THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 33, Issue of August 18, pp. 19563–19568, 1995
Printed in U.S.A.

Molecular Cloning of a Major Cockroach (*Blattella germanica*) Allergen, Bla g 2

SEQUENCE HOMOLOGY TO THE ASPARTIC PROTEASES*

(Received for publication, April 13, 1995, and in revised form, June 13, 1995)

L. Karla Arruda‡§, Lisa D. Vailes‡, Barbara J. Mann‡, John Shannon¶, Jay W. Fox¶, Thomas S. Vedvick‖, Mary L. Hayden‡, and Martin D. Chapman‡

From the Asthma and Allergic Diseases Center, Departments of ‡Internal Medicine and ¶Microbiology, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908 and ‖Sibia, Inc., La Jolla, California 92037

Inhalation of allergens produced by the German cockroach (*Blattella germanica*) elicits IgE antibody formation and the development of asthma in genetically predisposed individuals. We compared the allergenic importance of two cockroach (CR) allergens, Bla g 1 and Bla g 2, and determined the complete amino acid sequence of the major 36-kDa allergen, Bla g 2. A survey of 106 sera from CR allergic patients showed the prevalence of IgE antibodies to Bla g 1 and Bla g 2 to be 30.2% and 57.6%, respectively. Immediate skin tests on 7 selected patients gave positive reactions using 10^{-3} μg/ml either allergen, whereas controls showed no response to 10 μg/ml. Natural Bla g 2 was purified and the sequence of the NH_2 terminus and tryptic peptides, comprising 36% of the molecule, was determined. The cDNA for Bla g 2 was cloned from a *B. germanica* expression library and encoded a 24-amino acid signal peptide and a 328-amino acid mature protein, which showed sequence homology to aspartic proteases. Bla g 2 showed the highest degree of identity to mosquito (*Aedes aegypti*) lysosomal aspartic protease (30.8%), with similar identity to pepsin, cathepsins D and E, renin, and chymosin. Bla g 2 mRNA and protein were detected in *B. germanica*, but not in *Periplaneta americana*, the other principal domiciliary CR species in the U. S. High concentrations of Bla g 2 were found in CR digestive organs (esophagus, gut, and proventriculus). The results show that Bla g 2 is a major species-specific allergen of *B. germanica* and suggest that the allergen functions as a digestive enzyme in the cockroach.

Immediate hypersensitivity reactions to environmental allergens (*e.g.* pollens, dust mites, animal danders) occur in ~20% of Western populations and are a characteristic feature of common allergic illnesses, principally allergic rhinitis, asthma, and atopic dermatitis. These reactions are mediated by the production of IgE antibodies (Ab)[1] to low molecular mass (5–50 kDa) proteins or glycoproteins, with diverse structures and biologic functions, present in pollen grains, mite feces, animal hair, etc. Over the past few years, the application of molecular cloning techniques has elucidated the primary structures of allergens from these sources and, in many cases, this information has established their biologic function and allowed epitopes involved in T cell regulation of IgE Ab synthesis to be defined (1–10). These advances have led to the introduction of recombinant allergens and allergen peptides for improved allergy diagnosis and for the development of new forms of allergen-specific immunotherapy (11–13).

Infestation of houses with cockroaches (CR) results in the accumulation of high levels of potent allergens, which sensitize atopic individuals and induce the development of IgE Ab responses and asthma (14–18). This problem is particularly acute in the United States, where in some towns and cities up to 60% of patients with asthma are allergic to CR (15–20). Epidemiologic studies have shown that sensitization to CR allergens is an important risk factor for admission to hospital emergency rooms with asthma (21, 22). Indeed, asthma is the only disease that is consistently associated with CR-infested housing. The principal domiciliary CR species found in the U. S. are *Blattella germanica* (German CR) and *Periplaneta americana* (American CR). The molecular structure and biologic functions of allergens produced by either species are poorly understood. Moreover, in spite of the widespread use of CR in biology and in biomedical research, there is limited structural data on CR proteins.

Previous serologic studies, using IgE Ab and murine monoclonal antibodies (mAb), have identified allergens from both *B. germanica* (Bla g 1 and Bla g 2), and *P. americana* (Per a 1 and Per a 3) (16, 23–28). We have used molecular cloning techniques to determine the primary structures of allergens from *B. germanica* (the most common cause of allergic sensitization in the U. S.). We recently described the structure of a cockroach calycin allergen, termed Bla g 4 (29). Here, we report the complete nucleotide and amino acid sequence of the major *B. germanica* allergen, Bla g 2, and show that this allergen shares homology with the aspartic protease family of enzymes. The allergen is concentrated in organs of the digestive tract, suggesting that it may function as a digestive enzyme.

* This work was supported by National Institutes of Health Grants AI 32557 and AI 34607 and an Underrepresented Minority Investigator in Asthma and Allergy Award (to Dr. L. K. A.), sponsored by the National Institute of Allergy and Infectious Disease and by the American Academy of Allergy and Immunology. 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™/EMBL Data Bank with accession number(s) U28863.

