Am J Transl Res 2014;6(6):786-792
www.ajtr.org /ISSN:1943-8141/AJTR0002583

Original Article

Cloning, expression, and analysis of a cDNA coding for the *Dermatophagoides farinae* group 21 (Der f 21) allergen

Yubao Cui[1,2], Yongqian Jiang[3], Youlin Ji[3], Ying Zhou[2], Lili Yu[2], Nan Wang[2], Li Yang[2], Chengbo Zhang[2]

[1]Department of Laboratory Medicine, Shanghai Punan Hospital, Pudong New District, Shanghai 200125, P. R. China; [2]Department of Laboratory Medicine, Yancheng Health Vocational & Technical College, Jiangsu Yancheng 224006, P. R. China; [3]Department of Respiratory, The First People's Hospital of Yancheng City, Jiangsu Yancheng 224005, P. R. China

Received September 17, 2014; Accepted October 20, 2014; Epub November 22, 2014; Published November 30, 2014

Abstract: Domestic mite species like *Dermatophagoides farinae* induce allergies in people worldwide. Here, the cDNA coding for group 21 allergen of *Dermatophagoides farinae* (Hughes; Acari: Pyroglyphidae) from China was cloned, sequenced, and expressed in *E. coli* to aid in the development of diagnostic and treatment options for domestic mite hypersensitivity. First, the Der f 21 cDNA fragment was synthesized by RT-PCR; the confirmed full-length sequence comprised 411 nucleotides. The cDNA was ligated to the vector pCold-TF to construct an expression plasmid, pCold-TF-Der f 21. pCold-Tf-Der f 21 was transformed into *E. coli* BL21 cells, and its expression was induced by IPTG treatment. SDS-PAGE showed a specific band at the predicted molecular weight of Der f 21, demonstrating its successful expression. The recombinant fusion protein was obtained and its structure and molecular weight were confirmed by MALDI-TOF/TOF. Bioinformatics analysis revealed that the protein contained a signal peptide of 17 amino acids. The molecular weight of the mature Der f 21 allergen was approximately 14.16 kDa with a theoretical pI of 4.87. Its predicted secondary structure comprises a-helix (84.03%), extension chain (1.68%), and random coil (14.29%). The successful cloning of Der f 21 and a basic bioinformatics analysis of the protein provide a foundation for further study of this allergen in diagnosis and treatment of domestic mite hypersensitivity.

Keywords: *Dermatophagoides farinae*, recombinant allergen, molecular cloning, domestic mite, Der f 21

Introduction

Domestic mites, like those found in house dust, induce allergic diseases such as bronchial asthma, rhinitis, and atopic dermatitis in many people worldwide [1, 2]. Aqueous extracts of various mite species, which can be generated from eggs, mite bodies, culture media, and whole mites, have aided the diagnosis and treatment of these IgE-mediated allergic diseases. Most extracts comprise complex mixtures of proteins, only some of which exhibit allergenic characteristics. To date, at least 24 groups of house dust mite allergens have been identified (http://www.allergen.org/). Some individuals become sensitized to only one or two domestic mite allergens; others respond to a spectrum of them [3-7]. These diverse patterns of allergen recognition produce some limitations in both the safety and efficacy of mite extracts in allergen-specific immunotherapy [3-7]. The use of recombinant allergens that recapitulate their natural counterparts and can be produced as defined molecules in consistent quality and unlimited amounts is moving the field toward more effective treatments with fewer adverse events [8-13].

Domestic mite hypersensitivity, particularly to *Dermatophagoides farinae*, is common in China [14]. To better understand domestic mite hypersensitivity, our laboratory has cloned and expressed several allergens from *D. farinae* [15-22]. The current study sought to clone, for the first time, the full-length group 21 allergen from *D. farinae* (Der f 21). These results will provide a foundation for improved diagnosis and treatment of domestic mite allergy in China and worldwide.

