

Original article

Cloning, sequence analysis and expression in *E. coli* of the group 3 allergen of *Dermatophagoides farinae*

CUI Yu-bao, CAI Hong-xing, LI Li, ZHOU Ying, GAO Cui-xiang, SHI Wei-hong and YU Ming

Keywords: *Dermatophagoides farinae*; *Der f 3*; recombinant allergen; cloning; gene expression; bioinformatics

Background The dust mites, which are mostly represented by *Dermatophagoides spp.* (Acari: Pyroglyphidae), are the major sources of indoor allergens. Identification and characterization of these mite allergen molecules are an important step in the development of new effective diagnostic procedures and possible therapeutic strategies for allergic disorders associated with dust mites.

Methods Total RNA was extracted from *Dermatophagoides farinae*. The gene coding for *Der f 3* was amplified by RT-PCR with the primers designed based on previous sequence published in GenBank. The target gene was cloned intermediately into pMD19-T plasmid and finally into plasmid pET28a (+), expressed in *E. coli* BL21 at the aid of the inducer isopropyl-D-thiogalactopyranoside (IPTG). The physicochemical properties, spatial structure of the allergen were analyzed with bioinformatics software.

Results The cDNA coding for group 3 allergen of *Dermatophagoides farinae* from China was cloned and expressed successfully. Sequencing analysis showed that there were nineteen mismatched nucleotides in five *Der f 3* cDNA clones in comparison with the reference (GenBank Accession No. AY283291), which resulted in deduced amino acid sequence incompatibility in eleven residues. Bioinformatics analysis revealed that the *Der f 3* pro-protein was an extracellular hydrophobic protein, consisting of 259 amino acids with a 16 amino acid signal peptide. The protein was deduced to have three chymotrypsin active sites (53-68 AA, 108-122 AA and 205-217 AA), one N-glycosylation site, one cAMP- and cGMP-dependent protein kinase phosphorylation site, four protein kinase C phosphorylation sites, two casein kinase II phosphorylation sites, and five N-myristoylation sites.

Conclusions *Der f 3* is an extracellular hydrophobic protein which possesses multiple activation and phosphorylation sites. Polymorphism may exist in the *Der f 3* gene but this needs to be further confirmed in the future.

Chin Med J 2009;122(21):2657-2661

Extracts from the body, egg, feces, and culture medium of dust mites have been demonstrated to result in allergic skin reactions in patients with allergic asthma, rhinitis, and atopic dermatitis.^{1,2} More than thirty components that are able to bind to IgE in the sera from asthma patients have been identified in dust mite extracts by immunochemical methods.³⁻⁵ Allergen immunotherapy, also known as allergy shots, is often used to treat patients by injecting increasing amounts of a diluted mite allergen over several months.⁶ At present, the dust mites extract used in immunotherapy of allergic patients is believed to be associated with some side effects in clinical settings due to its complicated composition.⁷⁻⁹ Genetic engineering techniques have been used to produce the known major allergens, *Der f 1* and *Der f 2*, which have been demonstrated to be more effective than their nature molecules.¹⁰⁻¹²

Polymorphisms have been described in several dust mite allergens in different regions, which could influence the diagnosis of the dust mite allergens as well as the effectiveness in the immunotherapy of the allergic disorders mediated by these allergens.¹³ Therefore, it is necessary to characterize dust mites allergens in China and establish a high expression system to produce these allergens on a large scale for practical diagnosis and

therapeutic intervention.

The group 3 allergen was originally isolated from faecally enriched extracts of *Dermatophagoides farinae* by gel-filtration, and subsequent reports showed that it reacted with 16% sera of mite-allergic patients.¹⁴ Since it is estimated that mite allergy affects approximately 10% of the world population,¹⁵ the group 3 allergen should be regarded as a non-negligible allergen. In the present study, we sought to clone and characterize the gene encoding the group 3 allergen of *Dermatophagoides farinae* (*Der f 3*) and investigate its expression in *E. coli*.

