

Sequence Diversity of the Bla g 4 Cockroach Allergen, Homologous to Lipocalins, from *Blattella germanica*

Kyoung Yong Jeong^a Myung-hee Yi^a Kyoung-jin Jeong^a Haeseok Lee^a
Chein-Soo Hong^b Tai-Soon Yong^a

^aDepartment of Environmental Medical Biology and Institute of Tropical Medicine, Arthropods of Medical Importance Resource Bank, and ^bDepartment of Internal Medicine and Institute of Allergy, Brain Korea 21 Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea

Key Words

Bla g 4 · Cockroach · Polymorphism · RNA editing

Abstract

Background: The cockroach allergen Bla g 4, a putative lipocalin, is known to exhibit frequent sequence variations. However, the previously reported cDNA sequences are truncated at the N terminus. This study was undertaken to investigate the mechanisms by which these sequence variations are generated. **Methods:** Rapid amplification of cDNA ends PCR and RT-PCR were performed to obtain the full sequence of the Bla g 4 cDNA, and PCR was also used to clone the Bla g 4 genomic DNA. In addition, Bla g 4 protein variants were analyzed by two-dimensional gel electrophoresis. **Results:** Nine additional amino acid residues at the N terminus of Bla g 4 were identified, and 2 genes encoding Bla g 4, both of which consisted of 5 exons, were cloned. Examination of 34 clones of Bla g 4 cDNA obtained by RT-PCR revealed 14 variants. In particular, Bla g 4 sequences showed frequent clusters of variations in residues 38–45, 61–82 and 144–163. Differences in cDNA sequences may imply that RNA sequences are edited after transcription. More than 10 spots were identified between pH 5 and 7 upon two-dimensional gel electrophoresis, indicating that multiple variants of Bla g 4 are

produced at the protein level. **Conclusions:** Genetic polymorphisms among individual cockroaches, the existence of multiple genes and sequence variations caused by RNA editing produce sequence diversity of Bla g 4, which may influence its allergenicity. The sequence information obtained in this study will be helpful for the standardization of the cockroach allergen and thereby aid in the development of diagnostics and immunotherapeutics.

Copyright © 2008 S. Karger AG, Basel

Introduction

Various arthropods that infest human dwellings may elicit allergic disorders [1]. Identification and characterization of allergens are crucial for the diagnosis and remedy of the symptoms [2]. Sensitization to cockroach allergens is an important risk factor for asthma-related hospitalization [3]. Recently, molecular biological techniques have allowed for the identification of several groups of cockroach allergens [4, 5].

Bla g 4, a putative lipocalin, is one of the important cockroach allergens [6]. It is expressed in the adult male reproductive system and is speculated to bind pheromones. In a recent study, it was found that Bla g 4 is pro-

duced in the conglobate gland and the apical utricles (accessory glands of the male reproductive system) and transferred to the female's genital tract during copulation [7]. Bla g 4 is discarded with the spermatophore from the female 24 h after copulation.

Forty to 60% of cockroach-reactive asthmatics show serum IgE antibody to recombinant Bla g 4 [6]. The highest yield of recombinant Bla g 4 was obtained in the methylotrophic yeast *Pichia pastoris*, and this recombinant was shown to have excellent allergenic activity [8]. However, only 17.4% (20/115) and 20.5% (7/34) of cockroach-allergic subjects were shown to have serum IgE specific to the recombinant Bla g 4 in recent studies [9, 10].

Many animal allergens have been described to be lipocalins [11]. Important mammalian allergens such as those from cats (Fel d 4) [12], dogs (Can f 1 and Can f 2) [13], cows (Bos d 2 and Bos d 5) [14, 15], horses (Equ c 1 and Equ c 2) [16, 17], rats (Rat n 1) and mice (Mus m 1) [18] all belong to the lipocalin protein family. A major allergen of *Argas reflexus*, the European pigeon tick, can cause anaphylactic reactions upon biting and has also been shown to have significant homology to lipocalins (Arg r 1) [19]. Lipocalins exhibit diverse functions including roles in retinol transport, cryptic coloration, olfaction, pheromone transport, prostaglandin synthesis, regulation of cell homeostasis, modulation of immune responses and the general clearance of endogenous and exogenous compounds [20]. Despite their low amino acid sequence identity, lipocalins are known to share a similar crystal structure, namely 8–10 antiparallel β -sheets forming an internal ligand binding site at the N terminus and an α -helix at the C terminus [21].

