

# Novel Allergen Structures with Tandem Amino Acid Repeats Derived from German and American Cockroach\*

(Received for publication, July 9, 1998, and in revised form, September 1, 1998)

Anna Pomés‡, Erik Melén, Lisa D. Vailes, Jacques D. Retief§, L. Karla Arruda¶, and Martin D. Chapman

From the Asthma & Allergic Diseases Center, Department of Medicine, and the §ITC-Health Sciences, University of Virginia, Charlottesville, Virginia 22908, and the ¶Department Parasitologia, Microbiologia Imunologia, Faculdade de Medicina de Ribeirao Preto-USP, Av. Bandeirantes 3900, Ribeirao Preto, SP, Brazil

Cockroaches produce potent allergens that are an important cause of asthma. The two principal domiciliary cockroach species, *Blattella germanica* and *Periplaneta americana*, secrete major allergens, Bla g 1 and Per a 1. Here, we report the molecular cloning of three Bla g 1 cDNA clones, which showed 70% amino acid sequence identity with Per a 1. Plaque immunoassays with human IgE antibodies or murine monoclonal antibodies showed that these allergens were antigenically cross-reactive. The Bla g 1 sequences also showed homology to five previously undefined cockroach allergen sequences. An unusual feature of all these sequences was that they contained multiple tandem amino acid repeats of ~100 amino acid residues. Between one and seven repeat units were identified by dot-plot matrix analysis. The sequences also showed homology to a mosquito protein involved in digestion (ANG12 precursor) and to mitochondrial energy transfer proteins. High levels of Bla g 1 were found in cockroach hindgut and proventriculus. Amino acid sequencing of natural Bla g 1 and Per a 1 suggested that these allergens are cleaved by trypsin-like enzymes following secretion into the digestive tract. The repeat sequences appear to have evolved by duplication of an ancestral amino acid domain, which may have arisen from the mitochondrial energy transfer proteins.

Inhalation of environmental allergens from pollens, mites, animal danders, insects, and fungi induces IgE antibody responses in ~20% of the Western population. Immediate hypersensitivity reactions caused by IgE antibodies (Ab)<sup>1</sup> occur in genetically predisposed individuals and may result in the development of allergic diseases such as rhinitis, asthma, and atopic dermatitis. Most allergens are soluble low molecular proteins or glycoproteins (5–50 kDa), which rapidly diffuse through mucosal surfaces where they cross-link IgE Ab bound to mast cells and basophils (1). Over 200 allergen sequences

have now been cloned. Allergens are usually single polypeptide chains and occasionally occur as homodimers (e.g. *Alternaria*, Alt a 1) or heterodimers (cat, Fel d 1). Sequence comparisons, structural studies, and assays of biological function have shown that allergens are a diverse group of proteins, including enzymes, ligand binding proteins, and structural and regulatory proteins (2, 3). There is some evidence, especially for mite allergens, that proteolytic enzyme activity contributes to allergenicity (4, 5). Aerodynamic particle size, allergen dose, and sensitization through the respiratory tract are also important determinants for IgE responses.

Cockroaches produce potent allergens that are a common cause of asthma in United States cities (6–8). Several allergens have been cloned from German cockroach (*Blattella germanica*), including Bla g 2 (aspartic protease); Bla g 4 (lipocalin); and Bla g 5 (glutathione transferase) (9–11). These allergens elicit IgE Ab responses in 50–80% of cockroach allergic patients and are *B. germanica*-specific (12). Antigenically cross-reactive allergens have been identified in both German and American cockroach (Bla g 1 and Per a 1) but their structure was unknown (13–15). We recently cloned the cDNA encoding Per a 1 and here we report the sequences of three Bla g 1 cDNA clones (16). A novel feature of these allergens, which has not previously been described among allergen structures, is that they are composed of several tandem repeats of ~100 amino acid residues. The allergens show sequence homology to a mosquito protein associated with digestive function (ANG12 precursor), and the repeat sequences may have evolved from ancestral amino acid domains found in mitochondrial energy transfer proteins.

## EXPERIMENTAL PROCEDURES

**cDNA Library Screening**—A *B. germanica* cDNA library was prepared in the Uni-ZAP™ XR expression vector (Stratagene, La Jolla, CA) using 10 µg of mRNA isolated from adult cockroaches as described previously (9, 10). The library was screened with a murine monoclonal antibody (mAb) against Bla g 1, clone 10A6 (17). Positive plaques were detected using alkaline phosphatase-labeled goat anti-mouse IgG (1:1000) (KPL, Gaithersburg, MD) followed by color development with BCIP/nitro blue tetrazolium (KPL, Gaithersburg, MD) (9, 10). Three positive clones (Bla g 1 clone 1, clone 2, and clone 3) were identified and further screened by plaque immunoassay for IgE Ab using 15 sera from cockroach-allergic patients. Nine sera were obtained from patients having detectable IgE Ab to Bla g 1 by radioimmunoassay; four patients had IgE Ab to cockroach, but not to Bla g 1; and two sera were from nonallergic controls. Positive plaques were detected using alkaline phosphatase-labeled goat anti-human IgE (1:500) (KPL). In addition, the Bla g 1 cDNA clones were screened against a panel of high performance liquid chromatography-purified mAb to cockroach allergens: Bla g 1 (mAb 10A6), Bla g 2 (mAb 8F4), and Per a 1 (mAb Per a 1-03, kindly provided by Drs. Henmar and Schou, ALK-ABELLO, Horsholm, Denmark) (17). The mAb were used at 1–10 µg/ml for screening, and bound mAb was detected using alkaline phosphatase-labeled goat anti-mouse IgG Ab at 1/1000 dilution.

\* This work was supported by National Institutes of Health Grants AI 32557 and AI 34601. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF072219, AF072221, and AF072220.

‡ To whom correspondence and reprints should be addressed: Asthma & Allergic Diseases Center, University of Virginia, Box 225, Health Sciences Center, Charlottesville, VA 22908. Tel.: 804-982-3328; Fax: 804-924-5779; E-mail: ap8d@avery.med.virginia.edu.