DOI: 10.3760/cma.j.issn.0366-6999.2009.21.025

Department of Pathogenic Biology, Yancheng Health Vocational & Technical College, Yancheng, Jiangsu 224006, China (Cui YB, Cai HX, Zhou Y, Gao CX, Shi WH and Yu M)

Department of Laboratory Medicine, No. 1 Hospital of Shanghai, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China (Li L)

CUI Yu-bao and CAI Hong-xing contributed equally to this paper, being co-first author.

Correspondence to: Prof. LI Li, Department of Laboratory Medicine, No. 1 Hospital of Shanghai, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China (Tel: 86-21-63240090. Fax: 86-21-63240825. Email: annylish@hotmail.com)

This work was supported by grants from National Science Foundation of China (No. 30060166 and 30671939).

METHODS

Synthesis of full-length *Der f 3* cDNA by reverse transcription-polymerase chain reaction (RT-PCR)

House dust mites were cultured and isolated as previously described.^{16,17} Total RNA was obtained using RNA isolator (Code No. D312, TaKaRa Biotechnology Limited Company, Dalian, China). Based on the published sequence of *Der f 3* (GenBank Accession No AY283291), a pair of primers were designed and the specific sequences are: 5'-GGATCCATGATGATT-TTAACCATTGTCGTG-3' (forward) and 5'-CTCG-AGTCACTGTGAACGTTTTGATTCAATC-3' (reverse) with a BamH I site and an Xho I site at their 5' ends (underlined), respectively. RT was performed using the total RNA isolated from mites with High Fidelity PrimeScript™ RT-PCR Kit (Code No. DR027A, TaKaRa) on a PCR Thermal Cycler Dice (Code TP600, TaKaRa). The reaction mixture for RT contained total RNA (1 µl), 20 µmol/L of reverse primer (0.5 µl), 10 mmol/L of dNTP mixture (0.5 µl), and RNase free H₂O (3 µl). The mixture was incubated at 65°C for five minutes, followed by an ice-bath for two minutes. Then, 5 PrimeScript RT buffer (2 µl), 40 U/µl of RNase inhibitor (0.25 µl), PrimeScript RTase (0.5 µl), and RNase free dH₂O (2.25 µl) were added. The final reaction mixture (10 µl) was set at 30°C water-bath for 10 minutes, 42°C for 30 minutes and 70°C for 15 minutes. The RT product was used as a template for PCR on the same thermal cycler (Dice) with PrimeSTAR® HS DNA polymerase (Code No. DR010A, TaKaRa). The total reaction mixture contained RT products (3 µl), 5×PrimeSTAR PCR buffer (10 µl), 2.5 mmol/L of dNTP mixture (4 µl), 20 µmol/L of forward primer (0.5 µl), 20 µmol/L of reverse primer (0.5 µl), 2.5 U/µl of PrimeSTAR HS DNA polymerase (0.25 µl), and dH₂O (31.5 µl). PCR conditions used here included an initial incubation for 2 minutes at 94°C, and followed by 30 cycles of 10 seconds at 98°C, 30 seconds at 55°C and 40 seconds at 72°C. After a final incubation for 5 minutes at 72°C, 5 µl of the amplicons were analyzed by agarose gel electrophoresis (1.0%) and visualized with ImageMaster® VDS.

Cloning and DNA sequencing

The PCR-amplified DNA was recovered from the gel with Agarose Gel DNA Purification Kit Ver. 2.0 (Code No. DV805, TaKaRa) and a poly-A tail was added with DNA A-Tailing Kit (Code No. D404, TaKaRa). The poly-A tailed product was then cloned into the simple vector pMD19-T (Code No. D104, TaKaRa). *E. coli* JM109 (Code No. D9052, TaKaRa) was then transformed with the recombinant plasmid pMD19-T-*Der f 3* and positive clones were selected by blue/white screening on Luria-Bertani (LB) plates containing 100 µg/ml ampicillin. The inserts were confirmed by restriction enzyme analysis with BamH I and Xho I and by automatic DNA sequencing.