We have previously described the sequence diversity of Bla g 4 [22]. However, full sequence information for Bla g 4 is not yet available. Information on the sequence diversity of major allergens is very important for complete allergen characterization and standardization [23, 24]. In this study, we obtained full-length cDNA and genomic DNA sequence information for Bla g 4 in order to investigate the mechanism by which sequence variations could be generated.

Materials and Methods

Cockroaches

Male cockroaches were isolated from the colonies that are maintained in the insect rearing facility of the Korea National Arthropods of Medical Importance Resource Bank, Department of Environmental Medical Biology, Yonsei University College of

Medicine, Seoul, Korea. Mouse food was used as a diet. Cockroaches were collected after being euthanized with carbon dioxide gas and kept at -70°C until use.

Molecular Cloning of Bla g 4 cDNA

Total RNA was isolated from 6 adult male cockroaches using TRIzol reagent (Invitrogen, Carlsbad, Calif., USA). The full-length cDNA sequence encoding Bla g 4 was obtained using a SMARTTM rapid amplification of cDNA ends (RACE) cDNA amplification kit (BD Biosciences, Palo Alto, Calif., USA). A 5'-RACE CDS primer (adaptor primer, BD Biosciences) and Bla g 4 gene-specific primer (Bg4RC, 5'-GCGGGACAGCCTTCCACAATTGC-3') designed on the Bla g 4 internal sequence were used for the amplification of the 5' end. The PCR product was gel-purified and ligated into pGEM-T Easy vector (Promega, Madison, Wisc., USA), and the DNA sequence was determined at Solgent (Daejeon, Korea).

For the analysis of sequence variations of Bla g 4, the full-length cDNA sequence encoding Bla g 4 was amplified by RT-PCR. The oligonucleotides used for the amplification of the full-length cDNA encoding Bla g 4 were as follows: Bla g 4NtermF (forward), 5'-GGATCCCAGATGTGCATCACTGGA-3' and Bg4R (reverse), 5'-CTTAGTGACATGTGGAGTG-3'. The amplified DNA fragment was gel-purified and ligated into pGEM-T Easy, and DNA sequences were analyzed.

Cloning of Genomic DNA Encoding Bla g 4

Genomic DNA was isolated from 1 male cockroach using phenol-chloroform premixed with isoamyl alcohol (25:24:1; Amresco, Solon, Ohio, USA) after pulverization under liquid nitrogen. Genes encoding Bla g 4 were amplified (5 min at 95°C for initial denaturation, followed by 35 cycles of 30 s at 94°C , 30 s at 58°C and 1 min at 72°C , and 9 min at 72°C for the last extension) using the primers used for the full-length cDNA sequence of Bla g 4 with Ex Taq DNA polymerase, which has 3' to 5' exonuclease (proof-reading) activity (Takara, Japan). The PCR product was ligated into pCR4-TOPO vector (Invitrogen), and the DNA sequence was analyzed at Solgent.

Expression of Recombinant Bla g 4

Previously, the EF202172 cDNA sequence described as a major form of Bla g 4 [22] was used for the expression of recombinant protein. The oligonucleotide primer sequences used for the amplification were as follows: Bg4E1F, 5'-GGATCCGGCAGTTTTGGCACTA-3' and Bg4E5R, 5'-CTCGAGTTAGTTACATGTGAG-3'. The underlining indicates the *Bam*HI and *Xho*I restriction enzyme sites, respectively, introduced for directional cloning into the expression vector. PCR products were ligated into pGEM-T Easy vector, digested with the appropriate restriction enzymes and subcloned into the pET 28b vector (Novagen, Madison, Wisc., USA). Recombinant protein was expressed in *Escherichia coli* BL21 (DE3) and purified from the inclusion bodies. The insoluble fraction was solubilized in 20 mM Tris, 500 mM NaCl, 5 mM imidazole and 6 M urea, pH 7.9, and purified using a nickel-nitrilotriacetic acid resin (Qiagen, Valencia, Calif., USA) according to the manufacturer's instructions. Purified protein was separated on a 15% polyacrylamide gel containing sodium dodecyl sulfate under denaturing conditions and visualized by Coomassie brilliant blue.