Methods

Preparation of Der f 21 cDNA and polymerase chain reaction (PCR)

D. farinae were cultured and isolated as we reported previously [15-22]. 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 sequence of Der f 21 (GenBank Accession No. KF732965). The sequences of the forward and reverse primers were as follows: (F) 5'-TACCCTCGAGGGATCC ATGAAATTCATTATTTTCTGTGCC-3' and (R) 5'-TAGACTGCAGGTCGACTTA ATCATCCGATTTTAC-AGCTTT-3' with a *Bam* HI site and a *Sal* I site at their 5' end (underlined), respectively. Reverse transcription (RT) was performed with TaKaRa PrimeScript™ RT-PCR Kit (Code No. DR014A, TaKaRa) using total RNA isolated from mites in the PCR Thermal Cycler Dice (TaKaRa, TP600). Second, the RT product was used as the template for PCR in the same thermal cycler with PrimeSTAR® HS DNA Polymerase (TaKaRa, DR-044A). Finally, 5 mL of the PCR product were analyzed by agarose electrophoresis (1.0%) and visualized with ImageMaster® VDS (Pharmacia Biotech).

Cloning and DNA sequencing

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). *Escherischia coli* JM109 (TaKaRa Code No.D9052) cells were transformed with the recombinant plasmid pCold-TF-Der f 21, 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'-GCGGGTCTGGAA-GTTCTG TT-3') and VP2 (5'-CCAAATGGCAGGGA-TCTTAG-3').

Construction of expression plasmids of pCold TF-Der f 21

After sequencing, the verified pCold TF-Der f 21 plasmids were transformed into *E. coli* BL21 (DE3, Stratagene, La Jolla CA, USA). The *E. coli* BL21 carrying pCold TF-Der f 21 was grown on LB plates containing 100 μg/mL ampicillin at 37°C overnight. A single colony was inoculated into 2 mL LB and cultured at 37°C overnight; 10 mL of culture were added into a glass tube containing 2 mL LB with ampicillin and then cultured at 37°C. 100 mM Isopropyl-b-D- thiogalactopyranoside (IPTG, 50 mL, final concentration of 1 mM) was added, and the sample was incubated for 23 h at 15°C. The *E. coli* were harvested by centrifugation. After re-suspension in PBS buffer (160 mL/tube), cells were subjected to ultrasonic disruption until the suspension became transparent. 8 mL of the suspension were collected, 2 mL of 5 × SDS sample buffer were added, and the mixture was heated for 10 min at 95°C. In parallel, an empty pCold-TF plasmid was used as a control. The recombinant protein was expressed and isolated by sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE, 10%) and visualized by CBB-R250 staining.

Production and purification of recombinant fusion protein

The *E. coli* BL21 carrying pCold TF-Der f 21 was grown on 400 mL LB plates containing 100 μg/mL ampicillin, and IPTG was added to a final concentration of 100 mM. After culturing at 28°C, *E. coli* colonies were harvested by centrifugation (8000g ×), the precipitate was added with lysis buffer (10 mM imidazole, 50 mM PBS, 0.3 M NaCl, 5% glycerol) and disrupted by sonication. After the induced bacteria were collected and centrifuged, the supernatant was collected for purification. Before purification, lysis buffer (10 mM imidazole, 50 mM PBS, 0.3 M NaCl, 5% glycerol) was used to balance the His•Bind®column (Novagen company, Cat. No. 70971). The collected supernatant of the culture medium was added through the column at 1 mL/min, and the eluant was collected and stored at 4°C. The column was washed with washing buffer 1 (20 mM imidazole, 50 mM PBS, 0.3M NaCl, 5% glycerol) (5× volume), washing buffer 2 (50 mM imidazole, 50 mM PBS, 0.3M NaCl, 5% glycerol) (5× volume), and eluting buffer (250 mM imidazole, 50 mM PBS, 0.3M NaCl, 5% glycerol) (2× volume). Then 10 mL of the purification product were subjected to separation by SDS-PAGE.