§ To whom correspondence and reprint requests should be addressed: Box 225, Asthma and Allergic Diseases Center, University of Virginia Health Sciences Center, Charlottesville, VA 22908. Tel.: 804-982-3324; Fax: 804-924-5779.

[1] The abbreviations used are: Ab, antibody; mAb, monoclonal antibody; BBS, borate-buffered saline; CR, cockroach(es); PVDF, polyvinylidene difluoride; RAST, radioallergosorbent test; HPLC, high performance liquid chromatography; PAGE, polyacrylamide gel electrophoresis; ELISA, enzyme-linked immunosorbent assay; RIA, radioimmunoassay; kb, kilobase pair(s).

EXPERIMENTAL PROCEDURES

Purification of CR Allergens—B. germanica frass (feces, secretions, egg cases, and body parts) was extracted in borate-buffered saline, pH 8.0 (BBS), overnight at 4 °C. After centrifugation at 18,000rpm, the supernatant was dialyzed against BBS and ether-extracted. Bla g 1 was purified from frass extract by preparative isoelectric focusing and size exclusion HPLC (28). Bla g 2 was purified by affinity chromatography over mAb 8F4 immunosorbent (28), followed by elution from a C18 reverse phase HPLC column (Brownlee Labs, Santa Clara, CA) using a 0–80% gradient of acetonitrile in 0.1% trifluoroacetic acid. Alternatively, mAb affinity-purified Bla g 2 was further purified by electroelution from SDS-PAGE gels using a micro-electroelutor (Centrilutor, Amicon, Beverly, MA) according to the method of LeGendre and Matsudaira (30). Fifty μg of affinity-purified allergen were applied to 8 lanes of a 12% SDS-PAGE gel, and the 36-kDa bands were excised, electroeluted for 2 h, and concentrated by centrifugation. Purity was assessed by SDS-PAGE using either an 8–25% silver-stained gel on a PhastSystem (Pharmacia Biotech Inc.) or a Coomassie Blue-stained 12% gel.

Amino Acid Sequencing—Amino-terminal amino acid sequences of HPLC-purified allergens were determined by Edman degradation using a gas phase sequencer (model 470-A, Applied Biosystems, Foster City, CA). Seven tryptic peptides (7–22 residues) of electroeluted Bla g 2, comprising 84 amino acid residues, were sequenced. Affinity-purified Bla g 2 was also separated on a 12% SDS-PAGE gel and transferred to polyvinylidene difluoride (PVDF) membrane (Immobilon-P, Millipore, Bedford, MA). The NH_2-terminal amino acid residues of the 36- and 70-kDa bands were sequenced off the membrane (14 and 10 residues, respectively).

Molecular Cloning and Sequencing of Bla g 2 cDNA—Total RNA was extracted from adult *B. germanica* or *P. americana* of mixed sexes, with 5 M guanidinium thiocyanate using the method of Chirgwin *et al.* (31). Messenger RNA was obtained using a FastTrack kit (Invitrogen, San Diego, CA). A *B. germanica* cDNA library was prepared from 10 μg of mRNA in the UniZAP-XR expression vector (Stratagene, La Jolla, CA). The library was screened using polyclonal mouse anti-Bla g 2 antiserum, with an IgG Ab titer >100,000 as determined by ELISA. Recombinant plaques were grown on NZY agar, and protein expression was induced using nitrocellulose filters soaked in 10 mM isopropyl-1-thio-βD-galactopyranoside. Filters were incubated in blocking solution (1% dried milk, 0.2% bovine serum albumin, 0.4% goat serum, 0.03% gelatin), followed by a 1:5,000 dilution mouse anti-Bla g 2 antiserum, pre-absorbed with *Escherichia coli* lysate. Positive plaques were identified using 1:2,000 dilution alkaline phosphatase labeled anti-mouse IgG and 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium substrate (KPL, Gaithersburg, MD). A single positive plaque, containing a 1,317-base pair cDNA insert coding for Bla g 2, was isolated after screening 296,000 plaque-forming units of *B. germanica* cDNA library. Both strands of the Bla g 2 cDNA were sequenced by dideoxynucleotide chain termination using Sequenase (U. S. Biochemical Corp.) (32).

Sequence Analysis—Protein or peptide sequences were compared with the National Biomedical Research Foundation, Swiss-Prot, and GenBank data bases using FASTA, and sequence alignments were carried out using the GCG program (33).

IgE Antibodies and Immediate Hypersensitivity Skin Testing—Sera were obtained from 93 CR allergic patients who either had been recruited from the University of Virginia Allergy Clinics or had previously presented to hospital emergency rooms with asthma (21, 22, 34). An additional 13 sera were collected from CR allergic patients living in New York or Puerto Rico. All patients had asthma, and the majority were adults (>16 years old). Eight patients were children (aged 4–10). All patients had serum IgE Ab to CR allergens detectable by radioallergosorbent test (RAST) (CR RAST > 40 units/ml; 1 unit = ~0.1 ng of IgE) (34). Sera from 18 non-allergic individuals were used as controls.