<sup>1</sup> The abbreviations used are: Ab, antibody; mAb, monoclonal antibody; BCIP, 5-bromo-4-chloro-3-indolyl phosphate; ELISA, enzyme-linked immunosorbent assay; ER, endoplasmic reticulum.

```

Bla g 1.02  1  NAIEFLNNIHDLGLGIPHIPVTARKHHRRGVITGLIDDI IAILPVDDLYA  50

Bla g 1.01  1  -----NLEKLEKREKVD  12
                    ||:|:|:|
51  LFQEKLETSPEFKALYDAIRSPEFQSIGVTLEAMPEYQNLIQKLDKGV  100

13  VDKI IELIRALFGLTLNAKASRNLDLQDFLALIPVDQIIA IATDYLAN  62
    || ||||| :| : : . | : ||||| ||||| ||:|:| |||||
101 VDHI IELIHQIFNI...VRDTRGLPEDLQDFLALIPTDQVLAIAADYLAN  147

63  DAEVQA AVAYLQ SDEFETIVVTLDALPELQNFLNFLEANGLNAIDFLNGI  112
    |||.||| |||.||| |||.||| .|.||| .|||:|:|:|:|:|
148 DAEVKA AVEYLSDEFETIVVTVDSLPEFKNFLNFLTQTNGLNAIEFLNNI  197

113 HDLLGIPHIPVSGRKYHIRRGVITGLIDDVLA IAILPLDDLKALFNEKLET  162
    ||||| ||||| .||| |:|:|:|:|:|:|:|:|:|:|:|:|:|:|
198 HDLLGIPHIPVTGRK.HLRRGVITGLIDDI IAILPVDDLYALFQEKLET  246

163 SPDFLALYNAIKSPEFQSIGVQTLNAMPEYQNLLEKLEKREKGVVDVKI IELI  212
    |:| | |. |:|:|:|:|:|:|:| | | | |. |:|:|:|:|:|:|:|
247 SPEFKALYDAIRSPEFQSIGVETLKAMPEYQSLIQKLDKGVVDVHI IELI  296

213 RALFGLTLNAKASRNLDLQDFLALIPVDQIIA IATDYLANDAEVQA AV  262
    :| : : . | : ||||| ||||| |:|:|:|:|:|:|:|:|
297 HQIFNI...VRDTRGLPEDLQDFLALIPIDQILAI AADYLANDAEVQA AV  343

263 AYLQ SDEFETIVVTLDALPELQNFLNFLEANGLNAIDFLNGIHDLGLGIPH  312
    ||.||||| |||.||| .|||:|:|:|:|:|:|:|
344 EYLSKDEFETIVVTVDSLPEFKNFLNFLTQTNGLNAIEFINNIHDLGLGIPH  393

313 IPVSGRKYHIRRGVITGLIDDVLA IAILPLDDLKALFNEKLETSDFLALY  362
    || .||| |:|:|:|:|:|:|:|:|:|:|:|:|:|:|
394 IPATGRK.HVRRGVGINGLIDDVIA IAILPVDELYALFQEKLESSPEFKALY  442

363 NAIKSPEFQSIGVQTLNAMPEYQNLLEKLEKREKGVVDVKI IELIRALFGLTH  412
    .|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
443 DAIRSPEFQSIGVQTLKAMPEYQDLIQRLKDKGVVDVDFH IELIKLFLGSLH  492

```

FIG. 1. Homology between Bla g 1.01 (top) and Bla g 1.02 (bottom) amino acid sequences deduced from the nucleotide sequences of cDNA clones. Bla g 1.01 clone 2 sequence (original clone 3) is indicated in bold.

*In Vivo Excision of pBluescript from Uni-ZAP<sup>TM</sup> XR and Insert Sequencing*—The pBluescript phagemids containing cloned cDNA inserts were excised from the Uni-ZAP<sup>TM</sup> XR vector and plated with fresh *Escherichia coli* cells to produce colonies (ExAssist/SOLR System, Stratagene). The size of the inserts was assessed by double digestion with *Bam*HI and *Kpn*I. Double-stranded cDNA was isolated and sequenced using an ABI Prism 100 Model 377 automated DNA sequencer (Biomolecular Research Facility, University of Virginia). The nucleotide sequences of the three Bla g 1 cDNA clones were submitted to the GenBank with accession numbers AF072219 (Bla g 1.01 clone 1), AF072221 (Bla g 1.01 clone 2), and AF072220 (Bla g 1.02).<sup>2</sup>

*Sequence Analysis*—Sequences were analyzed using the GCG sequence analysis package (Wisconsin package, Oxford Molecular Group, Inc.). Protein sequences were compared with the GenBank, EBI, DDBJ, PDB, SwissProt, SPupdate, PIR data bases using FASTA and BLAST (18, 19). Sequence alignments were made with Clustal W (accessible via Internet site: <http://www2.ebi.ac.uk/clustalw/>) and GCG (Framealign, Gap) (20). The GCG package was used to search for motifs in the amino acid sequences and for dot-plot analysis of nucleotide or amino acid sequences.

Transmembrane spanning domains in the amino acid sequences were predicted using the program TMPredict, which is based on a data base of membrane-spanning protein segments, TMbase (21). Prediction of protein localization sites was done using the program PSORT (22). Both programs are available at ([http://www.isrec.isb-sib.ch/software/TMPRED\\_form.html](http://www.isrec.isb-sib.ch/software/TMPRED_form.html) and <http://psort.nibb.ac.jp/>).

*Amino Acid Sequencing*—Amino-terminal amino acid sequences of purified Bla g 1 and Per a 1 (kindly provided by Dr. Schou, ALK-

ABELLO) were determined by Edman degradation using a gas phase sequencer (model 470-A, Applied Biosystems, Foster City, CA) in the Biomolecular Research Facility (University of Virginia). The first 10 and 47 N-terminal amino acids were obtained for Bla g 1 and Per a 1, respectively.