Matrix-assisted laser desorption ionization (MALDI) mass spectrometry analysis of the recombinant fusion protein

The purified recombinant fusion protein was analyzed by MALDI-TOF/TOF mass spectrome-

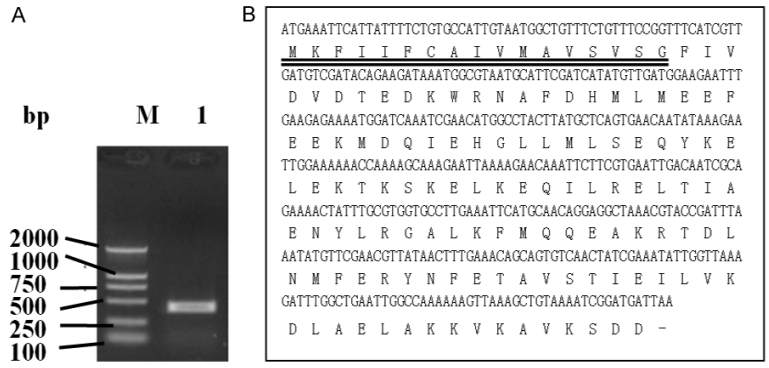

Figure 1. Cloning and sequencing of cDNA encoding Der f 21. A. Amplification of the cDNA of the gene encoding Der f 21 protein by RT-PCR. Total RNA was isolated from mites using RNA isolator and subjected to RT-PCR. The PCR products were separated on a 1% agarose gel containing ethidium bromide. *Lane M*, DNA Marker DL2, 000; *Lane 1*, PCR product (411 bp). B. The nucleotide sequencing results for Der f 21 and its deduced amino acid sequences (the signal peptide sequence is underlined).

ter (model 4800, Applied Biosystems). The spectra generated were mass-calibrated using known standards and the peaks de-isotoped. Databases were searched with the masses obtained using the MASCOT search engine (Swiss-Prot Database) and a 50 ppm mass tolerance window. Significant matches from the Peptide Mass Fingerprint data were confirmed by MS/MS analysis using the search criteria described and an MS/MS-tolerance window of 0.5 Da.

Bioinformatics analyses of the cloned Der f 21

The ORF was determined by using the ORF finder in the NCBI (National Center for Biotechnology Information) website. The amino acid sequence of Der f 21 was determined by Translate Tools in ExPaSy web server and its physio-chemical properties by ProtParam tool. After its signal peptide sequence predicted by SignalP 4.0 software was removed, its hydrophilicity, average flexibility, and relative mutability were determined by ProtScale tools; its nuclear export signal analysis by NetNES 1.1 server; phosphorylation site analysis by NetPhos2.0 server; and the secondary structure by GOR4.0.

Results

Obtaining the cDNA encoding Der f 21

Total RNA was isolated from adult mites, and Der f 21 cDNA fragments were amplified by RT-PCR. A band size indicating the expected 411 base pairs was detected following gel electrophoresis (**Figure 1A**). After the PCR-amplified DNA was recovered, the product was cloned into vector pCold-TF and transformed into *E. coli* competent cells (JM109). The product was confirmed by automatic DNA sequencing. Using the ORF Finder, a complete ORF was found within the Der f 21 cDNA; its length is 411 bp from the start codon ATG to the stop codon TAA (**Figure 1B**).

Expression and purification of recombinant protein Der f 21

After sequencing, the verified pCold-TF-Der f 21 plasmids were transformed into *E. coli* BL21, which were then grown on LB plates. Protein expression was induced by IPTG treatment. The optical density values were 0.61 and 1.28 before and after induction, respectively. A single, specific band matching the postulated molecular weight of 69 kDa was detected by SDS-PAGE (**Figure 2A** and **2B**). The molecular weight of the control plasmid, pCold-TF-DNA, was 55 kDa (**Figure 2A** and **2B**), therefore the molecular weight of Der f 21 was deduced to be 14 kDa. After chromatography, approximately 2.4 mg of the recombinant product was obtained, at a concentration of 2 mg/mL. The purified product was analyzed by MALDI-TOF/TOF. The peptide mass fingerprint (PMF) is shown in **Figure 2C** and **2D**, which was consistent with the structure of Der f 21. The molecular weight profiles by MALDI-TOF showed a broad peak near 68924.1 Da, which was estimated for the expression product of the plasmid pCold-TF-Der f 21 in *E. coli* BL21 cells (**Figure 2E**).