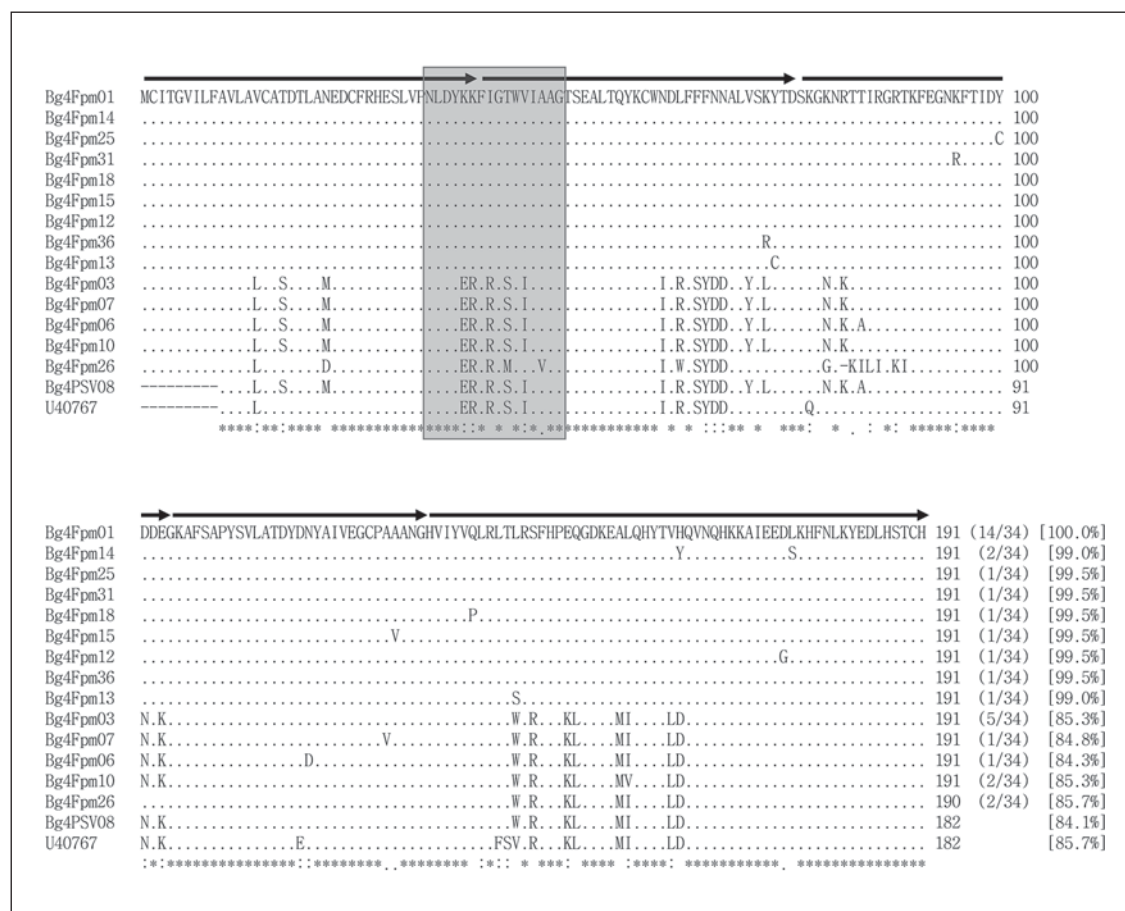


Fig. 1. Multiple sequence alignment of Bla g 4 deduced from cDNA sequences that were obtained by RT-PCR. The frequency and identity of each sequence are shown in parentheses. Arrows indicate the exons identified by genomic DNA analysis. The shaded area indicates the structurally conserved regions of lipocalin fold.