*Bla g 1 Measurements in Cockroach Tissues*—Cockroaches were dissected and body parts identified according to the method of Bell (23). Tissues were homogenized with a Polytron homogenizer and extracted overnight in 0.5 ml of BBS at 4 °C. After centrifugation at 13,200 × g for 15 min, extracts were stored at -20 °C until assayed. Bla g 1 levels in cockroach tissues were measured by mAb ELISA (17). Protein concentration was measured by Bradford assay using γ globulin as standard.

## RESULTS

*Molecular Cloning of Bla g 1*—Three cDNA clones encoding Bla g 1 were obtained by screening a *B. germanica* cDNA expression library with anti-Bla g 1 mAb, 10A6. The cDNA clones contained inserts of 1429, 1791, and 715 base pairs encoding 412, 492, and 188, respectively, amino acid fragments with calculated molecular masses of 45.8, 55.5, and 21.0 kDa, respectively, and isoelectric points of 4.31, 4.55, and 4.22, respectively. Bla g 1 clones 1 and 2 were isoallergens and showed 75% sequence identity (Fig. 1). Following the WHO/IUIS allergen nomenclature, the cDNAs were designated as Bla g 1.01 and Bla g 1.02. The third clone corresponded to a C-terminal fragment of clone 1, shown in bold in Fig. 1. IgE Ab reactivity to the three Bla g 1 clones, and to a recently isolated Per a 1 cDNA, was compared by plaque immunoassay (Fig. 2). The proportion of sera from nine patients allergic to Bla g 1 that recognized Bla g 1 clones 1, 2, and 3, and Per a 1 was 8/9 (89%), 9/9 (100%),

<sup>2</sup> These sequences have been scanned against the data base, and all sequences with significant relatedness to them identified and their accession numbers included.

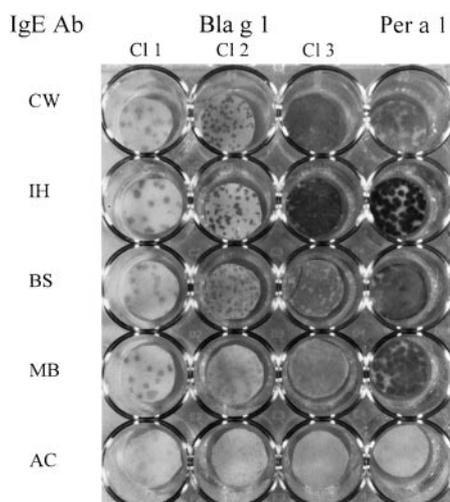


FIG. 2. Plaque immunoassay showing the reactivity of IgE Ab from patients allergic to Bla g 1 (first 4 rows) and to cockroach allergens but not to Bla g 1 (last row), against fusion proteins produced by the three Bla g 1 cDNA clones: Bla g 1.01, Bla g 1.02, and Bla g 1.01 clone 2 (first three columns), and by Per a 1 clone (last column).

7/9 (78%), and 8/9 (89%), respectively. Four sera from patients allergic to Bla g 2, but not to Bla g 1, and two sera from non-allergic patients were used as controls and gave negative results. Monoclonal antibodies against Bla g 1 (10A6), Bla g 2 (8F4), and Per a 1 (Per a 1-03) were also tested; 10A6 and Per a 1-03 reacted strongly with Bla g 1 and Per a 1, whereas 8F4 gave no reaction (data not shown).

**Sequence Analysis of Bla g 1 Clones, Homology with Mosquito ANG12 Precursor**—Initially, we compared the sequence similarity of the three Bla g 1 clones with the antigenically related allergen from *P. americana*, Per a 1 (16). The results showed 70–72% sequence identity between Per a 1, Bla g 1.01, and Bla g 1.02 (Table I). The Per a 1 and the Bla g 1 sequences showed 32% homology to mosquito (*Anopheles gambiae*) ANG12 precursor protein (Fig. 3). The ANG12 precursor is found exclusively in the midgut of female mosquito and is induced following a blood meal.

Further sequence comparisons revealed homology between the Bla g 1 sequences and other recently reported cockroach allergen sequences: Bla g Bd90K and four *P. americana* cDNA clones (24, 25) (Table I). A partial nucleotide sequence of Bla g Bd90K from *B. germanica* was previously published, and the full sequence is deposited in the GenBank. This long sequence (4058 base pairs) is untranslatable as a whole unit. However, Bla g 1.01 clearly aligns with the translatable part of Bla g Bd90K C terminus, including the sequence following the stop codon, and with other translated parts in different reading frames (Fig. 3). Translation of the N terminus in the third reading frame encodes for a methionine after the first five nucleotides, which presumably is the start of the open reading frame for Bla g Bd90K protein (Fig. 3).

The four *P. americana* clones described as Cr-PII (GenBank accession numbers U69261, U69260, U69957, and U78970) encode for sequences of 274, 446, 395, and 228 amino acids with calculated molecular masses of 31, 51, 45 and 26 kDa, respectively. These clones show 60–72% amino acid sequence homology to Bla g 1 (Table I). One of the four cDNA clones (Cr-PII3) encodes for a protein which N terminus aligns with the N termini of Bla g Bd90K and ANG12 precursor proteins (Fig. 3, top panel). Together, the results of these sequence comparisons show an extended family of *B. germanica* and *P. americana* allergen cDNAs encoding for proteins with deduced molecular

TABLE I  
Sequence homology of Bla g 1 cDNA clones with related cockroach allergens and ANG12 precursor

	Bla g 1.01	Bla g 1.01 (clone 2)	Bla g 1.02
Bla g 1/Bd90K (4058 base pairs)	94	95	74
Per a 1 (231 aa <sup>a</sup> )	72	70	70
Cr-PII (274 aa)	71	70	71
Cr-PII2 (446 aa)	60	64	60
Cr-PII3 (395 aa)	66	67	64
Cr-PII4 (228 aa)	72	69	70
ANG12precursor (211 aa)	32	35	32

<sup>a</sup> aa, amino acids.

masses of 20–90 kDa.

