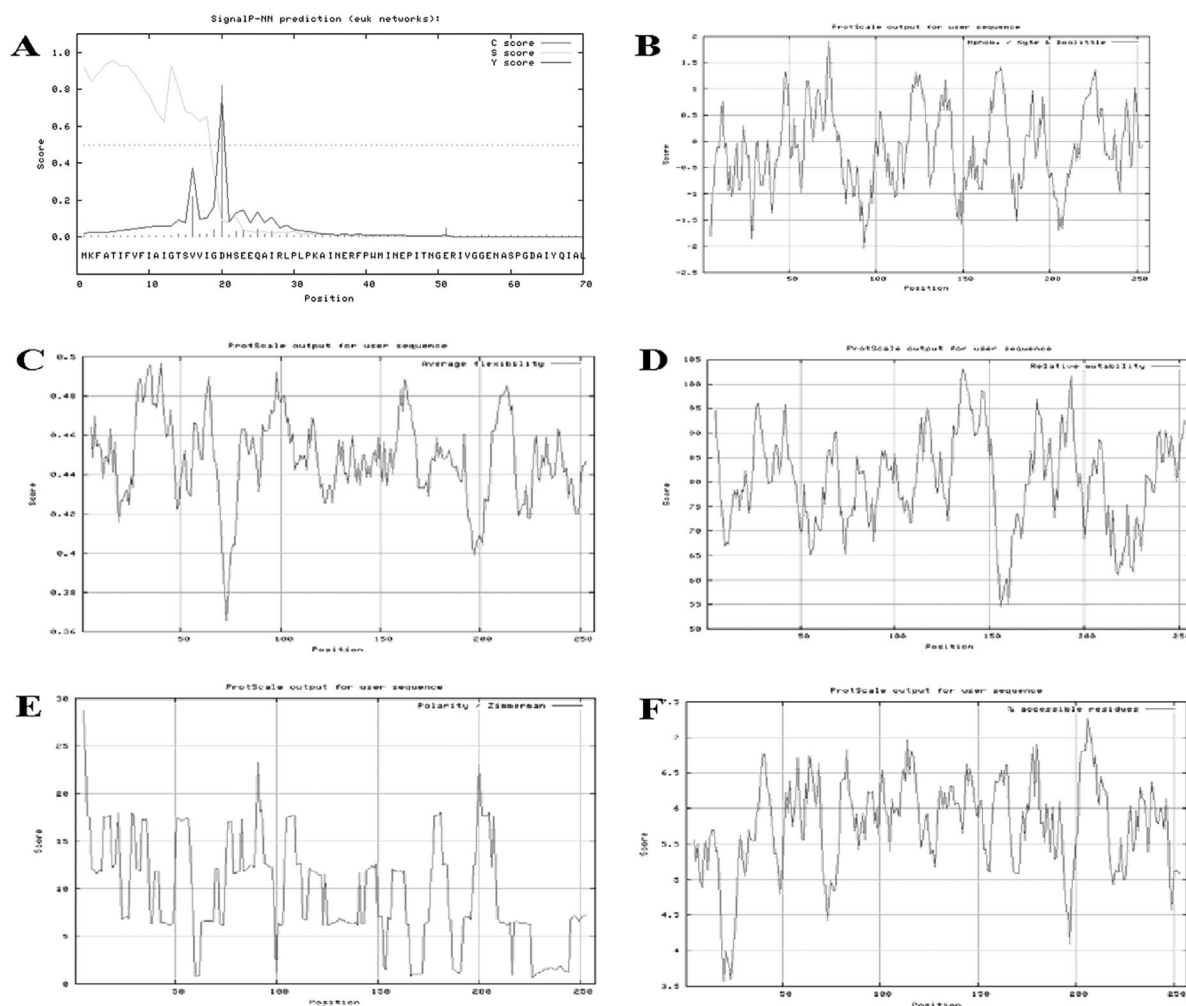


Fig. 5. Bioinformatic analyses using the predicted amino acid sequence of Der f 9. X-axis represents protein length from N- to C- terminal. Y-axis represents the score. The higher score, the higher probability. (A) SignalP-NN prediction (euk networks) by SignalP 3.0 software in SMART Server (<http://www.cbs.dtu.dk/services/SignalP/>). (B) Hydrophilicity analysis for Der f 9 allergen (ProtScale software, Kyte & Doolittle). (C) The average flexibility of Der f 9 (ProScale software). (D) The relative mutability of Der f 9 (ProScale software). (E) The Polarity/Zimmerman of Der f 9 (ProScale software). (F) The accessible residues of Der f 9 (ProScale software).