Construction of expression plasmids

pET28a(+)-*Der f 3*

The recombinant pMD19-T-*Der f 3* plasmid was digested with BamH I and Xho I to release the *Der f 3* fragment. After recovering from the gel using Agarose Gel DNA Purification Kit Ver. 2.0 (Code No. DV805, TaKaRa), the *Der f 3* cDNA fragment was sub-cloned into the expression vector pET28a(+) (Kit Lot No. N72770, Novagen, Germany) to create pET28a(+)-*Der f 3*, using the DNA Ligation Kit (Code No. D6023, TaKaRa). Competent *E. coli* JM109 cells (Code No. D9052, TaKaRa) were transformed with pET28a(+)-*Der f 3* plasmids, positive clones were selected by blue/white screening and confirmed by restriction enzyme analysis with BamH I and Xho I.

Expression of r*Der f 3* in *E. coli* BL21 (DE3)

Purified pET28a (+)-*Der f 3* plasmid (0.5 µl) was used to transform 100 µl of *E. coli* BL21 (DE3, Stratagene, USA). The pET28a (+)-*Der f 3*-carrying *E. coli* BL21 were grown on LB plates containing 50 µg/ml of kanamycin at 37°C overnight. A single colony was inoculated into 2 ml LB plates at 37°C overnight, 100 µl of which was added into a glass tube containing 5 ml LB fluid containing kanamycin and then cultured at 37°C. One hundred mmol/L of isopropyl-β-D-thiogalactopyranoside (IPTG, 50 µl, final 1 mmol/L) was added to induce the tac promoter. The *E. coli* cells were harvested by centrifugation, and PBS buffer (200 µl/tube) was added in. After resuspension, ultrasonic disruption was used until a transparent fluid was obtained, 50 µl of which was taken as the whole cell lysate. The remaining fluid was centrifuged for the supernatant and pellet, which was regarded as the soluble and insoluble protein samples, respectively.

Western blotting

The whole cell lysate, soluble and insoluble protein samples were loaded (10 µl per well) onto a 12.5% sodium dodecylsulfate-polyacrylamide gel (SDS-PAGE). After electrophoresis, the proteins were transferred to a PVDF membrane (TIANGEN, Biotech Co. Ltd, Beijing, China) blot. The blot was then incubated with a His-Tag Mab (QIAGEN, Germany) primary antibody overnight at 4°C, followed by incubation with the horseradish peroxidase (HRP)-rabbit anti-mouse IgG (H+L) (ZYMED Laboratories, USA) secondary antibody for one hour at room temperature. The proteins were then visualized using 3 ml of BCIP/NBT (Roche, Switzerland).

Nucleotide sequence, inferred amino acid sequence, and bioinformatics

The sequences were edited to remove the vector sequence and the extra restriction sites. The open reading frame (ORF) was obtained using the ORF finder at the NCBI (National Center for Biotechnology Information) Website. The amino acid sequence of *Der f 3* was deduced using the Translate Tools in the ExPaSy web server; its

physico-chemical properties were analyzed by ProtParam Tools, its signal peptide sequence by SignalP 3.0 software, hydrophilicity by ProtScale tools, the secondary structure by GOR4.0, the protein subcellular localization by CELLO v.2.5, and the functional site by InterProScan and PROSCAN software.

RESULTS

Cloning and sequencing of *Der f* 3 cDNA

Eight clones were positive based on the restriction enzyme mapping with BamH I and Xho I, five of which were subject to sequencing with primers pMD18F and pMD18R. The sequences of these five clones were compared for their homology with those published in GenBank by some other authors (access numbers D63858, EU312162, EU312163 and AY283291). As shown in Figure 1, a total of nineteen variant positions were found among these nine sequences. When these nucleotide sequences were translated into amino acid sequences using the Translate Tools in the ExPaSy web server, there were eleven incompatible amino acid (AA) positions (Figure 2). Some of these positions might represent substitution sequences.