Preparation of Mouse Anti-Bla g 4 Sera

Six-week-old female BALB/c mice were obtained from Japan SLC, Inc. (Shizuoka Prefecture, Japan). Four mice were intraperitoneally immunized with 30 µg of recombinant Bla g 4 mixed with complete Freund's adjuvant (Sigma-Aldrich, St. Louis, Mo., USA). The same amount of antigen mixed with incomplete Freund's adjuvant (Sigma-Aldrich) was intraperitoneally administered twice at 3-week intervals. The production of anti-Bla g 4 antibodies was monitored by an enzyme-linked immunosorbent assay, and the mice were sacrificed 3 days after the second booster injection. The mouse anti-Bla g 4 sera produced from 4 mice were pooled and used to investigate the variants of Bla g 4 at the protein level.

Two-Dimensional Gel Electrophoresis and Western Blot Analysis

In order to investigate Bla g 4 at the protein level, proteins were extracted from the abdomen of adult male cockroaches and analyzed by two-dimensional gel electrophoresis and Western blot analysis. Abdomens were dissected from frozen cockroach-

es that had been reared at the insect rearing facility of the Arthropods of Medical Importance Resource Bank. In total, 2.76 g of cockroach abdomen were homogenized under liquid nitrogen, and the proteins were extracted in phosphate-buffered saline as previously described [25]. Two-dimensional gel electrophoresis was performed after desalting with trichloroacetic acid. One milligram of the extract was loaded onto the 18-cm tube for the first-dimension gel. After isoelectric focusing (pH 3–10), second-dimension gel electrophoresis was performed in a 9–17% gradient polyacrylamide gel containing sodium dodecyl sulfate. The separated proteins were either stained with Coomassie brilliant blue or electrophoretically transferred onto a nitrocellulose membrane (Osmonics, Westborough, Mass., USA). After blocking with 3% skim milk, Bla g 4 proteins were probed with mouse anti-Bla g 4 sera (1:250) for 1.5 h. After washing in PBS, the membrane was incubated with goat antimouse IgG conjugated with alkaline phosphatase (Sigma, St. Louis, Mo., USA) for 1 h, and the color was developed with nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl-phosphate (Promega).

Fig. 2. A schematic representation of the genomic DNA organization encoding Bla g 4.