substitutions (Fig. 7). Sequence similarity was assessed using VECTOR NTI v11.0. The greatest similarity (85%) was found between Der f 9 and Der p 9; Der f 9 is 62%, 64%, and 58% similar to Blo t 9, Sui m 9, and Ale o 9, respectively (Table 3).

DISCUSSION

Since the 1980s, a considerable number of cDNAs encoding allergens have been obtained and expressed to produce recombinant allergens. Structural and

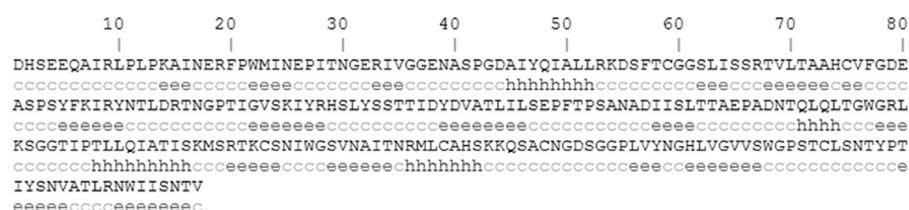


Fig. 6. Analysis of the secondary structure of Der f 9 allergen (GOR4 software) H, Alpha helix; E, extended strand; C, random coil.

TABLE 1—Protein Subcellular Localization for Der f 9 (CELLO2.5)

Support vector machine	Localization	Reliability
Amino acid comp.	Plasma membrane	0.855
N-peptide comp.	Extracellular	0.697
Partitioned seq. comp.	Plasma membrane	0.917
Physico-chemical comp.	Extracellular	0.843
Neighboring seq. comp.	Plasma membrane	0.592
CELLO prediction:	Plasma membrane	2.455
	Extracellular	1.781
	Nuclear	0.301
	Mitochondrial	0.187
	Chloroplast	0.108
	Cytoplasmic	0.055
	Lysosomal	0.045
	Vacuole	0.029
	Peroxisomal	0.014
	Golgi	0.011
	ER	0.008
	Cytoskeletal	0.006

immunological similarities between recombinant allergens and natural allergens have provided the opportunity to use these proteins instead of natural allergen extracts in the clinic for the diagnosis and therapy of IgE-mediated allergic disease. Here, we report the full-length cDNA fragment encoding the group 9 allergen from *D. farinae* isolated from house dust. The reference sequence, GenBank Accession No. AY283282, was obtained from expressed sequence tags concluded from the GenBank record. Thus, there are no reported data describing its molecular weight, biochemical name, allergenicity, or other immunological characterization. Thus, we cannot be certain whether reference sequence AY283282 is

TABLE 2—Functional Sites or Motifs of the Mature Der f 9 Allergen

Prediction tool	Functional sites or motifs	Amino acid position
InterPro 5.0	Peptidase S1A, chymotrypsin-type	60–75
	Trypsin-like cysteine/serine peptidase domain	29–256
	Peptidase S1	34–252
		33–252
		34–257
ScanProsite	Peptidase S1, trypsin family, active site	206–217
		70–75
	Serine proteases, trypsin domain profile (TRYPSIN_DOM)	34–257
	Serine proteases, trypsin family, histidine active site (TRYPSIN_HIS)	70–75
	Serine proteases, trypsin family, serine active site (TRYPSIN_SER)	206–217

accurate. Further, this allergen has not been included in the official website of the systematic allergen nomenclature that is approved by the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-committee (<http://www.allergen.org/>).