Construction of plasmid pET28a (+)-*Der f* 3 and its expression in *E. coli* BL21

Among the five sequenced recombinant pMD19-T-*Der f* 3 clones in our study, clone number 5 had the highest homology (99.99%) with the reference sequence (accession No. AY283291). There was only one base pair difference, at position 489 (C→A). When this clone was subsequently sub-cloned into the expression vector pET28a(+) and transformed into *E. coli* BL21, a single specific band around 32 kD was detected by SDS-PAGE Western blotting (Figure 3 A and B).

Der f 3 inferred amino acid sequence and its structural and functional prediction

There was one open reading frame in clone 5 with the length of 780 bp from the start codon ATG to the stop codon TGA, which should encode a protein comprising 259 amino acids with a 16 amino acid signal peptide analyzed by SignalP 3.0 software. After removal of signal peptide, the mature protein has a deduced molecular weight of 26033.2, a theoretical isoelectric point (pI) of 5.56, an instability index (II) of 35, and a grand average of hydropathicity (GRAVY) score of -0.290, indicating that the protein was a hydrophobic protein that confirmed by hydrophilic analysis by ProtScale tools. The analysis of the complete amino acid sequence translated from clone 5 sequencing by the GOR4 on-line software revealed that the secondary structure of *Der f* 3 comprised an extended strand (79 AA, 30.50%), random coil (163 AA, 62.93%), and alpha helix (17AA, 6.56%). The subcellular localization of *Der f* 3 was predicted to be extracellular by CELLO v.2.5 (Table 1). The InterProScan sequence search tool predicted three chymotrypsin active sites in the pro-protein (53–68 AA, 108–122 AA and