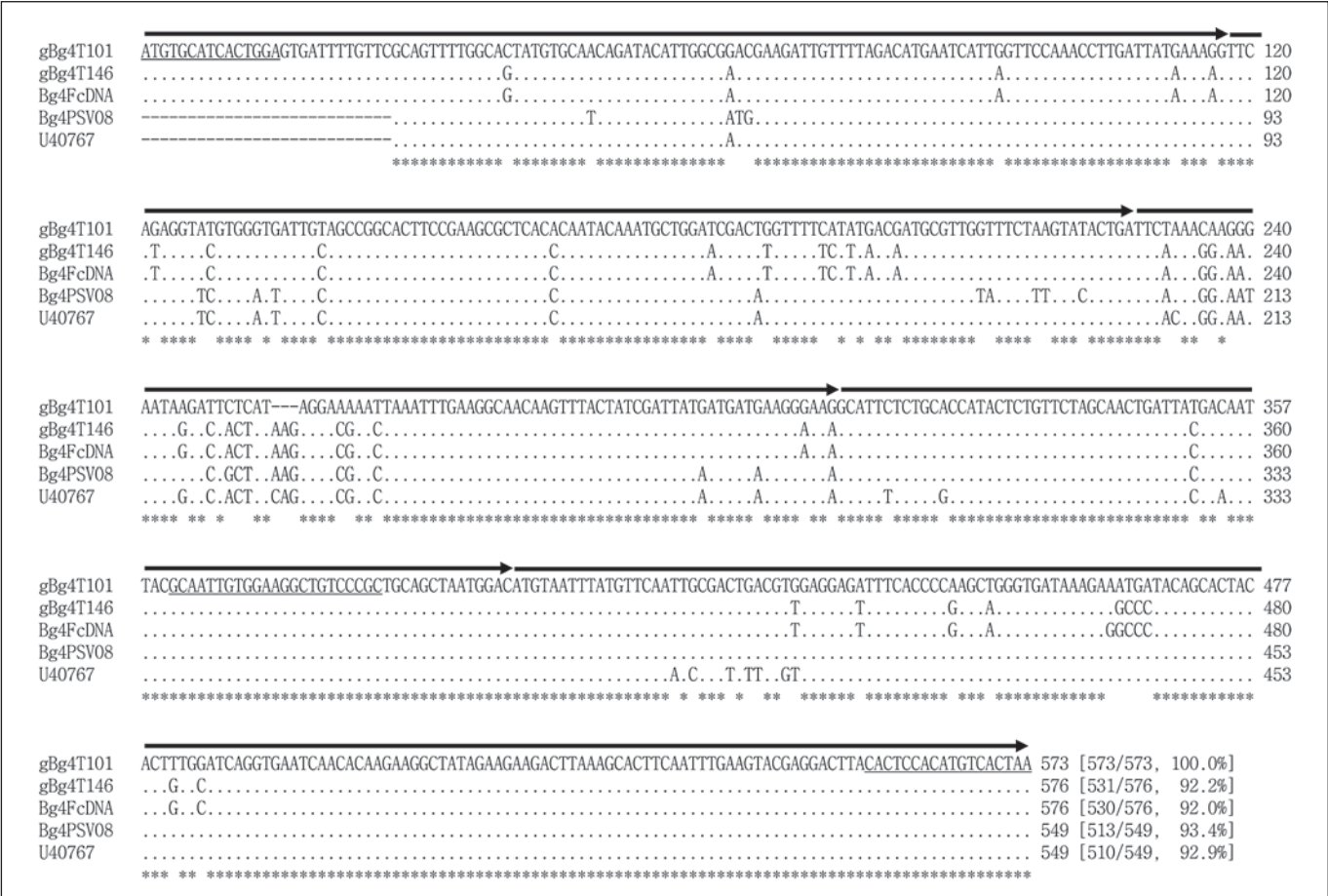
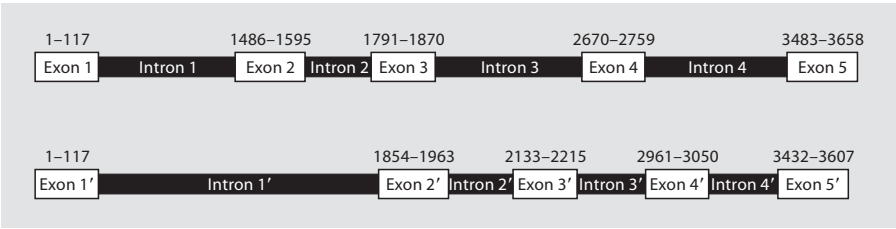


Fig. 3. Comparison of DNA sequences with exons from genomic DNA and the Bla g 4 cDNA sequences of predominant forms. Arrows indicate coding regions of genomic DNA. Underlining indicates the regions where primers were designed for PCR amplification.

Results

Cloning and Sequence Analysis of Full-Length Bla g 4 cDNA

RACE-PCR allowed the isolation of the complete coding sequence of Bla g 4, which contains 9 additional amino acids at the N terminus. Subsequently, RT-PCR was

performed to investigate the sequence variations of Bla g 4. The deduced amino acid sequences from 34 clones showed 14 different sequences (fig. 1). The most frequently encountered form, Bg4Fpm01, occurred in 14 of 34 clones (41.2%) and showed 84.3–99.5% sequence identity with other variants. Moreover, it shared 84.1 and 85.7% identity with previously described Bla g 4 sequenc-