The sequence for Der f 9 that we obtained following RT-PCR and RACE was 99.85% identical with the reference sequence (AY283282); however, the two sequences exhibit differences in the 5'-end. A later search identified a second sequence, GenBank No. KM010013, which is consistent with our reported sequence. After being translated to proteins, searched with BLAST tools and aligned, our sequence has similar sequence with Der p 9 (AAP57077) in the N-term part. Therefore, our sequence is likely to be accurate.

After removal of signal peptide sequence, the gene encoding the mature allergen was then sub-cloned into plasmid pET-28b (+) and transformed into *E. coli* BL21 (DE3) for expression. Our recombinant protein demonstrated good IgE-binding activity and immunogenicity by ELISA and dot blotting. The recombinant Der f 9 exhibited IgE reactivity in nearly 57% of mite-allergic pediatric patients with asthma. This is an initial evaluation of its IgE-binding ability as an allergen in mite-allergic individuals with asthma in China. One limitation of our study is that we did not conduct inhibition studies to confirm that the cloned version is similar to the native form because we have not the native Der f 9 allergen. The present cloning and expression work provides a foundation for the large-scale production of recombinant allergens to diagnose and treat one of the most common causes of allergic symptoms.

With the increasing allergy-related data through functional genomics, gene microarrays, and proteomics, bioinformatics methods are becoming indispensable for the management of allergen data and detailed analysis of their sequences, structures, functional properties, and allergenicity. After sequencing the plasmid pCold TF-Der f 9, a 19-residue signal peptide and an *Nco* I site in the full-length Der f 9 were predicted by bioinformatics. A signal peptide is generally required for the secretion of a target protein through a cell membrane and into the medium; for recombinant protein expression in bacteria, the presence of a signal peptide does not ensure efficient protein secretion.¹⁸ The pET expression vector is one of the most widely used systems for the cloning and in vivo expression of proteins in *E. coli* due to the strong selectivity of the bacteriophage T7 RNA polymerase for its cognate promoter sequences, the high level of activity of the polymerase, and the high efficiency of translation mediated by the T7 gene 10 translation initiation signals. The pET expression vectors contain an *Nco* I sequence at the 5'-end of the multiple cloning site which affects the full-length Der f 9 subcloned into pET-28b(+) via *Nco* I

```

Der_p_9  ----EQAIRLPLPDAITEKFPWMINEPLNDERE----RIVGGSNASPGDAVYQIALFRKD
Der_f_9  DHSEEQAIRLPLPKA INERFPWMINEPITN-GE----RIVGGENASPGDAIYQIALLRKD
Sui_m_9  -----DLRLPLPEVIQREYPWLIDEPINSDNDAYSARIVGGSNVGPGEAKHQIALLRSG
Blo_t_9  -----SELRLPLPGLIRTLHPHLENEMINTNAP--FVRVVGNDAGYGDAPYQIAMIRSG
Ale_o_9  -----HLRLPLPKDIQARHPHLMTEPINTDAD--FVKVVGSAVSDGEARHQIAMLRSG
          :***** * . * : * : .          : :***. . . * : * :***: :*.

Der_p_9  SFTCGGSLISSRTVLTAAHCVFGDEATPSYFKIRYNTLDRTNGPPIGVSKIYRHNLYSSS
Der_f_9  SFTCGGSLISSRTVLTAAHCVFGDEASPSYFKIRYNTLDRTNGPTIGVSKIYRHSLYSST
Sui_m_9  SFTCGGSLISSRTVLTAAHCVYGNENKPTTFTIRYNTLDRTSGPTIGVKKVNRHASYSNN
Blo_t_9  SFICGGSLIGSQTVLTAAHCVHGYENNPSVFNVRYGTNDRTRGPTVAVKKVNRHPQYDSY
Ale_o_9  SFICGGSLIGASTVLTAAHCVYGYENRPTSFSVRYNTLDRTSGPTLAVAKVNRHPSYSNN
          ** ***** : ***** * * * : * : ** * *** ** : . * : ** * *

Der_p_9  PIDYDVATLILSQPFTPSANADIIPLTSEPADGTLKQITGWGRKSGGTLPTILQIAS-
Der_f_9  TIDYDVATLILSEPFTPSANADIISLTAEPADNTQLQLTGWGRKSGGTIPTLLQIAT-
Sui_m_9  TIDYDVATLTLSSAFAPGDNAAVIGLASTEPAANEVLKVTGWGRLSAGGSLPTKLQGAN-
Blo_t_9  TIDYDVATFVLSEPFTPSANAIVIPLATSVPADGSSLITGWGRLSGGTLPTKLQKATE
Ale_o_9  TIDYDIATFSLASPFSPGTNAAVVELAKSRPADDSAVQVTGWGRLSGGTLPTKLQKADT
          *****: *:. *. *. ** : : *:. : ** : :*****. :*: :** * *

Der_p_9  VTKMSRTKCSSTWGSVNAITNRMLCAHNSNQASCNGDSGGPLVSNHGLVGVSWSGPSTCL
Der_f_9  ISKMSRTKCSNIWGSVNAITNRMLCAHKKQSACNGDSGGPLVYNGHLVGVSWSGPSTCL
Sui_m_9  LNTISRSDCQKRWGSTNTITARMICAHSTTQSACNGDSGGPITKNGLLVGVWSGSSSCL
Blo_t_9  LKVVPKTTCAQAWSSVNSITNRMLCAHKKQSACNGDSGGPLIQDNVQVGVSWSGSSSCL
Ale_o_9  LKVVSKAECQKRWGSTNTITDRMLCAHSTTQSACNGDSGGPLTQNNVLVGVSWSGSSSCL
          :. : : : * . *. * : ** * : ** : :*****: : ***** * : **

Der_p_9  STKYPTIYSNVANLRNWIISNTV
Der_f_9  SNTYPTIYSNVATLRNWIISNTV
Sui_m_9  HQTYPNVYANVANLKNWINANTV
Blo_t_9  HATYPNVYANVANLRSWIQANS-
Ale_o_9  HATYPNVYASVADLADWIAANP-
          . ** : * : . ** * . ** : *