Inferred amino acid sequence and its structural and functional prediction

Bioinformatics tools were used to analyze the properties of Der f 21 based on the full-length nucleotide sequence. The amino acid sequence of Der f 21 was predicted using the Translate Tools on the ExPaSy web server; the protein sequence comprises 136 residues (**Figure 1B**).

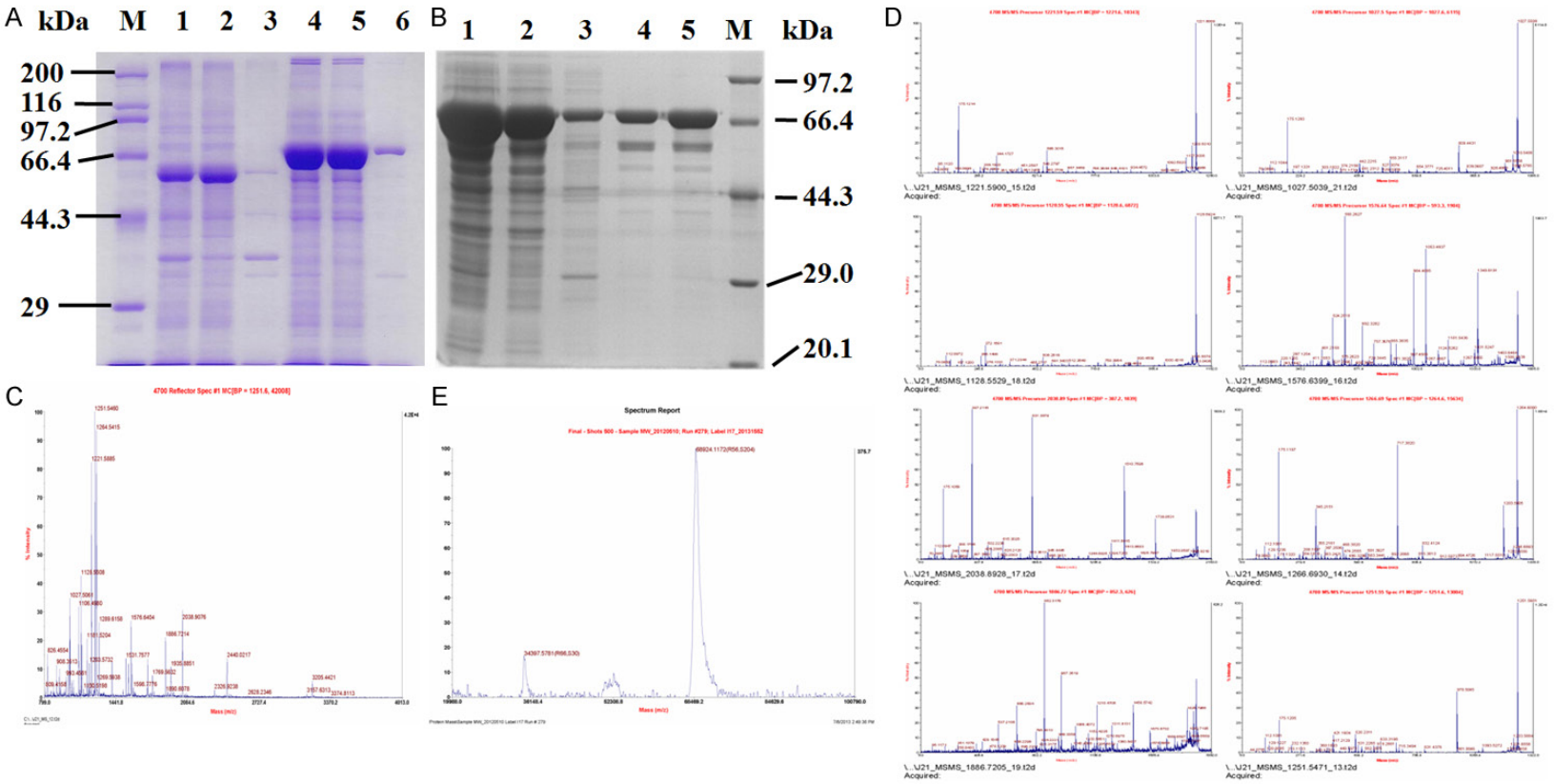

Figure 2. Expression and purification of recombinant protein Der f 21. A. SDS-PAGE analysis of the protein expressed from the pCold-TF- Der f 21 recombinant plasmid in *E. coli* BL21 cells. *Lane M*, TaKaRa Protein Marker (Broad); *Lane 1*, the whole cell lysate of *E. coli* BL21 cells containing pCold-TF; *Lane 2*, the supernatant of cells containing pCold-TF; *Lane 3*, the pellet of cells containing pCold-TF; *Lane 4*, the whole-cell lysate of *E. coli* BL21 cells containing pCold-TF-Der f 21; *Lane 5*, the supernatant of cells containing pCold-TF-Der f 21; *Lane 6*, the pellet of cells containing pCold-TF-Der f 21. B. SDS-PAGE analysis of purified recombinant protein from *E. coli* BL21 cells. *Lane 1*, the supernatant of cell containing pCold-TF-Der f 21; Lane 2, the flow through the column; *Lanes 3, 4, and 5,* Eluted fractions with 20, 50, and 250 mmol/L imidazole elution buffer, respectively. *Lane M*: TaKaRa Protein Marker (Broad). C. The first-order structure of the purified recombinant fusion protein from *E. coli* BL21 cells by MALDI-TOF-MS. D. The second-order structure of the purified recombinant fusion protein from *E. coli* BL21 cells by MALDI-TOF-MS. E. Determination of molecular weight of the purified recombinant fusion protein from *E. coli* BL21 cells by MALDI-TOF-MS.