**The Bla g 1 and Per a 1 Allergens Are Composed of Multiple Amino Acid Repeats**—A unique feature of the primary structure of Bla g 1.01 was the existence of multiple amino acid repeats, which have not previously been found in allergen sequences. Close inspection of the sequences revealed segments that were repeated every ~100 amino acids and suggested that the sizes of the Bla g 1, Per a 1, and related cDNAs varied with the number of repeat sequences they contained. The repeats were compared using a dot-plot matrix analysis that plots a point file with the coordinates of the points in common between two sequences that are compared and a diagonal where two sequences are identical. Repeated fragments show as parallel lines above the diagonal. Dot-plot matrix analysis showed that there were four entire amino acid repeats in Bla g 1.01, five in Bla g 1.02, and two in Bla g 1.01 clone 2 and Per a 1 (Fig. 4). Each repeat is about 100 amino acids long. In Bla g 1.01, the first and third and the second and fourth repeats are identical (Fig. 5, panel A), and in Bla g 1.02, the first, third, and fifth and the second and fourth repeats are almost identical (Fig. 5, panel B).

The Bla g Bd90K sequence contained seven repeats of ~576 nucleotides (192 amino acids) (Fig. 4, and Ref. 24). A repeat in Bla g Bd90K corresponds to two different consecutive repeats of Bla g 1.01, which occur within Bla g 1.01 clone 2 (Fig. 5, panel C), and this clone encodes for the basic unit that constitutes the 90-kDa protein (Fig. 5, panel C). Dot-plot analysis for ANG12 precursor, a less closely related sequence, shows absence of the clear structure of 100 amino acids repeats. Remains of that structure are apparent as smaller repeated fragments (Fig. 4).

**Structural Motifs of Bla g 1**—The Bla g 1.01 and Bla g 1.02 sequences contained myristoylation, amidation, and phosphorylation sites, with no obvious N-linked glycosylation sites (Fig. 6). Myristoylation sites are localized at the beginning of each repeat. Bla g 1.01 also has a motif that is characteristic of mitochondrial energy transfer proteins (P-x-[DE]-x-[LIVAT]-[RK]-x-[LRH]-[LIVMFY]), which is present in Per a 1 and degenerates in Bla g 1.02, and the ANG12 precursor. Proteins belonging to this family are evolutionarily related membrane proteins with three repeated sequences, each containing two transmembrane domains. However, analysis of Bla g 1 using the program TMpredict showed that Bla g 1 is not a membrane protein. The program PSORT predicts that Bla g 1 and Bla g Bd90K proteins would be directed to the outside of the cell following the secretory pathway through the endoplasmic reticulum (ER). The N terminus of Bla g Bd90K contains an ER targeting signal (KLALIFLAFL) that would initiate transport across the ER (Fig. 6).

**Production and Cleavage of Bla g 1 in the Cockroach Digestive Tract**—Natural Bla g 1 and natural Per a 1 were partially sequenced by Edman degradation. The results showed that each natural allergen preparation contained three amino acid

Bla g Bd90K 3rf	<u>M</u> KLALIFLAFGLTTCGKS--IPSTRNLDDDLQDFLKLVPVDQIIATATDYIANDAENVQAAV	59
CR-PII3	<u>M</u> KLPIMILAVLGVAFGKS-LPN-RNLQDLDLDFLALLPVDEITAVMDYLANDAENVQBAV	58
ANG12 prec	<u>M</u> KIAAFVVAACLVATSAVSCAPTTRPLTDFDDFVGLLPLNDLLDLAMRYLLTDKVVQQL	60
	* * . : . : * * . . * * . * * * * : * * * * : * * * * : . . * * * * : .	
Bla g Bd90K 3rf	AYLQSDDEFETIVTVTDALPELQNFLEFLENGLNADFLNGIHDLG--IPHIPVSGRKY--	117
CR-PII3	AYLQGEFPHKIVFTVEGLQEFGNFVQFLEDHGLDAVGYINRLHSVFGWDPYVPSKRK--	116
ANG12	LYLQGEFSAVWDQFFELSAVRDLQYLEEAGVPAYESLNVVADFGLSPLKPTSVRSL	120
	***. : * * : . * . : : : * * * * : * * : * * : . : * * * * .	
Bla g 1.01	RNLQDDLQDFLALIPVDQIIAATDYLANDAENVQAAVAYLQSDDEFETIVTVTDALPELQN	93
Bla g 1.02	RGLPELDQDFLALIPIDQILAIADYLANDAENVQAAVAYLKSDEFETIVTVTDALPELQN	366
Bla g Bd90K 3rf	RNLDDDLQDFLKLVPVDQIIAATDYIANDAENVQAAVAYLQSDDEFETIVTVTDALPELQN	82
Per a 1	RNLQDDLNDLALIPDQILAIAMDYLANDAENVQELVAYLQSDDFHKIITTEIGLPEFAN	106
CR-PII	RNLQDDLNDLALIPDQILAIAMDYLANDAENVQELVAYLQSDDFHKIINTIEALPEFAN	149
ANG12 prec	RPLTDFDDFVGLLPLNDLLDLAMRYLLTDKVVQQLLQYQGEFSAVWDQFFELSAVRD	83
	* * * * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * :	
Bla g 1.01	FLNLEANGLNADFLNGIHDLGIPHIPVSGRKY--HIRRGVITGLIDDDVLAAILPLDD	151
Bla g 1.02	FLNFLQTNGLNAIEFNNIHDLGIPHIPATGRK---HVRGNGVINGLIDDDVLAAILPVDE	422
Bla g Bd90K 3rf	FLNLEANGLNADFLNGIHDLGIPHIPVSGRKY--HIRRGVITGLIDDDVLAAILPLDD	140
Per a 1	FYNFLKEHGLDVADFNEIHSIIGLPPFPVPSRR---HARRGVINGLIDDDVLAAILPVDE	163
CR-PII	FYNFLKEHGLDVANYINEIHSIIGLPPFPVPSRR---HARRGVINGLIDDDVLAAILPVDE	206
ANG12 prec	LLQYLEEAGVPAYESLNVVADFGLSPLKPTSVRSLAARTGGLNGLEALAMPAE	143
	: : * * : * . : * : . : * . : . : * * * * : * * * * : * * * * : * * * * :	
Bla g 1.01	LKALFNEKLETSPDFLALYNAIKSPEFQSIQVTLNAMPEYQNLLEKREKRGVDVVKIIEI	211
Bla g 1.02	LYALFQEKLESPDFKALYDAIRSPEFQSIQVTLKAMPEYQDLIQRKDRGVDVDFHIEL	492
Bla g Bd90K 3rf	LKALFNEKLETSPDFLALYNAIKSPEFQSIQVTLNDMPEYQNLQKREKRGVDVVKIIEI	200
Per a 1	LKALFQEKLETSPDFKALYDAIRSPEFQSIISTLNAMPEYQDLQNLDRGVDVDFHIEL	223
CR-PII	LKTLFQEKLETSPDFKALYDAIRSPEFQSIISTLNAMPEYQELLQNLDRGVDVDFHIEL	266
ANG12 prec	LEAMFEKMKSTEFKALFEKMQNFDHKLRLALYESSTEVQNMHKLSESLGVDVDFHIVEV	203
	* : : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * :	
Bla g 1.01	LNGIHDLGIPHIPVSGRKY--YHIRRG-VGITGLIDDDVLAAILPLDDLALFNEKLETSP	356
Bla g 1.02	INNIHDLGIPHIPATGRK---KHVRRG-VGINGLIDDDVLAAILPVDELYALFQEKLESP	436
Bla g Bd90K 1rf	RSYLEYLT---FQFSGXQ---YHNPQEELELLXSFDDVLAAILPLDDLKTFLFNEELETSP	59
Per a 1	LNETHSIIGLPPFPVPSR---RHARRG-VGINGLIDDDVLAAILPVDELYALFQEKLETSP	176
CR-PII	LNETHSIIGLPPFPVPSR---RHARRG-VGINGLIDDDVLAAILPVDELYALFQEKLETSP	219
ANG12 prec	LNVVADFGLSPLKPTSVRSLAARTG---GLNGLEALAMPAEAMPEKMKST	156
	. :	
Bla g 1.01	DFLALYNAIKSPEFQSIQVTLNAMPEYQNLLEKREKRGVDVVKIIEILRALFGLTH*	412
Bla g 1.02	EFKALYDAIRSPEFQSIQVTLKAMPEYQDLIQRKDRGVDVDFHIELIKKFLGSLH*	492
Bla g Bd90K 1rf	DFLALYNAIKSPEFQSIQVTLNAMPEYQNLLEKREKRGVDVVKIIEILRALFGLTH*	115
Per a 1	DFKALYDAIRSPEFQSIISTLNAMPEYQDLQNLDRGVDVDFHIELIRSLFGLP*	231
CR-PII	DFKALYDAIRSPEFQSIISTLNAMPEYQELLQNLDRGVDVDFHIELIRSWFGLP*	274
ANG12 prec	EFKALFERMQNFDHKLRLALYESSTEVQNMHKLSESLGVDVDFHIVEVLDKDFFGWN*	211
	: * * * * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * :	