```

Fig. 7. Alignment between Der f 9 and homologous amino acid sequences by ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/>). Notes: “*” means that the residues or nucleotides in that column are identical in all sequences in the alignment; “:” indicates conserved substitutions; “.” indicates semi-conserved substitutions. Group 9 allergens of *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Suidasia medanensis*, *Aleuroglyphus ovatus*, and *Blomia tropicalis* are abbreviated as Der p, Der f, Sui m, Ale o, and Blo t, respectively.

TABLE 3—Amino Acid Similarity Between Der f 9 and Related Proteins in Other Species: Der p 9, Blo t 9, Sui m 9, and Ale o 9 [VECTOR NTI (v.11.5.1)]

	Ale o 9	Blo t 9	Der f 9	Der p 9	Sui m 9
Ale o 9		70	58	55	71
Blo t 9	81		62	60	62
Der f 9	74	75		85	64
Der p 9	74	74	93		62
Sui m 9	87	79	79	78	

Similarity is above the diagonal line, and divergence is below. Group 9 allergens of *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Suidasia medanensis*, *Aleuroglyphus ovatus*, and *Blomia tropicalis* are abbreviated as Der p, Der f, Sui m, Ale o, and Blo t, respectively.

and *Xho* I sites. Therefore, we deleted the signal peptide sequence and mutated the *Nco* I site in Der f 9 cDNA by PCR amplification to ensure better quality of recombinant Der f 9. In fact, SDS-PAGE identified a specific band at ~28 kDa after induction of the *E. coli* BL21 (DE3) containing the plasmid pET-28b-Der f 9, which is consistent with the estimated molecular weight of 27.8 kDa by bioinformatics software and validates our DNA manipulation RT-PCR and RACE. Further, mass spectrometry analyses of the structure by MALDI-TOF/TOF confirmed the recombinant fusion protein was consistent with the Der f 9 allergen.