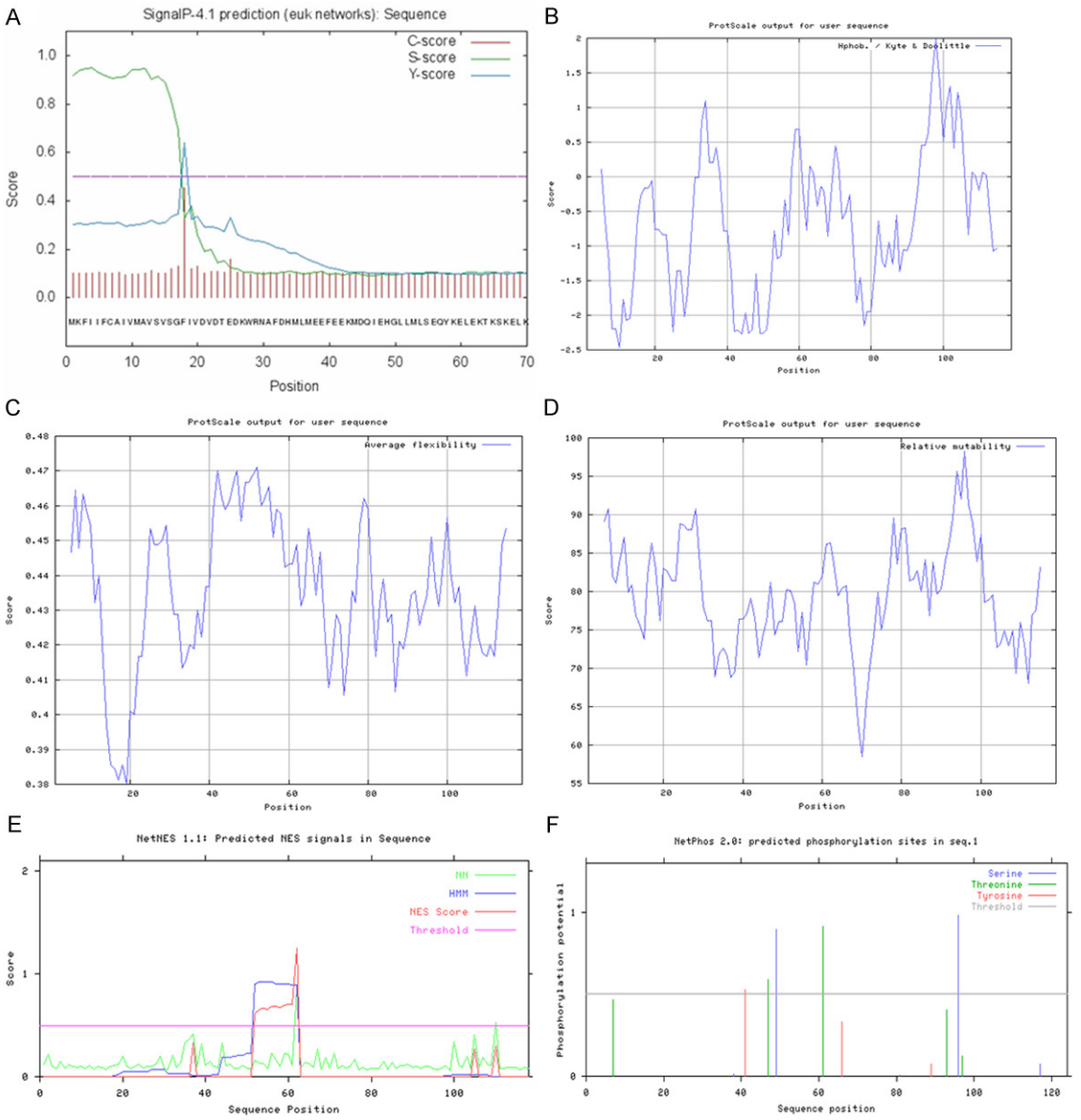

Figure 3. Bioinformatics analyses using the mature amino acid sequence of Der f 21. X-axis represents protein length from *N*- to *C*-termini. Y-axis represents the score. The higher score, the higher probability. A. SignalP-NN prediction (eukaryotic networks) by SignalP 3.0 software in SMART Server (http://www.cbs.dtu.dk/services/SignalP/) B. Hydrophilicity analysis (ProtScale software, Kyte & Doolittle). C. The average flexibility (ProScale software). D. The relative mutability (ProScale software). E. Nuclear export signal analysis by NetNES 1.1 Server. F. Phosphorylation sites analysis by NetPhos2.0 Server.

A signal peptide sequence from amino acid residues 1-17 was predicted by SignalP 4.0 software (**Figure 3A**). After removal of this sequence, the predicted mature Der f 21 comprises 119 amino acid