Quantitative intradermal skin tests were carried out using serial 10-fold dilutions of *B. germanica* extract (1/20, w/v; Allergy Laboratories of Ohio, Columbus, OH), or purified Bla g 1 or Bla g 2, from 10 to 10^{-6} μg/ml, as described previously (28). Skin testing, and collection of sera for use in these studies, was approved by the Human Investigation Committee of the University of Virginia.

Immunoassays for IgE Antibodies to CR Allergens—Serum IgE Ab to Bla g 1 and Bla g 2 were measured using a mAb-based solid-phase RIA (28). Briefly, 1 μg of mAb 10A6 (anti-Bla g 1) or 8F4 (anti-Bla g 2) was coated onto plastic microtiter wells, followed by successive incubations with 0.5 μg of *B. germanica* frass extract and human sera, diluted 1:2 and 1:10. Bound IgE Ab were detected using 2 ng of ^{125}I-labeled anti-human IgE. The assays were quantitated using sera from two CR allergic patients (B. S. and M. A.) with high levels of IgE Ab to Bla g 1 or Bla g 2, respectively, to construct control curves. Each serum was arbitrarily assigned a value of 1,000 units/ml of IgE Ab.

Inhibition RIA for Comparing Bla g 2 Antigen Expression in CR Species—Twenty μg of electroeluted Bla g 2 were radiolabeled with 1 mCi ^{125}I using the chloramine T technique (specific activity 18.5 μCi/μg) (35). Serial doubling dilutions of Bla g 2, *B. germanica* frass extract, or *B. germanica* commercial extract (Greer Laboratories, Lenoir, NC) were used to inhibit the binding of ^{125}I-Bla g 2 to mouse IgG anti-Bla g 2 Ab. Allergens were incubated for 2 h with 0.1 ml of a 1:4,000 dilution of mouse anti-Bla g 2 antiserum, followed by 3 ng of ^{125}I-Bla g 2 for 2 h, and precipitated overnight at 4 °C with 1:15 dilution goat anti-mouse IgG (Chemicon, El Segundo, CA). Precipitates were washed with BBS and counted in a γ counter. Expression of Bla g 2 was compared in *P. americana* extracts. Whole body extract was prepared by homogenizing 133 g of CR with BBS in a blender, extracting at 1:5 (w/v) overnight at 4 °C, and decanting the supernatant after centrifugation (6,000 rpm × 30 min). Frass extract was prepared as for *B. germanica*. Commercial *P. americana* extracts (n = 11) were obtained from 10 United States allergen manufacturers: Allergy Laboratories of Oklahoma (Oklahoma City, OK), Center Laboratories (Port Washington, NY), Greer Laboratories (Lenoir, NC), Miles Laboratories (Elkhart, IN), Antigen Labs (Liberty, MO), Nelco (Deer Park, NY), ALK/Berkeley (Milford, CO), Bencard (Bencard, MS), Meridian (Round Rock, TX), and Iatric (Tempe, AZ).

Northern Analysis of the Expression of Bla g 2 mRNA—For Northern blots, *B. germanica* and *P. americana* mRNA was isolated from total RNA using the Poly(A)Tract mRNA isolation system (Promega, Madison, WI). Samples containing 0.5–6 μg of mRNA were electrophoresed in formaldehyde denaturing 1% agarose gels, followed by transfer to a Zetabind nylon membrane (Cuno, Meridien, CT). A 1.3-kb *SmaI/KpnI* restriction fragment from a pBluescript phagemid DNA comprising the complete Bla g 2 cDNA, and an 8.6-kb *BamHI* fragment containing *Neurospora crassa* ribosomal DNA (pRW528) were labeled with [α-^{32}P]dCTP by random priming and used to probe the blots (36). Hybridization was carried out at 37 °C, as described previously (6).

Bla g 2 Measurements in CR Tissues—To prepare tissue extracts, CR were dissected and body parts identified according to the method of Bell (37). Tissues were homogenized with a Polytron homogenizer and extracted overnight in 0.5 ml of BBS at 4 °C. After centrifugation at 12,000 rpm for 15 min, extracts were stored at −20 °C until assayed. Bla g 2 levels in CR tissues were measured by mAb ELISA (38).

RESULTS

Allergenic Importance of Bla g 1 and Bla g 2—Sera from 106 CR allergic patients with asthma living in different geographic areas in the United States were analyzed for IgE Ab to Bla g 1 and Bla g 2 by mAb-based RIA. The results showed that 30.2% and 57.6% of these patients had detectable IgE Ab to Bla g 1 or Bla g 2, respectively. The prevalence of IgE Ab to Bla g 1 and Bla g 2 was higher among patients with IgE Ab to CR of >200 RAST units/ml (40% and 70%, respectively, n = 58), as compared to the group with <200 RAST units/ml (19% and 42%, respectively, n = 48) (Fig. 1). The biologic activity of purified Bla g 1 and Bla g 2 was assessed by quantitative intradermal skin testing of seven selected CR allergic patients. These patients gave positive immediate skin reactions (>8 × 8-mm wheal) to allergen concentrations down to $10^{-3.5}$ μg/ml, whereas non-allergic controls showed no reaction to concentrations up to 10 μg/ml (and had no detectable serum IgE Ab) (Table I). Skin test reactivity correlated with serum IgE Ab to Bla g 1 and Bla g 2, and distinct patterns of reactivity to the allergens were observed. Some patients had positive skin tests and IgE Ab to both Bla g 1 and Bla g 2 (S. W. and M. B), whereas others were exclusively sensitive to either allergen (compare B. S. and R. M.; Table I).