**FIG. 3. Sequence alignments of cockroach allergens.** The N terminus third reading frame translation of Bla g Bd90K from *B. germanica* (GenBank accession number L47595) aligned with the N terminus of Cr-PII3 from *P. americana* (U69957), and with ANG12 precursor from *A. gambiae* (Z22925) (*upper panel*). The N termini for the three proteins are very similar, and methionines (in *bold* and *underlined*) clearly align. The two other panels show alignment of Bla g 1.01 and Blag g 1.02 amino acid sequences with Bla g Bd90K, Per a 1, Cr-PII (U69261), and ANG12 precursor. *Middle panel*, Bla g Bd90K third reading frame translation of the sequence near to the N terminus. *Lower panel*, alignments with the first reading frame translation of Bla g Bd90K C terminus.

sequences. The N-terminal residues of each of these sequences were preceded by arginine residues (deduced from the cDNA sequence), suggesting that the natural allergen had been cleaved by trypsin-like enzymes (Fig. 6). The most abundant of the natural Bla g 1 sequences was located in the first and third Bla g 1.01 repeats, and the other two minority sequences were located in the second and third repeats. The results of these experiments suggest that Bla g 1 and Per a 1 are secreted in the cockroach digestive tract and cleaved by trypsin-like enzymes before being excreted. To localize sites of Bla g 1 production, cockroach tissues were dissected and assayed for Bla g 1 content by ELISA. The results showed that Bla g 1 was predominantly found in digestive organs, with highest amounts in the *hindgut* (633 units/mg protein), followed by the *proventriculus* (623 units/mg protein), and *esophagus* (334 units/mg protein).

#### DISCUSSION

Previous studies established that Bla g 1 and Per a 1 are important allergens that elicit IgE Ab responses in ~50% of cockroach allergic patients. They are the only allergens identified to date that show antigenic cross-reactivity between *B. germanica* and *P. americana*. In keeping with this, the cDNA cloning studies showed extensive amino acid sequence homology between Bla g 1 and Per a 1 (70–72% sequence identity), confirming that these are structurally related proteins. The three Bla g 1 cDNA clones encode the sequences of two isoallergens, Bla g 1.01 and Bla g 1.02, with 75% amino acid sequence identity. Antigenic cross-reactivity between bacterially expressed Bla g 1 and Per a 1 clones was confirmed by plaque immunoassay; IgE antibodies against Bla g 1 and the monoclonal anti-Bla g 1 antibody reacted strongly with Per a 1 fu-

sion proteins, and, conversely, anti-Per a 1 mAb bound to Bla g 1 fusion proteins.

The novel and unique feature of the Bla g 1 (and Per a 1) allergen sequences is that they contain multiple tandem repeats of ~100 amino acid residues. This type of structure has not previously been described for any allergen but does occur in other proteins. An earlier report by Helm *et al.* described a long (4058 base pair) cockroach sequence with seven 576-nucleotide repeats (24). However, the full sequence was untranslatable, and the identity of the putative ~90 kDa protein encoded by the sequence could not be established. Our data clearly shows that Bla g 1, Per a 1, Bla g Bd90K, and the four recently described Cr-PII sequences (25), form a family of structurally and antigenically related cockroach allergens, which differ in the number of amino acid repeats they contain. Collectively, these sequences belong to the "Group 1" allergens produced by *B. germanica* and *P. americana*. It also seems likely that these allergens occur in other cockroach species.