residues and has a molecular weight of 14.16 kDa. Its isoelectric point is 4.87, and its instability index (II) is 35.58, suggesting a stable protein molecule. The grand average of hydropathicity (GRAVY) was postulated to be -0.626, suggesting that it is a hydrophilic protein, as demonstrated by ProtScale tool (**Figure 3B**). Additional sequence and structural analysis provided more insight to this protein. **Figure 3C** shows the average flexibility indexes for each amino acid residue. The lowest was 0.380, positioned at residue 19; the highest was 0.471 at residue 52, which is leucine. **Figure 3D** shows the relative mutability for each amino acid residue. The lowest was 58.4 positioned at residue 70, whereas the

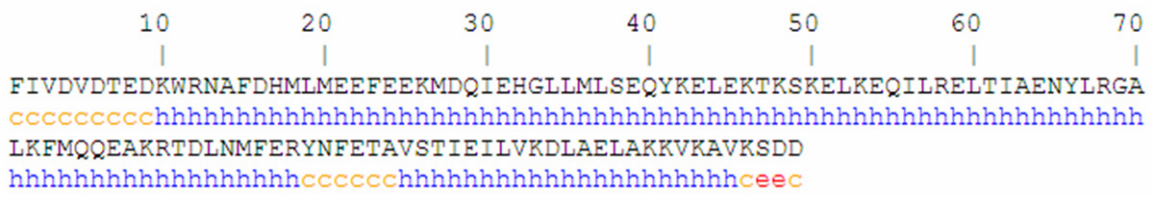

Figure 4. The secondary structure analysis of the mature allergen Der f 21 by GOR 4 software.