Complete Nucleotide and Amino Acid Sequence of Bla g 2—We focused on sequencing Bla g 2, because of its allergenic importance and because this protein has been used extensively as a marker of environmental CR allergen exposure (21, 22, 38). The Bla g 2 sequence was determined using a combination of protein sequencing and cDNA cloning. Previous studies showed that single-step mAb affinity chromatography yielded 36-kDa Bla g 2, as well as additional bands of 20 and 70 kDa,

which co-purified with the allergen (28). For amino acid sequencing, Bla g 2 was further purified by reverse-phase HPLC or electroelution from SDS-PAGE gels. Three protein peaks were obtained on reverse phase HPLC, which separated Bla g 2 from the 20-kDa contaminant (Fig. 2, *panel A*). The NH_2-terminal amino acid sequence of Bla g 2 (35 residues) was determined from HPLC peak 2. This sequence was subsequently confirmed by sequencing Bla g 2 (14 residues) that had been electroeluted onto PVDF membrane (Figs. 2*B* and 3). The NH_2-terminal sequence of the 20-kDa protein did not share significant homology to Bla g 2 or other proteins, suggesting that it was not a breakdown product of Bla g 2 (data not shown). Tryptic peptides were prepared from electroeluted Bla g 2, and internal peptide sequences (comprising 84 residues) were obtained, which, together with the NH_2-terminal sequence, comprised 36.3% of the entire molecule. In addition, the first 10 residues of the 70-kDa protein sequenced off PVDF membrane were identical to the NH_2-terminal sequence of Bla g 2, suggesting that the 70-kDa band was a dimer of Bla g 2.

Although ~60% of CR allergic patients had IgE Ab to Bla g 2, attempts to screen the *B. germanica* cDNA library with pooled IgE Ab, to identify a Bla g 2 cDNA clone, were unsuccessful. The cDNA coding for Bla g 2 was identified using mouse polyclonal IgG anti-Bla g 2 Ab. The full-length cDNA contained an open reading frame of 1,056 nucleotides, encoding a 24-amino acid putative signal peptide and a 328-amino acid protein, with a predicted molecular mass of 35,939 Da (Fig. 3). Inspection of the nucleotide sequence identified a polyadenylation signal 22 nucleotides upstream from the poly(A) tail and three potential *N*-linked glycosylation sites. However, the close agreement between the molecular mass obtained by sequencing and by SDS-PAGE analysis suggests that the allergen is not glycosylated. The deduced amino acid sequence of Bla g 2 showed 91% identity to the amino acid sequences determined by Edman degradation from Bla g 2 protein (Fig. 3).

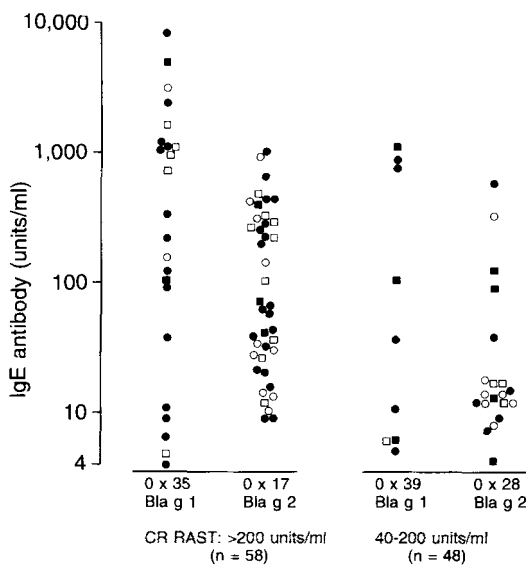

FIG. 1. **IgE antibodies to Bla g 1 and Bla g 2 in sera from CR allergic patients with asthma.** Sera obtained from 106 patients living in Charlottesville, VA (●), Wilmington, DE (○), Atlanta, GA (□), or New York or Puerto Rico (■) were compared by mAb-based RIA. Results are expressed as IgE Ab units/ml relative to control curves constructed using sera containing high levels of IgE Ab. Data from selected patients and 18 non-allergic controls, expressed as counts/min ^{125}I anti-IgE bound, are shown in Table I.

Homology of Bla g 2 to Aspartic Proteases—Sequence similarity searches showed significant homology between Bla g 2 and aspartic proteases from several species (39–43). The highest degree of identity was between Bla g 2 and mosquito (*A. aegypti*) lysosomal aspartic protease (30.8%); however, identity was observed with other aspartic proteases including human cathepsin D (27%), human pepsinogen (26.7%), human renin (25.8%), and bovine chymosin (29.5%) (Fig. 4*A*). The two aspartic acid residues involved in the catalytic activity of aspartic proteases (at positions 31 and 215 in Bla g 2), as well as surrounding amino acid residues, are conserved in Bla g 2 (Fig. 4, *A* and *B*). Alignments of these sequences revealed that Bla g 2 is likely to be synthesized as a pro-enzyme, but the propeptide sequence is much shorter than in other aspartic proteases (Fig. 4*A*).