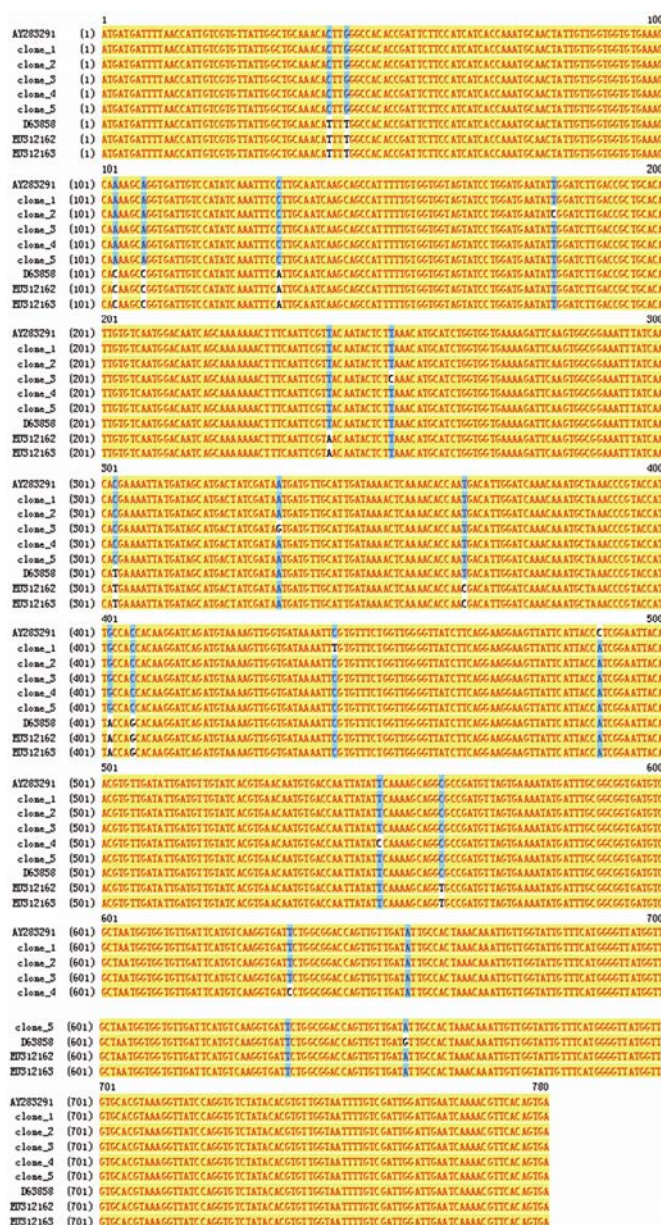


Figure 1. Nucleotide sequence alignment among different clones obtained in this study and the *Der f* 3 sequences published in GenBank. Non-similar sequences were indicated with foreground and automatic background, conservative with foreground and background, block of similar with foreground and background, identical with foreground and background, weakly similar with foreground and automatic background.

205–217 AA). Furthermore, the protein might also contain one N-glycosylation site, one cAMP- and cGMP-dependent protein kinase phosphorylation site, four protein kinase C phosphorylation sites, two casein kinase II phosphorylation sites, and five N-myristoylation sites (Table 2).

DISCUSSION

Polymorphism has been reported in several dust mite allergens, and previous studies suggest that the *Der f* 3 variation may be less frequent than the group 1 and 2



Figure 2. Alignment results among amino acid sequences deduced from different plasmids sequencing in this paper and other published *Der f 3* in GenBank. Non-similar sequences were expressed with foreground and automatic background, conservative with foreground and background, block of similar with foreground and background, identical with foreground and background, weakly similar with foreground and automatic background.

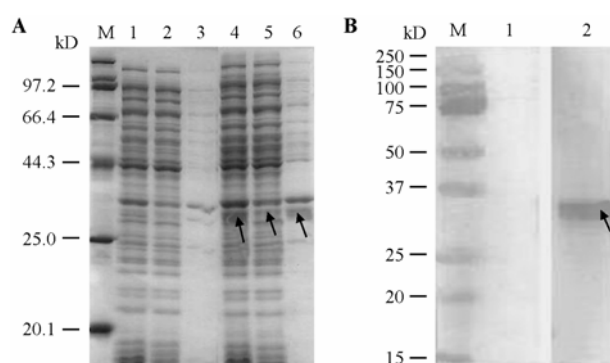


Figure 3. Expression rDer f 3 in *E. coli* BL21 cells. *E. coli* BL21 cells were transformed with either pET28a(+)-*Der f 3* or empty vector pET28a(+) as control. SDS-PAGE (stained by CBB-R250) and Western blotting analysis of rDer f 3 protein were performed as described in "Methods". **A:** SDS-PAGE analysis of rDer f 3 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 3*; Lane 5: the supernatant of cell containing pET28a(+)-*Der f 3*; Lane 6: the pellet of cells containing pET28a(+)-*Der f 3*. **B:** Western blotting analysis of rDer f 3 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 3*. Lane M: precision plus protein standards. The specific rDer f 3 band is indicated by arrows.

allergens because two *Der f 3* sequences from commercial mites in Australia differ from each other at only one position.¹⁸ In the present study, five positive cDNA clones of *Der f 3* were obtained by PCR, in which a total of 19 variant positions were identified after comparison with four published *Der f 3* sequences from GenBank. Seven of these 19 mismatched nucleotides were kept in the same among three or more sequences of these nine, i.e.

Table 1. Protein subcellular localization prediction result for *Der f 3* by CELLO v.2.5

Items	Localization	Reliability
Support vector machine		
Amino acid composition	Extracellular	0.870
N-peptide composition	Nuclear	0.223
Partitioned sequence composition	Extracellular	0.984
Physico-chemical composition	Extracellular	0.943
Neighboring sequence composition	Extracellular	0.388
CELLO prediction		
	Extracellular	3.395
	Nuclear	0.402
	Cytoplasmic	0.354
	Chloroplast	0.271
	Lysosomal	0.213
	Mitochondrial	0.108
	Vacuole	0.059
	Peroxisomal	0.056
	Golgi	0.045
	PlasmaMembrane	0.043
	Endoplasmic reticulum	0.039
	Cytoskeletal	0.015