highest was 98.2 at residue 96. The higher the relative mutability, the more probable is the mutation of an amino acid. The relative mutability for Der f 21 was scored the highest at residue 96, which is glutamic acid. NetNES 1.1 was used to analyze the Der f 21 amino acid sequences. The N-terminus, from residues 52 to 62, contains a nuclear export signal (**Figure 3E**). A nuclear export signal (NES) is a short amino acid sequence with 4 hydrophobic residues that targets a protein for export from the cell nucleus to the cytoplasm through the nuclear pore complex using nuclear transport. The deduced NES for Der f 21 was LKEQILRELTI, of which leucine (L) and isoleucine (I) are hydrophobic residues. NetPhos2.0 Server was used to analyze the Der f 21 amino acid sequence. Five phosphorylation sites were inferred (**Figure 3F**), including two serine sites positioned at residues 49 and 96, two threonine sites at residues 47 and 61, and one tyrosine site at residue 41. GOR4 structural analysis predicted a secondary structure with an a-helix (100 peptides, 84.03%), a main extension chain (2 peptides, 1.68%), and a random coil (17 peptides, 14.29%; **Figure 4**).

Discussion

Since recombinant allergens offer the potential to overcome the shortcomings of treatments dependent on native allergens, many allergens have been cloned, sequenced, and expressed. Here, we amplified the cDNA fragment encoding the group 21 allergen from *D. farinae* isolated from house dust in China by RT-PCR, and inserted it into the pCold-TF vector to obtain the cloned plasmids. After nucleotide sequencing of the pCold-TF-Der f 21, the full-length sequence of Der f 21 was comprised of 411 base pairs. Transformed *E. coli* BL21 were induced to express the recombinant protein by treatment with IPTG. Protein expression was detected by SDS-PAGE, wherein a single specific band matching the postulated molecular weight was obtained. Further mass spectrometry analyses of the structure and molecular weight profiles by MALDI-TOF/TOF confirmed the recombinant fusion protein was consistent with the Der f 21 allergen. Thus, we successfully developed a fusion construct of pCold-TF-Der f 21 and expressed it in *E. coli*. This work was performed to develop a foundation for the large-scale production of recombinant allergens for diagnosis and treatment of the mite-allergic patients.

With the ever-increasing allergy-related data obtained through functional genomics, gene microarrays, and proteomics, bioinformatics tools are already crucial to allergy research, particularly the characterization of allergens by identification of structural motifs and B- and T-cell epitopes [23-25]. Our bioinformatics analysis of the sequence of Der f 21 revealed that the encoded 14-kDa protein contains stable and hydrophilic residues. The high average flexibility index is characterized by high average hydrophilicity, higher average absolute net charge, and high total charge, whereas the average flexibility index is significantly enriched in hydrophobic residues [26]. The allergen comprises predominantly alpha helix (84.03%), a common structure of proteins, and is a right-hand coiled or spiral conformation, as well as having extended strand (1.68%) and random coils (14.29%).

In sum, this work provides a foundation for future investigation of Der f 21 and the potential use of the recombinant Der f 21 for clinical application in allergy diagnosis and treatment.

Acknowledgements

This work was supported by National Sciences Foundation of China (NSFC 30060166, NSFC-81001330, NSFC31272369).

Disclosure of conflict of interest

None to declare.

Address correspondence to: Dr. Yubao Cui, Department of Laboratory Medicine, Shanghai Punan Hospital, Pudong New District, No. 519 Nanmatou Road, Shanghai 200125, China. Tel: 86-21-58892700; E-mail: ybcui1975@hotmail.com

References

[1] Thomas WR, Hales BJ, Smith WA. House dust mite allergens in asthma and allergy. Trends Mol Med 2010; 16: 321-328.

[2] Nadchatram M. House dust mites, our intimate associates. Trop Biomed 2005; 22: 23-37.