Expression of Bla g 2 in CR Species—To investigate whether Bla g 2 was also expressed by *P. americana*, we compared antigenic cross-reactivity and expression of Bla g 2 mRNA in the two CR species. Binding of ^{125}I-Bla g 2 to polyclonal mouse IgG anti-Bla g 2 was inhibited >90% by Bla g 2 and by *B. germanica* extracts in a dose-dependent manner. In contrast, *P. americana* extracts, including frass and whole body extracts prepared in our laboratory and 11 commercial extracts marketed for allergy diagnosis, gave <5% inhibition (Fig. 5).

It was possible that the murine IgG antibodies were directed against "species-specific " epitopes on Bla g 2 and did not recognize the allergen in *P. americana* extracts. To investigate this possibility, mRNA expression was compared by Northern analysis. A 1.8-kb mRNA encoding Bla g 2 was detected by hybridization to a ^{32}P-labeled Bla g 2 cDNA probe. This message was detected using 0.5–2 μg of *B. germanica* mRNA, whereas no Bla g 2 message was detected using up to 6 μg of mRNA from *P. americana* (Fig. 6). These results strongly suggest that *P. americana* does not express a protein that is closely homologous to Bla g 2.

Detection of Bla g 2 in Tissues—The tissue distribution of Bla g 2 was assessed by comparing the ratios of Bla g 2 to total

TABLE I
Immediate skin tests and serum IgE antibodies to B. germanica allergens

Patient	Skin test to *B. germanica*[a]	Cockroach RAST[b]	Bla g 1		Bla g 2	
			Skin test	Serum IgE Ab[c]	Skin test	Serum IgE Ab[c]
		units/ml	*μg/ml*	*cpm bound*	*μg/ml*	*cpm bound*
B. S.	10^{-6}	1,075	$10^{-3.5}$	35,000	10^{0}	997
R. M.	10^{-6}	1,065	10^{0}	148	10^{-3}	13,976
S. W.	10^{-6}	1,290	$10^{-3.5}$	25,092	$10^{-2.5}$	16,192
M. B.	$10^{-6.5}$	1,465	$10^{-3.5}$	21,104	$10^{-2.5}$	12,961
S. T.	10^{-6}	1,080	>10	118	10^{-1}	8,251
J. S.	10^{-6}	1,395	10^{-1}	183	10^{-3}	17,776
R. Mo.	10^{-6}	300	10^{-2}	96	10	116
Controls (n = 4)	$>10^{-2}$	<10	>10	128 ± 61[d]	>10	153 ± 52[d]

[a] Using *B. germanica* extract from Allergy Labs of Ohio, Inc., Columbus, OH.
[b] One RAST unit = 0.1 ng of IgE antibody (34).
[c] Measured by mAb-based RIA.
[d] Mean ± S.D. of 18 non-allergic controls.

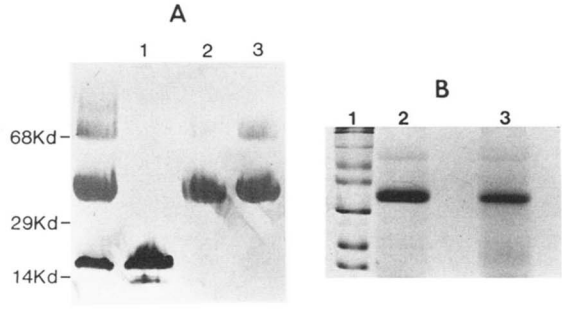

FIG. 2. **Purification of Bla g 2 protein for amino acid sequencing.** *Panel A*, silver-stained SDS-PAGE of single-step affinity-purified Bla g 2 (*left lane*) and HPLC peaks 1–3. The 20-kDa protein, which co-eluted with Bla g 2 from the mAb column, eluted as a single peak (*Peak 1*) from the C18 column. *Peaks 2* and *3* contained 36-kDa Bla g 2 and a 70-kDa band. *Panel B*, Coomassie Blue-stained SDS-PAGE of two preparations of electroeluted Bla g 2 (*lanes 2* and *3*). *Lane 1*, molecular mass markers.

protein in dissected *B. germanica* tissues and body parts. Although Bla g 2 was present in all tissues, the highest concentrations relative to protein content were detected in organs of the digestive system, principally esophagus, proventriculus, and gut (Table II).

DISCUSSION

We report the complete nucleotide sequence of cDNA coding for a major cockroach allergen, Bla g 2. The allergenic importance of Bla g 2 had been suggested by previous studies showing IgE binding to a 36-kDa *B. germanica* allergen on immunoblotting (24). In the present study, an extensive survey of sera from several localities in the United States confirmed a high prevalence of IgE Ab to Bla g 2 among CR allergic patients (~60%). The results also showed that the allergen exhibited classical immediate hypersensitivity responses on skin testing, and that these responses were specific. Although the prevalence of IgE Ab to Bla g 1 (~30%) was significantly lower than to Bla g 2, some patients showed comparable skin test reactivity to the two allergens, or to Bla g 1 in the absence of responses to Bla g 2, suggesting that Bla g 1 can be an important allergen for some individuals.