After removal of the signal peptide sequence, the mature Der f 9 should be stable and hydrophilic due to its predicted instability index (II) of 35.64 and its grand average of hydropathicity (GRAVY) of -0.147. These

physicochemical properties, including of MW, PI, stability, and hydrophilicity, support the previous reports on Der p 9 allergen as a collagenolytic serine protease with an apparent MW of ~28 kD isolated from feces-enriched extracts of the dust mite *D. pteronyssinus* by ion-exchange chromatography and affinity chromatography.¹⁹ Besides these properties, we computed the average flexibility, relative mutability, the polarity, and accessible residues by ProtScale tools. Structural flexibility is essential for protein activity, and flexible regions are common in catalytic sites, binding sites, antigenic regions, sites susceptible for proteolytic cleavage, allosteric hinge, etc.²⁰ The average flexibility was calculated with a maximum score at residue 40 for Der f 9, which was predicted as the active site for serine proteases by InterPro 5.0 and ScanProsite. Relative mutability is the probability that the given amino acid will change in the evolutionary period of interest. The relative mutability of Der f 9 was calculated with a maximum score at residue 136. The residue accessibility is the sum of the accessibilities of the atoms, so active sites are often located on the surface of the protein. The accessible residues of Der f 9 were predicted with max score at residue of 206, a serine site. These bioinformatics characterizations will be helpful for understanding structure and function relationships.

Based on the deduced amino acid sequence with 257 residues, the secondary structure comprises alpha helix (10.89%), extended strand (32.30%), and random coils (56.81%). For most proteins, alpha helices and extended beta sheets are the most common secondary structures, whereas the random coil is not a true specific shape but a statistical distribution of shapes for all the chains, where the monomer subunits are oriented randomly while still being bonded to adjacent units. In biochemistry and structural biology, secondary structure is the general three-dimensional form of local segments of proteins, which is defined by patterns of hydrogen bonds between backbone amide and carboxyl groups. That 56.81% of residues were involved in random coil may be associated with some particular feature of the Der f 9 allergen, which should be explored in future work.

Functional sites or motifs analysis regarded Der f 9 as a chymotrypsin-type protein family member with trypsin-like cysteine/serine peptidase domain. BLAST search and alignment identified proteins similar to group 9 allergens from other mites. Perhaps unsurprisingly, the greatest similarity (85%) was found between Der f 9 and Der p 9. Der p 9 was described previously as similar to chymotrypsins and cathepsin G-like enzymes from both vertebrate and invertebrate sources as judged by cleavage of substrates containing either C-terminal phenylalanine or leucine residues.¹⁸ Like Der p 9, these group 9 allergens

may be involved in digestion of collagenous substrates in the form of human skin scales.¹⁸

In conclusion, we obtained the full-length Der f 9 cDNA and expressed it in *E. coli* successfully for the first time. Subsequent bioinformatics analyses provide a foundation to advance our understanding of this allergen, but much work remains to be done to improve the utility of this knowledge in the clinic.

ACKNOWLEDGMENTS

This work was supported by National Natural Sciences Foundation of China (NSFC 30060166, NSFC81001330, NSFC31272369, NSFC31572319). We thank Dr. Jiayong Li, Department of Laboratory, the first affiliated hospital of Shanghai Jiao Tong University, Shanghai, P. R. China, who provided sera for specific-IgE antibody detection.