[3] Marth K, Focke-Tejkl M, Lupinek C, Valenta R, Niederberger V. Allergen peptides, recombinant allergens and hypoallergens for allergen-specific immunotherapy. Curr Treat Options Allergy 2014; 1: 91-106.

[4] Vrtala S, Huber H, Thomas WR. Recombinant house dust mite allergens. Methods 2014; 66: 67-74.

[5] Jutel M, Solarewicz-Madejek K, Smolinska S. Recombinant allergens: the present and the future. Hum Vaccin Immunother 2012; 8: 1534-1543.

[6] Valenta R, Niespodziana K, Focke-Tejkl M, Marth K, Huber H, Neubaruer A, Niederberger V. Recombinant allergens: what does the future hold? J Allergy Clin Immuno 2011; 127: 860-864.

[7] Focke-Tejkl M, Valenta R. Safety of engineered allergen-specific immunotherapy vaccines. Curr Opin Allergy Clin Immunol 2012; 12: 555-563.

[8] Schmid-Grendelmeier P, Crameri R. Recombinant allergens for skin testing. Int Arch Allergy Immunol 2001; 125: 96-111.

[9] Valenta R, Linhart B, Swoboda I, Niederberger V. Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. Allergy 2011; 66: 775-783.

[10] Valenta R, Niederberger V. Recombinant allergens for immunotherapy. J Allergy Clin Immunol 2007; 119: 826-830.

[11] Valenta R, Linhart B, Swoboda I, Niederberger V. Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. Allergy 2011; 66: 775-783.

[12] Cromwell O, Häfner D, Nandy A. Recombinant allergens for specific immunotherapy. J Allergy Clin Immunol 2011; 127: 865-872.

[13] Cromwell O, Suck R, Kahlert H, Nandy A, Weber B, Fiebig H. Transition of recombinant allergens from bench to clinical application. Methods 2004; 32: 300-312.

[14] Li J, Sun B, Huang Y, Lin X, Zhao D, Tang G, Wu J, Zhao H, Cao L, Zhong N, China Alliance of research on respiratory allergic disease. Allergy 2009; 64: 1083-1092.

[15] Cui YB, Zhou P, Peng JL, Peng M, Zhou Y, Lin Y, 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.

[16] Cui YB, Zhou Y, Shi WH, Ma GF, Yang L, Wang YG. Cloning, expression, and analysis of the group 2 allergen from *Dermatophagoides farinae* from China. An Acad Bras Cienc 2010; 82: 1-11.

[17] 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 Dermatophgoides farinae. Chin Med J (Engl) 2009; 122: 2657-2661.

[18] Cui YB, Zhou Y, Ma GF, Yang L, Wang YG. 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.

[19] Cui YB, Zhou Y, Shi WH, Ma GF, Yang L, Wang YG, 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.

[20] Cui YB, Cai HX, Zhou Y, Gao CX, Shi WH, Yu M, Li L. Cloning, expression and characterization of Der f 7, an allergen of Dermatiophagoides farinae from China. J Med Entomol 2010; 47: 868-876.

[21] Cui YB, Gao CX, Zhou Y, Peng M, Lin YZ, Peng JL. Phylogenetic analysis of house dust mites. Cent Eur J Med 2010; 5: 69-74.

[22] 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. J Med Entomol 2013; 50: 205-208.

[23] Brusic V, Petrovsky N. Bioinformatics for characterisation of allergens, allergenicity and allergic crossreactivity. Trends Immunol 2003; 24: 225-228.

[24] Brusic V, Petrovsky N, Gendel SM, Millot M, Gigonzac O, Stelman SJ. Computational tools for the study of allergens. Allergy 2003; 58: 1083-1092.

[25] Song J, L Li, Z Liu, Li Q, Ran P. Sequence analysis and expression of a cDNA clone encoding tropomysin in Sinonovacula constricta. Mol Biol Rep 2009; 36: 315-321.

[26] Radivojac P, Obradovic Z, Smith DK, Zhu G, Vucetic S, Brown CJ, Lawson JD, Dunker AK. Protein flexibility and intrinsic disorder. Protein Sci 2004; 13: 71-80.