Sequence analysis revealed that Bla g 2 shares homology to the aspartic proteases: a widespread group of enzymes that have two essential aspartic acid residues at their catalytic site (44). Most aspartic proteases are single-chain enzymes with a molecular mass of ~35,000 Da and are active at low pH. The group includes both intracellular enzymes such as cathepsin D (40) and cathepsin E (45), and extracellular digestive enzymes, such as pepsin (42) and chymosin (43), and human renin (41). Assignment of Bla g 2 to the aspartic proteases was based on sequence homology and presence of the aspartic acid and adjacent amino acid residues in conserved positions. Several aspartic proteases are secreted as pro-enzymes and undergo self-activation upon exposure to acidic pH. Upon activation, an NH_2-terminal pro-peptide of up to 50 amino acids long is released. Sequence alignment showed homology between the NH_2-terminal sequence of Bla g 2 and the pro-peptide sequences, particularly bovine chymosin, and suggested that Bla g 2 may be produced as a zymogen with a short pro-peptide. Pro-peptide sequences as short as 6 amino acid residues occur in albumin and trypsin (46). In keeping with these observations, purification of natural Bla g 2 has always been performed at high or neutral pH (28). Further studies are being carried out to determine the effects of low pH treatment on the stability of Bla g 2 and on release of the pro-peptide.

The homology of Bla g 2 to aspartic proteases raised the possibility that this allergen functions as a digestive enzyme in

```
1    AA ATG ATT GGC CTA AAG CTA GTG ACA GTT CTC TTT GCG GTT GCT      44
        M  I  G  L  K  L  V  T  V  L  F  A  V  A

45      ACC ATA ACA CAT GCA GCT GAG CTT CAA CGT GTT CCA TTG TAC      86
        T  I  T  H  A  A  E  L  Q  R  V  P  L  Y

87      AAA TTG GTG CAC GTT TTC ATT AAC ACT CAA TAC GCT GGT ATA     128
        K  L  V  H  V  F  I  N  T  Q  Y  A  G  I

12      ACC AAG ATT GGA AAC CAG AAC TTC CTA ACA GTA TTC GAT AGC     170
        T  K  I  G  N  Q  N  F  L  T  V  F  D  S

171     ACC TCA TGC AAT GTA GTC GTT GCC AGT CAA GAA TGC GTT GGT     212
        T  S  C  N  V  V  V  A  S  Q  E  C  V  G

213     GGA GCT TGT GTA TGT CCA AAT CTA CAA AAA TAT GAG AAA CTT     254
        G  A  C  V  C  P  N  L  Q  K  Y  E  K  L

255     AAA CCG AAG TAT ATC TCT GAT GGG AAT GTA CAG GTG AAA TTC     296
        K  P  K  Y  I  S  D  G  N  V  Q  V  K  F

297     TTC GAC ACT GGT AGC GCA GTT GGT AGA GGC ATT GAA GAT TCC     338
        F  D  T  G  S  A  V  G  R  G  I  E  D  S

339     CTT ACG ATT TCT AAC CTC ACG ACA TCT CAA CAA GAC ATT GTC     380
        L  T  I  S  N  L  T  T  S  Q  Q  D  I  V
                    ■

381     CTT GCC GAT GAA CTC AGT CAA GAA GTC TGC ATT CTA TCT GCT     422
        L  A  D  E  L  S  Q  E  V  C  I  L  S  A

423     GAC GTA GTT GTA GGA ATA GCA GCC CCA GGA TGC CCT AAT GCA     464
        D  V  V  V  G  I  A  A  P  G  C  P  N  A

465     CTG AAA GGA AAA ACT GTT CTC GAA AAC TTT GTC GAA GAA AAT     506
        L  K  G  K  T  V  L  E  N  F  V  E  E  N

507     CTT ATT GCG CCT GTC TTT TCT ATT CAT CAT GCT AGA TTT CAA     548
        L  I  A  P  V  F  S  I  H  H  A  R  F  Q

549     GAT GGA GAA CAT TTC GGA GAA ATT ATT TTC GGA GGT TCT GAT     590
        D  G  E  H  F  G  E  I  I  F  G  G  S  D

591     TGG AAA TAC GTT GAT GGT GAA TTC ACT TAT GTT CCA CTT GTG     632
        W  K  Y  V  D  G  E  F  T  Y  V  P  L  V

633     GGT GAT GAT TCC TGG AAG TTC AGG CTG GAT GGT GTG AAA ATA     674
        G  D  D  S  W  K  F  R  L  D  G  V  K  I

675     GGT GAC ACA ACT GTT GCT CCA GCA GGT ACA CAG GCC ATC ATC     716
        G  D  T  T  V  A  P  A  G  T  Q  A  I  I

717     GAC ACA AGC AAA GCT ATC ATT GTC GGA CCT AAA GCC TAT GTT     758
        D  T  S  K  A  I  I  V  G  P  K  A  Y  V

759     AAT CCA ATC AAC GAA GCT ATT GGG TGT GTA GTG GAA AAG ACA     800
        N  P  I  N  E  A  I  G  C  V  V  E  K  T

801     ACA ACC AGG AGA ATA TGC AAG CTT GAC TGC AGC AAG ATA CCA     842
        T  T  R  R  I  C  K  L  D  C  S  K  I  P

843     TCT CTC CCT GAT GTC ACA TTT GTG ATC AAT GGC AGG AAT TTC     884
        S  L  P  D  V  T  F  V  I  N  G  R  N  F

885     AAC ATC AGC TCA CAA TAT TAC ATC CAA CAG AAC GGG AAC TTG     926
        N  I  S  S  Q  Y  Y  I  Q  Q  N  G  N  L
                                          ■

927     TGC TAT TCC GGC TTC CAA CCA TGC GGT CAC TCC GAT CAC TTT     968
        C  Y  S  G  F  Q  P  C  G  H  S  D  H  F

969     TTT ATT GGT GAC TTC TTT GTT GAT CAT TAT TAT TCT GAA TTC    1010
        F  I  G  D  F  F  V  D  H  Y  Y  S  E  F

1011    AAC TGG GAG AAC AAG ACC ATG GGA TTC GGC CGT TCA GTA GAA    1052
        N  W  E  N  K  T  M  G  F  G  R  S  V  E
                ■

1053    AGC GTC TAA GAA TTT CAA CAT CAA GAT GGA CTT CAG AGA TTA    1094
        S  V  ***

1095    CTT CGG AAT CAC TAA TAA GAC ATT CAC GAG ACT TAC GAA GAC    1136
1137    CAC TAC AGT TTT GGA TAT GAA TGA TGA CAA ATA ACT GAA GAC    1178
1179    TTT TCA TTA TAT GAC ATG GAG AAG ATT TTT TTA AAG TCG CCT    1220
1221    ATT ATT ACT TTT TTC GCA CAC TTT TAT GTA TAC AGC TAC TGA    1262
1263    TGT CTT AAA ATA AAC TGG AAA TAT TTT GAA TTT TCT A19        1317
```