REFERENCES

1. Eifan AO, Calderon MA, Durham SR. Allergen immunotherapy for house dust mite: clinical efficacy and immunological mechanisms in allergic rhinitis and asthma. *Expert Opin Biol Ther* 2013;13:1543–1556.
2. Walgraffe D, Matteotti C, Bakkoury M, Garcia L, Marchand C, Bullens D, Vandenbranden M, Jacquet A. A hypoallergenic variant of Der p 1 as a candidate for mite allergy vaccines. *J Allergy Clin Immunol* 2009;123:1150–1156.
3. Bordas-Le Floch V, Bussi eres L, Airouche S, Lautrette A, Bouley J, Berjont N, Horiot S, Huet A, Jain K, Lemoine P, et al. Expression and characterization of natural-like recombinant Der p 2 for sublingual immunotherapy. *Int Arch Allergy Immunol* 2012;158:157–167.
4. Vrtala S, Huber H, Thomas WR. Recombinant house dust mite allergens. *Methods* 2014;66:67–74.
5. Valenta R, Niespodziana K, Focke-Tejkl M, Marth K, Huber H, Neubauer A, Niederberger V. Recombinant allergens: what does the future hold? *J Allergy Clin Immunol* 2011;127:860–864.
6. Valenta R, Linhart B, Swoboda I, Niederberger V. Recombinant allergen for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. *Allergy* 2011;66:775–783.
7. Li J, Sun B, Huang Y, Lin X, Zhao D, Tan G, Wu J, Zhao H, Cao L, Zhong N. China alliance of research on respiratory allergic disease. A multicentre study assessing the prevalence of sensitization in patients with asthma and/or rhinitis in China. *Allergy* 2009;64:1083–1092.
8. Cui YB, Zhou P, Peng JL, Peng M, Zhou Y, Lin YZ, Liu L. Cloning, sequence analysis, and expression of cDNA coding for the major house dust mite allergen, Der f 1, in *Escherichia coli*. *Braz J Med Biol Res* 2008;41:380–388.
9. Cui YB, Zhou Y, Shi W, Ma G, Yang L, Wang Y. Cloning, expression, and analysis of the group 2 allergen from *Dermatophagoides farinae* from China. *An Acad Bras Ci nc* 2010;82:1–11.
10. Cui YB, Cai HX, Li L, Zhou Y, Gao CX, Shi WH, Yu M. Cloning, sequence analysis and expression in *E. Coli* of the group 3 allergen of *Dermatophagoides farinae*. *Chin Med J (Engl)* 2009;122:2657–2661.
11. Cui YB, Zhou Y, Ma GF, Li Yang, Wang Y. Molecular cloning, bioinformatic analysis, and expression of a house dust mite

- allergen Der f 5 of *Dermatophagoides farinae* from China. *Braz J Med Biol Res* 2012;45:746–752.
12. Cui YB, Zhou Y, Shi WH, Ma GF, Yang L, Wang Y, Li L. Molecular cloning, expression, sequence analyses of dust mite allergen Der f 6 and its ige-binding reactivity with mite allergic asthma patients in southeast China. *Mol Biol Rep* 2012;39: 961–968.
 13. Cui YB, Cai HX, Zhou Y, Gao CX, Shi WH, Yu M, LIL. Cloning, expression and characterization of Der f 7, an allergen of *Dermatophagoides farinae* from China. *J Med Entomol* 2010;47: 868–876.
 14. Cui YB, Gao CX, Zhou Y, Peng M, Lin Y, Peng J. Phylogenetic analysis of house dust mites. *Cent Eur J Med* 2010;5:69–74.
 15. Cui YB, Zhou Y, Wang YG, Ma GF, Yang L. The group 10 allergen of *Dermatophagoides farinae* (Acari: pyroglyphidae): cDNA cloning, sequence analysis, and expression in *E. coli* BL21 (DE3). *J Med Entomol* 2013;50:205–208.
 16. Yubao Cui, Lili Yu, Feixiang Teng, Nan Wang, Ying Zhou, Li Yang, Chengbo Zhang. Dust mite allergen Der f 4: expression, characterization, and IgE binding in pediatric asthma. *Pediatr Allergy Immunol* 2016;27:391–397.
 17. Nicklas RA. National and international guidelines for the diagnosis and treatment of asthma. *Curr Opin Pulm Med* 1997;3:51–55.
 18. Li YL, Huang LM, Lin WP, Tsai CC, Lin TS, Hu YH, Chen HS, Han JM, Wang HJ, Liu YT. Secretion of biologically active human epidermal growth factor from *Escherichia coli* using *Yersinia pestis* Caf1 signal peptide. *J Microbiol Immunol Infect* 2006;39: 366–371.
 19. King C, Simpson RJ, Moritz RL, Reed GE, Thompson PJ, Stewart GA. The isolation and characterization of a novel collagenolytic serine protease allergen (Der p 9) from the dust mite. *Dermatophagoides pteronyssinus*. *J Allergy Clin Immunol* 1996;98: 739–747.
 20. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins* 1994;19:141–149.