Table 2. Function site analysis of *Der f 3* by PROSCAN software

Function site	AA position	AA sequences
N-glycosylation site	25–28	NATI
cAMP- and cGMP-dependent protein kinase phosphorylation site	75–78	KKLS
Protein kinase C phosphorylation site	73–75	SAK
	78–80	SIR
	83–85	TLK.
	255–257	SKR.
Casein kinase II phosphorylation site	55–58	SILD
	88–91	SGGE
Tyrosine kinase phosphorylation site	35–41	KAGDCPY
	176–183	REQCDQLY
N-myristoylation site	30–35	GGVKAK
	187–192	GADVSE
	197–202	GGDVAN
	203–208	GGVDSC
	204–209	GVDSCQ
Serine proteases, trypsin family, histidine active site	63–68	LTAHC
Serine proteases, trypsin family, serine active site	206–217	DSCQGDGGPVV

41 bp (C→T), 44 bp (G→T), 103 bp (A→C), 108 bp (C→A), 132 bp (C→A), 402 bp (G→A), and 406 bp (C→G). Although some of these variations might be caused by PCR errors, others should be inherent in dust mites, especially those resulting in amino acid variation. These seven variant positions were the same among three or more sequences; therefore, they could be polymorphic sites, resulting in four variant amino acid positions, which can be regarded as substitution sequences (14 (T→I), 15 (W→L), 35 (K→Q), and 136 (P→A)). The results of this work allude the sequence polymorphism in the group 3 allergen for the first time.

In the present study, the *Der f 3* gene was obtained by RT-PCR, the plasmid pET28a (+)-*Der f 3* was constructed and then transformed into *E. coli* BL21. Expression was induced by IPTG and rDer f 3 was expressed successfully, as confirmed by SDS-PAGE and Western blotting. This work lays the foundation for the large-scale production of recombinant allergens, for further study on *Der f 3*-IgE

binding activity and the mechanisms of the *Der f 3* allergic reaction.

With the compiling allergy-related data through functional genomics, gene microarrays and proteomics, bioinformatics tools are already crucial to allergy research in allergen characterization such as identification of structural motifs and B- and T- cell epitope.^{19,20} The InterProScan sequence search tool predicted three chymotrypsin active sites in the *Der f 3* pro-protein (53-68 AA, 108-122 AA and 205-217 AA) in our study.

When the complete sequence of *Der f 3* cDNA was analyzed, one open reading fragment (ORF) coding for a full-length cDNA of 780 bp from the start codon ATG to the stop codon TAA was found in clone 5, which was reckoned to encode a protein with 259 amino acid residues with the estimated molecular weight and theoretical isoelectric point of 26 033.2 Da and 5.56 respectively by ProtParam tool in ExPaSy web server. The instability index of this preprotein was calculated as 35, revealing that it was stable. The grand average of hydropathicity was calculated as -0.290, indicating that it was hydrophobic. In addition, the *Der f 3* protein was deduced to be a hydrophobic extracellular protein with three chymotrypsin active sites and some other function sites, which would provide the basis for further experimental study on its molecular characteristics whereas the amino acid positions of the functional sites might give a clue to the locations of epitopes to IgE antibodies of this allergen.

The secondary structure prediction results showed that the *Der f 3* allergen was consisted of an extended strand (30.50%), random coil (62.93%), and alpha helix (6.56%). Given that the most common secondary structures are alpha helices and extended beta sheets, the random coil may not be a true specific shape, but is a statistical distribution of shapes for all the chains, where the monomer subunits are oriented randomly while still being bonded to adjacent units. This may be the reason why little is known about the three-dimensional structure of the group 3 allergens.