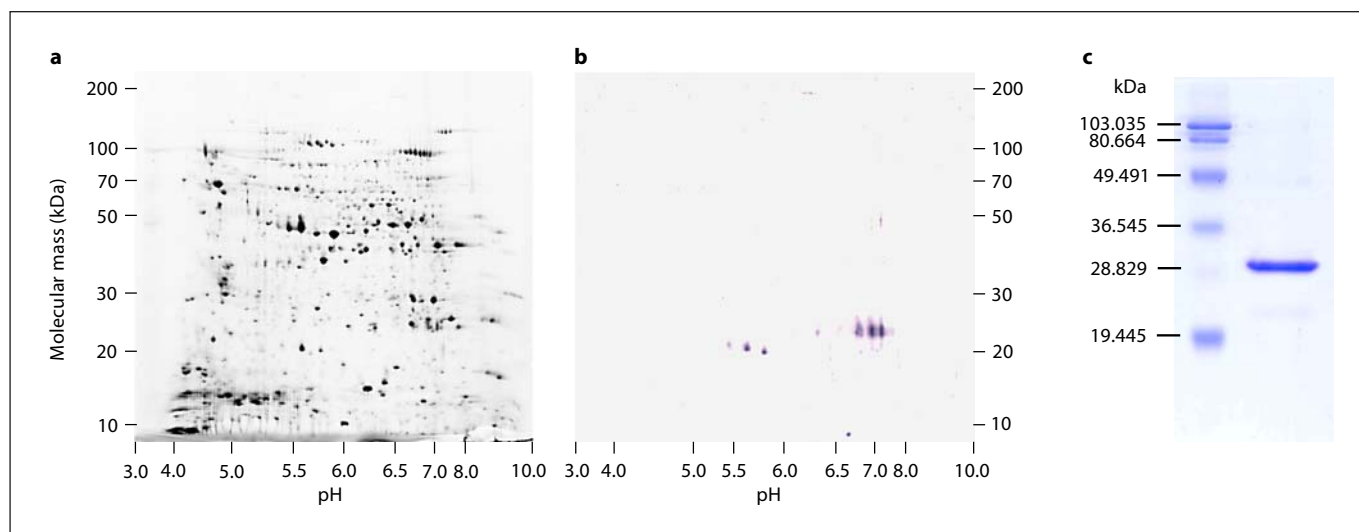


Fig. 4. Two-dimensional gel electrophoresis of German cockroach male extract (a), identification of Bla g 4 variants by Western blot analysis (b) and SDS-PAGE analysis of purified recombinant Bla g 4 which was used for the production of mouse anti-Bla g 4 sera (c).

es in GenBank under accession numbers EF202172 (Bg4PSV08) and U40767, respectively. The second most predominant form encountered, Bg4Fpm03, was found in 5 of the 34 clones (14.7%). One of these clones, Bg4Fpm32, showed a substitution at amino acid 7, which may be the result of a primer error, and was essentially identical with Bg4Fpm03. Interestingly, the amino acid at position 81 of variant BgFpm26 was found to be a deletion compared to the remaining variants. One of the regions which is well conserved structurally (MOTIF1) among the lipocalins is found to be well conserved in Bla g 4 sequences.

Molecular Cloning of Genomic DNA Encoding Bla g 4

Two genes encoding Bla g 4 were identified by the PCR amplification of genomic DNA from 1 male cockroach (fig. 2). These 2 genes were identified by the analysis of 9 clones. Two of the clones were found to be identical, while 7 of them were found to be derived from the same gene. Each gene consisted of 5 exons and 4 introns comprising 3,658 and 3,607 bp, respectively. The DNA sequences from the genomic DNA encoding Bla g 4 were deposited in GenBank under accession numbers EU635774 (3,607 bp) and EU635775 (3,658 bp). Comparison of exons from genomic DNA and cDNA sequences showed 92.0–99.8% identity (fig. 3). However, none of the cDNA sequences showed the same sequence as the genomic DNA sequences. This observation may suggest that the Bla g 4 mRNA sequences are somehow modified after transcription.