FIG. 3. **Nucleotide and deduced amino acid sequence of Bla g 2.** The putative signal peptide is shown in *bold*, and three potential *N*-linked glycosylation sites (■) and the termination codon TAA (*) are indicated. A polyadenylation signal (AATAAA), in the 3′-noncoding region, is underlined. Amino acid sequences determined by Edman degradation are *underlined*; 90.7% (108/119 residues) of the amino acid sequence was identical to that derived from nucleotide sequence. Differences (*underlined*) were found at the amino terminus (VPLYKLVSVFINTQYAGIT - - GNQDFLLVFDTTS - N - VV) and in two tryptic peptides (YYEGEFTYAP and IADDSWQFR).

CR. Unlike the *A. aegypti* aspartic protease, which is located in lysosomes and concentrated in the insect fat body, Bla g 2 is concentrated in the digestive organs, particularly the gut, with much lower levels in the fat body. Our experiments are consistent with previous immunofluoresence studies using IgE Ab, which showed localization of CR allergens to epithelial cells in the intestinal tract and to Malpighian vessels (the major excretory organs in the CR) (47). Previous RAST inhibition studies showed that CR feces are a potent source of allergens (48). Thus we speculate that Bla g 2 is a digestive enzyme that is secreted or excreted along with CR feces.

B. germanica and *P. americana* are the most prevalent domiciliary CR species found in the United States. Although CR allergic patients usually give positive skin prick tests to extracts of both species, the molecular basis of the antigenic relationships between them is poorly understood. Northern analyses have consistently shown that mRNA encoding Bla g 2