REFERENCES

- Arlian LG, Platts-Mills TA. The biology of dust mites and the remediation of mite allergens in allergic disease. *J Allergy Clin Immunol* 2001; 107: s406-s413.
- Fernández-Caldas E. Dust mite allergens: mitigation and control. *Curr Allergy Asthma Rep* 2002; 2: 424-431.
- 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.
- Thomas WR, Smith WA, Hales BJ. The allergenic specificities of the house dust mite. *Chang Gung Med J* 2004; 2: 7563-7569.
- Sun JL, Zhang HY, Guo ZY, Ying WT, Qian XH, Wang JL. Comparison of three methods of protein extraction from *Dermatophagoides pteronyssinus* for two-dimensional electrophoresis. *Chin Med J* 2009; 24: 64-68.
- Chen ZG, Li M, Chen YF, Ji JZ, Li YT, Chen W, et al. Effects of *Dermatophagoides pteronyssinus* allergen-specific immunotherapy on the serum interleukin-13 and pulmonary functions in asthmatic children. *Chin Med J* 2009; 122: 1157-1161.
- Mauro M, Russello M, Alesina R, Sillano V, Alessandrini A, Dama A, et al. Safety and pharmacoeconomics of a cluster administration of mite immunotherapy compared to the traditional one. *Eur Ann Allergy Clin Immunol* 2006; 38: 31-34.
- Zhang L, Wang C, Han D, Wang X, Zhao Y, Liu J. Comparative study of cluster and conventional immunotherapy schedules with *Dermatophagoides pteronyssinus* in the treatment of persistent allergic rhinitis. *Int Arch Allergy Immunol* 2009; 148: 161-169.
- Cox L. Accelerated immunotherapy schedules: review of efficacy and safety. *Ann Allergy Asthma Immunol* 2006; 97: 126-137.
- Jeong KY, Hongb CS, Yong TS. Recombinant allergens for diagnosis and immunotherapy of allergic disorders, with emphasis on cockroach allergy. *Curr Protein Pept Sci* 2006; 7: 57-71.
- Mothes N, Valenta R, Spitzauer S. Allergy testing: the role of recombinant allergens. *Clin Chem Lab Med* 2006; 44: 125-132.
- Qiu J, Li GP, Liu ZG, Ran PX, Zhong NS. DNA vaccine encoding Der p2 allergen down-regulates STAT6 expression in mouse model of allergen-induced allergic airway inflammation. *Chin Med J* 2006; 119: 185-190.
- Piboonpocanun S, Malainual N, Jirapongsananuruk O, Vichyanond P, Thomas WR. Genetic polymorphisms of major house dust mite allergens. *Clin Exp Allergy* 2006; 36: 510-516.
- Stewart GA, Ward LD, Simpson RJ, Thompson PJ. The group III allergen from the house dust mite *Dermatophagoides pteronyssinus* is a trypsin. *Immunology* 1992; 75: 29-35.
- Derewenda U, Li J, Derewenda Z, Dauter Z, Mueller GA, Rule GS, et al. The crystal structure of a major dust mite allergen Der p 2, and its biological implications. *J Mol Biol* 2002; 318: 189-197.
- Cui YB, Zhou P, Peng JL, Peng M, Zhou Y, Lin YZ, et al. 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.
- Cui YB, Peng J, Zhou P, Peng M, Qian SY. Bioinformatic studies on the group 2 allergens of *Dermatophagoides farinae* from China. *Asian Pac J Allergy Immunol* 2007; 25:199-206.
- Smith WA, Thomas WR. Sequence polymorphisms of the Der p 3 house dust mite allergen. *Clin Exper Allergy* 1996; 26: 571-579.
- Brusic V, Petrovsky N. Bioinformatics for characterisation of allergens, allergenicity and allergic crossreactivity. *Trends Immunol* 2003; 24: 225-228.
- Brusic V, Petrovsky N, Gendel SM, Millot M, Gigonzac O, Stelman SJ. Computational tools for the study of allergens. *Allergy* 2003; 58: 1083-1092.

(Received July 15, 2009)
Edited by WANG De