Analysis of Bla g 4 Proteins

A proteomic approach was carried out to find out whether different Bla g 4 variants are produced at the protein level. Bla g 4 was detected using mouse antisera raised against recombinant Bla g 4 from two-dimensionally separated proteins of male German cockroach extract. More than 10 proteins from pH 5 to 7 showed immunoreactivity, indicating the presence of multiple isoforms of Bla g 4 (fig. 4). However, the possibility of multiple proteins resulting from posttranslational modifications rather than from posttranscriptional modifications could not be excluded by this experiment.

Discussion

Standardization of allergens is essential both for the diagnosis and immunotherapy of allergic diseases [24]. Allergenic proteins are often discovered to be a mixture of similar molecules with some variations in the amino acid sequences, which are called isoallergens or variants depending on the degree of similarity [26]. This phenomenon, often called polymorphism, has an influence on allergenicity. The presence of several alleles of the genes or gene families may produce sequence polymorphisms. Hence, the standardization of allergens may require the identification of isoforms and their proportions.

In this study, we analyzed sequence variations with the full-sequence information for Bla g 4, which was obtained by RACE-PCR. While variations were found in both the previous and present studies, we observed less frequent variations as compared to the previously published experiments. It is possible that the primers used for PCR may have some effect on the results, because the primer sequence used for the RT-PCR in the previous study was designed on the U40707 sequence which has Leu at position 14. PCR errors could have contributed to the variability of the sequences, as we did not use DNA polymerase which has proofreading activity for RT-PCR. The insertion or deletion of amino acids was also described in the previous study. Here, the conceptually translated amino acid sequence of 1 of our variants, BgFpm26, revealed the deletion of 1 amino acid at position 81. Analysis of 2 different genomic DNA sequences isolated by PCR using the same primers that were used for the RT-PCR showed 5 exons (fig. 2). However, none of the DNA sequences from genomic DNA coding regions were found to have the same sequence obtained by RT-PCR (fig. 3). Moreover, more than 10 Bla g 4 variants were detected by two-dimensional Western blot analysis (fig. 4), but it does not seem that there is a separate gene encoding each variant. These results may indicate that the Bla g 4 sequences are modified after transcription. RNA editing may explain these observations. Pre-RNA

editing by adenosine deaminase is a widespread phenomenon in animal tissues [27]. Different degrees of RNA editing may result in different degrees of sequence variations because a small number of cockroaches (6) was used for the analysis of polymorphisms by RT-PCR. If this is the case, the sequence previously described (EF202172) may better represent the predominant variant of Bla g 4.

In summary, sequence variations of Bla g 4 were investigated using the full cDNA sequence information of Bla g 4. Two different genes coding for Bla g 4 were identified. The difference in DNA sequences between the cDNA and the coding region of the genomic DNA implies that the modification of RNA sequences follows transcription. Multiple variants were also found to be produced by the proteomic approach, suggesting that the effects of sequence variations of Bla g 4 on its allergenicity need to be examined.

Acknowledgements

We wish to acknowledge technical support from the Yonsei Proteome Research Center (<http://www.proteomix.org>). This work was supported by a Korea Science and Engineering Foundation grant funded by the Korean government (No. R01-2006-000-10997-0).