Dot-plot analysis showed that two consecutive amino acid repeats from Bla g 1.01 corresponded to one of the Bla g Bd90K nucleotide repeats. Analysis of these repeat amino acid sequences suggests that they probably derived from an original domain of approximately 100 amino acids by duplication. Both repeats in the "duplex" would diverge in a way that they are recognized by dot-plot analysis as two amino acid repeats but the corresponding nucleotide sequence does not show any repeat. Despite their divergence during evolution, both repeats have conserved amino acids that may be important for biological function (26). The "duplex" would replicate up to seven times to form the long DNA with seven tandem repeats in the Bla g Bd90K nucleotide sequence. The related mosquito pro-

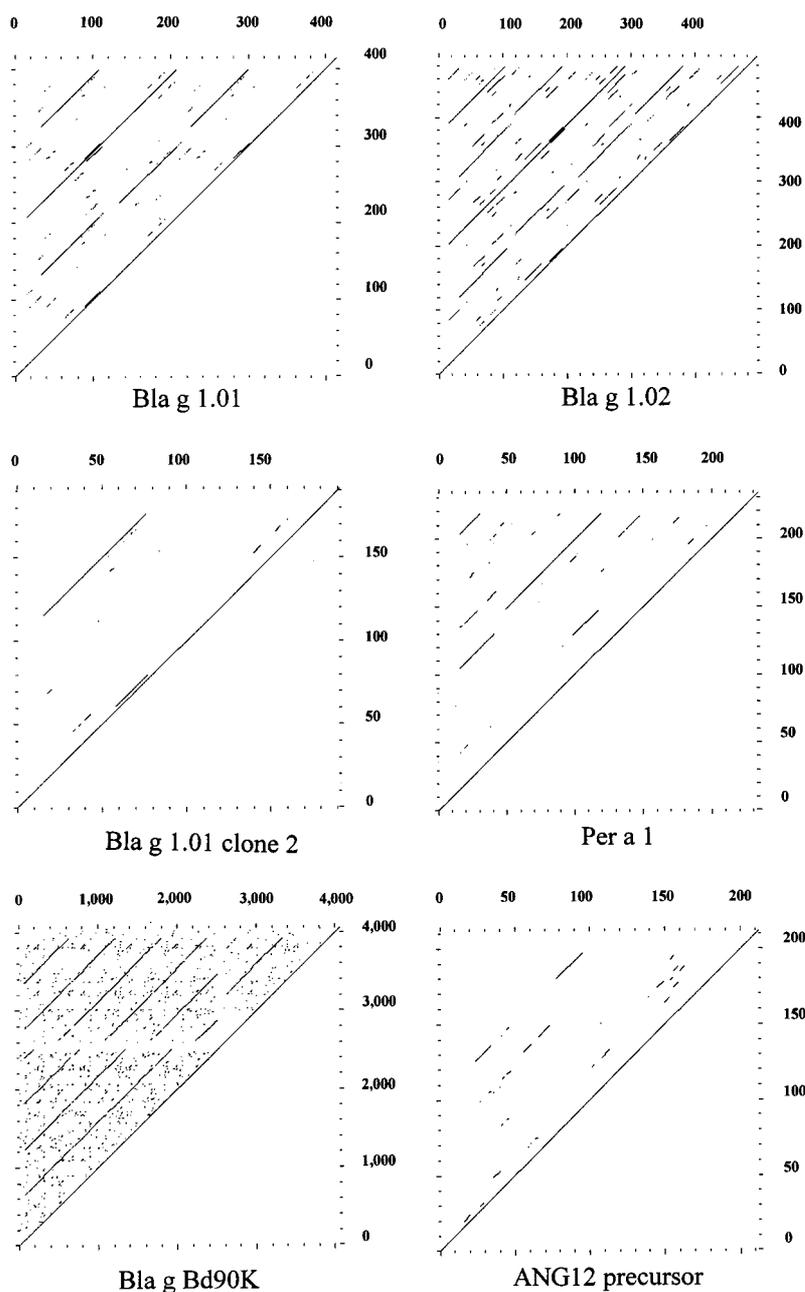


FIG. 4. Dot-plot matrix analysis of Bla g 1, Per a 1, and ANG12 precursor amino acid sequences and Bla g Bd90K nucleotide sequence.

tein ANG12 precursor has lost the clear structure of 100 amino acid repeats, indicating that ANG12 precursor had degenerated from the original structure during evolution. This would be consistent with the fact that mosquitoes appeared later in evolution than cockroaches in which the repeat allergen structure was conserved. The similarity between Bla g 1 and mitochondrial energy transfer proteins is interesting, because these proteins share amino acid motifs in a structure involving repeat sequences. This raises the possibility that the allergen may have evolved from a primordial mitochondrial sequence.

Several lines of evidence indicate that the Bla g Bd90K cDNA clone contains the open reading frame for Bla g Bd90K protein: (i) labeled Bla g Bd90K insert hybridized to a single-size mRNA of approximately the same size, 4kb (24); (ii) there is a methionine in the third reading frame translation of the Bla g Bd90K N terminus; (iii) a 90-kDa protein was reported in cockroach extracts (24); and (iv) the N terminus of the third reading frame translation of Bla g Bd90K aligned with the N termini of the related proteins Cr-PII3 and ANG12 precursor.

Despite the evidence for production of a 90-kDa protein, a broad range of molecular masses (6–37 kDa) has previously been reported for purified Bla g 1 and Per a 1 and for cross-reacting IgE Ab binding bands in cockroach extracts (13–15, 24). Natural Bla g 1 and Per a 1 appear to be produced following trypsin cleavage. A 90-kDa protein could be a precursor (like ANG12 precursor) that would be post-translationally modified and cleaved in certain sites such as myristoylation sites. The number of repeats contained in trypsin-cleaved fragments or in fragments derived from post-translational modifications would explain this broad range of molecular masses. If cockroach trypsin-like enzymes have the same characteristics as mosquito trypsin-like enzymes, they may have been co-purified with Bla g 1 and Per a 1 (27). Then, the trypsin cleavage of the natural allergen could take place not only in the digestive tract but also during and after purification. Plaque immunoassay with Bla g 1.01 clone 2 also showed that a 21-kDa unit, which contained two amino acid repeats, is sufficient to be recognized by Bla g 1- and Per a 1-specific IgE antibodies.

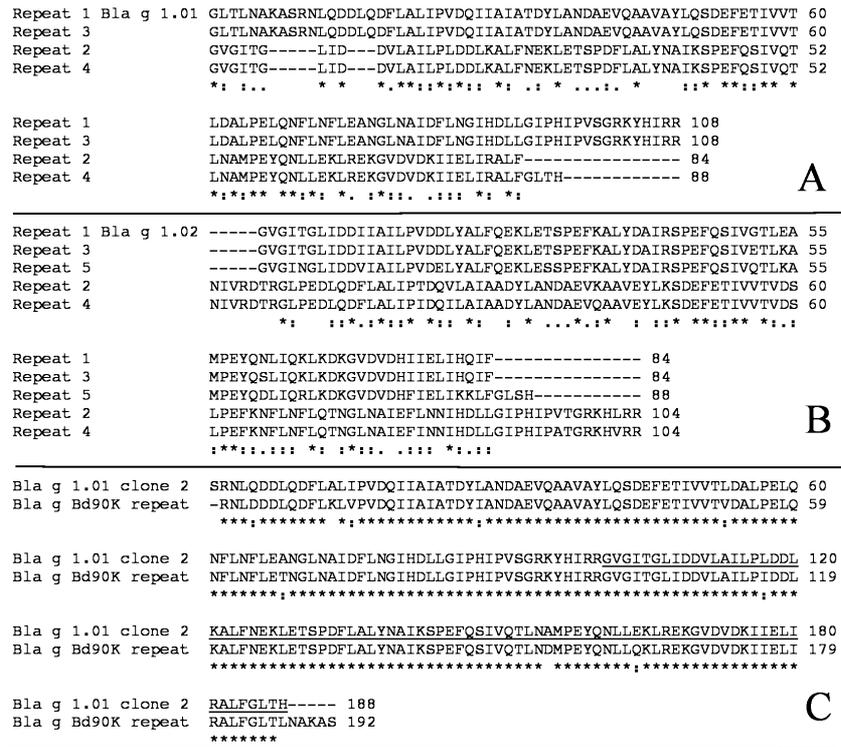


FIG. 5. Alignment of Bla g 1.01 (panel A) and Bla g 1.02 (panel B) entire repeats, and of Bla g 1.01 clone 2 with Bla g Bd90K translated repeat previously published by Helm et al. (panel C) (24).

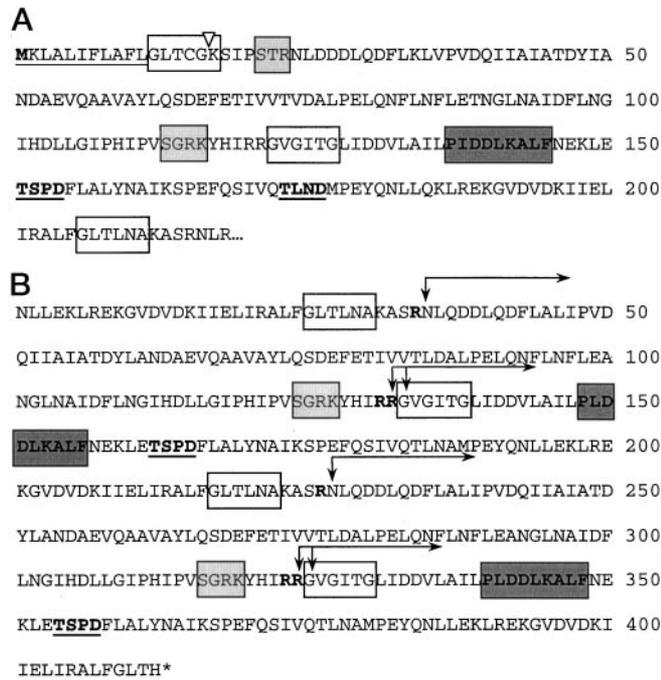


FIG. 6. Amino acid sequence motifs present in the third reading frame translation of Bla g Bd90K N terminus (panel A) and in Bla g 1.01 (panel B). 1) N-myristoylation sites (open box); 2) amidation sites (light gray boxes with four amino acids) and protein kinase C phosphorylation sites (first three amino acids in light gray boxes); 3) casein kinase II phosphorylation site (bold and underlined); and 4) mitochondrial energy transfer proteins motif (dark gray boxes). A McGleod cleavage prediction site is indicated by an inverted triangle, and the ER targeting signal is underlined after the methionine (panel A).

Therefore, cloning the “duplex” represents cloning the basic allergenic unit of Bla g 1.

Despite the presence in Bla g 1 of a signature motif for mitochondrial energy transfer membrane proteins, Bla g 1 is not a membrane protein. Rather, Bla g 1 is predicted to follow a secretory pathway through the endoplasmic reticulum, for

which it has a targeting signal, to the outside of the cell and into the digestive tract. Accordingly, high levels of Bla g 1 were measured by ELISA in the proventriculus and hindgut, in keeping with a previous report of localization of cockroach allergens in gastrointestinal cells using IgE Ab (28). From the digestive tract, Bla g 1 would be excreted into the feces, where it becomes a potential allergen if inhaled by humans.

The homology between Bla g 1 and mosquito ANG 12 suggests that Bla g 1 may have a digestive function. Mosquitos (*A. gambiae*) also produce trypsin-like proteins after taking blood meals (29–30). We speculate that similar enzymes may be produced in the cockroach digestive tract and give rise to the different molecular forms of the Bla g 1 and Per a 1 allergens. Proteases are found in commercial cockroach allergen extracts and can reduce allergenic potency when mixed with other extracts (31–32). The precise digestive function of Bla g 1 and ANG 12 remains to be determined.