References

- Jeong KY, Hong CS, Yong TS: Domestic arthropods and their allergens. *Protein Pept Lett* 2007;14:934–942.
- Jeong KY, Hong 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.
- Kang B, Vellody D, Homburger H, Yunginger JW: Cockroach cause of allergic asthma. *J Allergy Clin Immunol* 1979;63:80–86.
- Gore JC, Schal C: Cockroach allergen biology and mitigation in the indoor environment. *Annu Rev Entomol* 2007;52:439–463.
- Pomes A, Wunschmann S, Hindley J, Vailes LD, Chapman MD: Cockroach allergen: function, structure and allergenicity. *Protein Pept Lett* 2007;14:960–969.
- Arruda LK, Vailes LD, Hayden ML, Benjamin DC, Chapman MD: Cloning of cockroach allergen, Bla g 4, identifies ligand binding proteins (or calycons) as a cause of IgE antibody responses. *J Biol Chem* 1995; 270:31196–31201.
- Fan Y, Gore JC, Redding KO, Vailes LD, Chapman MD, Schal C: Tissue localization and regulation by juvenile hormone of human allergen Bla g 4 from the German cockroach, *Blattella germanica* (L). *Insect Mol Biol* 2005;14:45–53.
- Vailes LD, Kinter MT, Arruda LK, Chapman MD: High-level expression of cockroach allergen, Bla g 4, in *Pichia pastoris*. *J Allergy Clin Immunol* 1998;101:274–280.
- Satinover SM, Reefer AJ, Pomes A, Chapman MD, Platts-Mills TA, Woodfolk JA: Specific IgE and IgG antibody-binding patterns to recombinant cockroach allergens. *J Allergy Clin Immunol* 2005;115:803–809.
- Slater JE, James R, Pongracic JA, et al: Biological potency of German cockroach allergen extracts determined in an inner city population. *Clin Exp Allergy* 2007;37:1033–1039.
- Mantjarvi R, Rautianinen J, Virtanen T: Lipocalins as allergens. *Biochim Biophys Acta* 2000;1482:308–317.
- Smith W, Butler AJ, Hazell LA, et al: Fel d 4, a cat lipocalin allergen. *Clin Exp Allergy* 2004;34:1732–1738.
- Konieczny A, Morgenstern JP, Bizinkauskas CB, et al: The major dog allergens, Can f 1 and Can f 2, are salivary lipocalin proteins: cloning and immunological characterization of the recombinant forms. *Immunology* 1997;92:577–586.
- Rautiainen J, Auriola S, Rouvien J, et al: Molecular and crystal properties of Bod s 2, an allergenic protein of the lipocalin family. *Biochim Biophys Res Commun* 1998;247: 746–750.
- Wu SY, Perez MD, Puyol P, Sawyer L: Beta-lactoglobulin binds palmitate within its central cavity. *J Biol Chem* 1999;274:170–174.
- Lascombe MB, Gregoire C, Poncet P, et al: Crystal structure of the allergen Equ c 1. A dimeric lipocalin with restricted IgE-reactive epitopes. *J Biol Chem* 2000;275:21572–21577.

- 17 Bulone V, Krogstad-Johnsen T, Smestad-Paulsen B: Separation of horse dander allergen protein by two-dimensional electrophoresis – molecular characterization and identification of Equ c 2.0101 and Equ c 2.0102 as lipocalin proteins. *Eur J Biochem* 1998;253:202–211.
- 18 Bocskai Z, Groom CR, Flower DR, et al: Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* 1992;360:186–188.
- 19 Hilger C, Bessot JD, Hutt N, et al: IgE-mediated anaphylaxis caused by bites of the pigeon tick *Argas reflexus*: cloning and expression of the major allergen Arg r 1. *J Allergy Clin Immunol* 2005;115:617–622.
- 20 Flower DR: The lipocalin protein family: structure and function. *Biochem J* 1996;318:1–14.
- 21 Rouvinen J, Rautiainen J, Virtanen T, et al: Probing the molecular basis of allergy. Three-dimensional structure of the bovine lipocalin allergen Bos d 2. *J Biol Chem* 1999;274:2337–2343.
- 22 Jeong KY, Lee H, Shin KH, et al: Sequence polymorphisms of major German cockroach allergens Bla g 1, Bla g 2, Bla g 4, and Bla g 5. *Int Arch Allergy Immunol* 2008;145:1–8.
- 23 van Ree R: Analytical aspects of the standardization of allergenic extracts. *Allergy* 1997;52:795–805.
- 24 van Ree R: Indoor allergens: relevance of major allergen measurements and standardization. *J Allergy Clin Immunol* 2007;119:270–277.
- 25 Jeong KY, Lee J, Lee IY, Ree HI, Hong CS, Yong TS: Allergenicity of recombinant Bla g 7, German cockroach tropomyosin. *Allergy* 2003;58:1059–1063.
- 26 King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W: Allergen nomenclature. WHO/IUIS Allergen Nomenclature Subcommittee. *Int Arch Allergy Immunol* 1994;105:224–233.
- 27 Nishikura K: Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Biol* 2006;7:919–931.