In conclusion, the molecular cloning of Bla g 1 has revealed a new family of “Group 1” cockroach allergens. The novel feature of these allergens is the occurrence of multiple tandem amino acid repeats, which has not previously been described among allergen structures. We have recently expressed Per a 1 in the yeast, *Pichia pastoris*. Manipulation and overexpression of the sequences in appropriate vectors will facilitate structural and immunologic studies and the production of recombinant allergens for diagnostic and therapeutic purposes.

**Acknowledgments**—We are grateful to Drs. John Shannon and Jay Fox of the Biomolecular Research Facility, University of Virginia, for support with amino acid and nucleotide sequencing. We also thank Wanda Harvey for secretarial help and Tracy Glime for technical support.

REFERENCES

- Chapman, M. D. (1998) in *Encyclopedia of Immunology* (Roitt, I. M., and Delves, P. J., eds) 2nd Ed., Vol. 1, pp. 64–69, Academic Press, London
- Scheiner, O., and Kraft, D. (1995) *Allergy* **50**, 384–391
- Stewart, G. A., and Thompson, P. J. (1996) *Clin. Exp. Allergy* **25**, 1020–1044
- Hewitt, C. R. A., Brown, A. P., Hart, B. J., and Pritchard, D. I. (1995) *J. Exp. Med.* **182**, 1537–1544
- Schultz, O., Laing, P., Sewell, H. F., and Shakib, F. (1995) *Eur. J. Immunol.* **25**, 3191–3194
- Chapman, M. D., Smith, A. M., Vailes, L. D., and Arruda, L. K. (1997) *Allergy*

- 52, 374–379
7. Call, R. S., Smith, T. F., Morris, E., Chapman, M. D., and Platts-Mills, T. A. E. (1992) *J. Pediatr.* **121**, 862–866
  8. Rosenstreich, D. L., Eggleston, P., Kattan, M., Baker, D., Slavin, R. G., Gergen, P., Mitchell, H., McNiff-Mortimer, K., Lynn, H., Ownby, D., and Malveaux, F. (1997) *N. Engl. J. Med.* **336**, 1356–1363
  9. Arruda, L. K., Vailes, L. D., Mann, B. J., Shannon, J., Fox J. W., Vedvick, T. S., Hayden, M. L., and Chapman, M. D. (1995) *J. Biol. Chem.* **270**, 19563–19568
  10. Arruda, L. K., Vailes, L. D., Hayden, M. L., Benjamin, D. C., and Chapman, M. D. (1995) *J. Biol. Chem.* **270**, 31196–31201
  11. Arruda, L. K., Vailes, L. D., Platts-Mills, T. A. E., Hayden, M. L., and Chapman, M. D. (1997) *J. Biol. Chem.* **272**, 20907–20912
  12. Chapman, M. D., Vailes, L. D., Hayden, M. L., Platts-Mills, T. A. E., and Arruda, L. K. (1997) in *Allergy and Allergic Diseases* (Kay, A. B., ed) pp. 942–951, Blackwell Science Ltd., Oxford
  13. Schou, C., Lind, P., Fernandez-Caldas, E., Lockey, R. F., and Löwenstein, H. (1990) *J. Allergy Clin. Immunol.* **86**, 935–946
  14. Pollart, S. M., Mullins, D. E., Vailes, L. D., Hayden, M. L., Platts-Mills, T. A. E., Sutherland, W. M., and Chapman, M. D. (1991) *J. Allergy Clin. Immunol.* **87**, 511–521
  15. Wu, C.-H., Luo, S.-F., and Wong, D. W. (1997) *Allergy* **52**, 411–416
  16. Melén, E., Vailes, L. D., Pomés, A., Arruda, L. K., and Chapman, M. D. (1998) *J. Allergy Clin. Immunol.* **101**, 156
  17. Pollart, S. M., Smith, T. F., Morris, E. C., Gelber, L. E., Platts-Mills, T. A. E., and Chapman, M. D. (1991) *J. Allergy Clin. Immunol.* **87**, 505–510
  18. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
  19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
  20. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
  21. Hofmann, K., and Stoffel, W. (1993) *Biol. Chem. Hoppe-Seyler* **374**, 166
  22. Nakai, K., and Kanehisa, M. (1992) *Genomics* **14**, 897–911
  23. Bell, W. J. (1981) in *The Laboratory Cockroach: Experiments in Cockroach Anatomy, Physiology, and Behavior* (Chapman and Hall, eds) 1st Ed., pp. 33–38, University of Kansas, Lawrence, KS
  24. Helm, R., Cockrell, G., Stanley, J. S., Brenner, R. J., Burks, W., and Bannon, G. A. (1996) *J. Allergy Clin. Immunol.* **98**, 172–180
  25. Wu, C.-H., Wang, N. M., Lee, M.-F., Kao, C.-Y. Y., and Luo, S.-F. (1998) *J. Allergy Clin. Immunol.* **101**, 832–840
  26. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) *Science* **247**, 1306–1310
  27. Graf, R., Boehlen P., and Briegel H. (1991) *Experientia* **47**, 603–609
  28. Zwick, H., Popp, W., Sertl, K., Rauscher, H., and Wanke, T. (1991) *J. Allergy Clin. Immunol.* **87**, 626–630
  29. Müller, H.-M., Catteruccia, F., Vizioli, J., della Torre, A., and Crisanti, A. (1995) *Exp. Parasitol.* **81**, 371–385
  30. Müller, H.-M., Crampton, J. M., della Torre, A., Sinden, R., and Crisanti, A. (1993) *EMBO J.* **12**, 2891–2900
  31. Esch, R. E. (1991) in *Regulatory Control and Standardization of Allergenic Extracts* (Klein, R., ed) pp. 171–179, Gustav Fischer Verlag, Stuttgart, Germany
  32. Nelson, H. S., Iklé, J. D., and Buchmeier, A. (1996) *J. Allergy Clin. Immunol.* **98**, 382–388