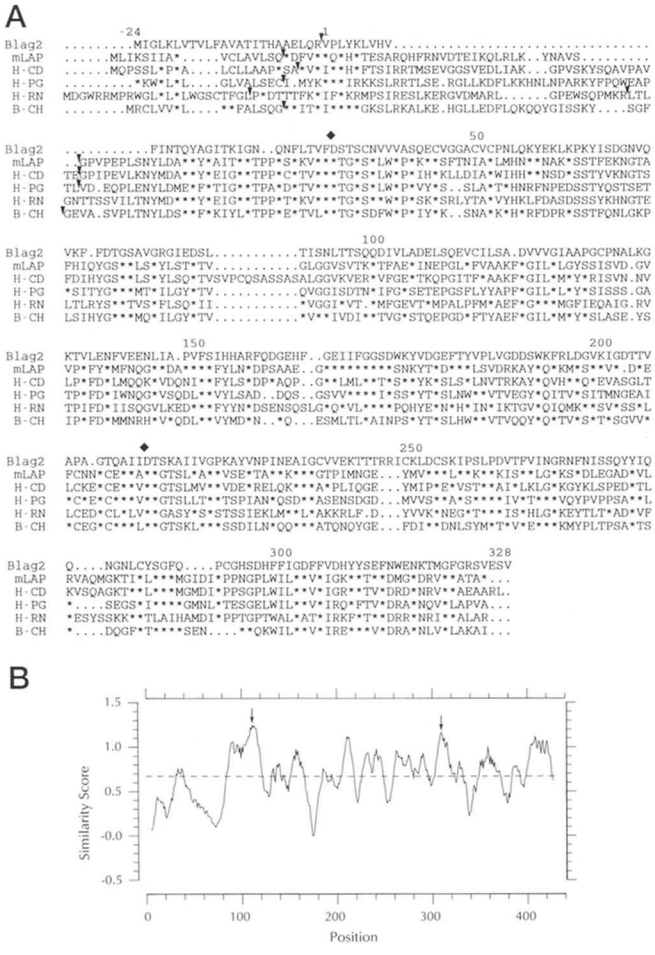

FIG. 4. *A*, alignments of the amino acid sequences of Bla g 2, mosquito lysosomal aspartic protease (*mLAP*), human cathepsin D (*H-CD*), human pepsinogen (*H-PG*), human renin (*H-RN*), and bovine chymosin (*B-CH*). Sequences were aligned using the GCG computer program. Residues identical to Bla g 2 are indicated (*), and gaps (. . .) were introduced by the program for optimal alignment. The *first* and *second arrowheads* represent the signal peptide and pro-peptide cleavage sites in each sequence, respectively. The two conserved aspartic acid residues (*D*) are shown at positions 31 and 215 of the Bla g 2 sequence. *B*, similarity plot of the aspartic protease sequences. The position of the two conserved aspartic acid residues (*D*) in Bla g 2 is indicated. *Numbering* refers to the original computer alignment (*1*, first residue in human renin; *428*, last residue in Bla g 2). The highest degree of similarity was observed around the two enzymatic catalytic sites.

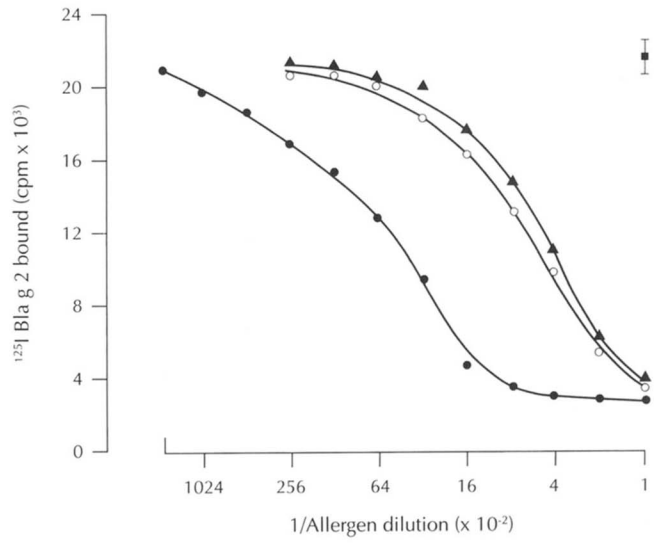

FIG. 5. **Inhibition RIA for Bla g 2.** Dilutions of *B. germanica* frass extract (▲), commercial extract (Greer) (○), or affinity-purified Bla g 2 (●) were used to inhibit binding of ^{125}I-labeled Bla g 2 to mouse polyclonal IgG Ab. The *solid square* (■) represents the mean ± S.D. of results from 13 *P. americana* extracts (including frass, whole body extract, and 11 commercial extracts).

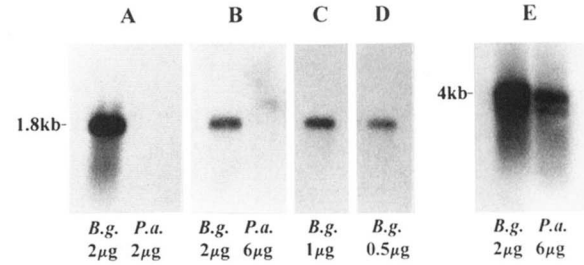

FIG. 6. **Northern blot analysis of *B. germanica* and *P. americana* mRNA.** *B. germanica* (*B.g.*) or *P. americana* (*P.a.*) mRNA was hybridized with ^{32}P-labeled Bla g 2 cDNA probe (*panels A–D*). A ^{32}P-labeled *N. crassa* DNA (pRW528), which hybridizes to highly conserved 18S and 26S ribosomal RNA, was used as a control (*panel E*).

TABLE II

Concentration of Bla g 2 in B. germanica Tissues

B. germanica tissues were disrupted with a Polytron homogenizer, followed by extraction in 0.5 ml of BBS overnight at 4 °C. Extracts were centrifuged, and Bla g 2 levels and total protein concentration were determined.

Tissue	Bla g 2[a]	Protein concentration[b]	Bla g 2: protein
	units/ml	*mg/ml*	
Esophagus	6.1	0.1	61
Crop	8.5	<0.06	
Proventriculus	10.0	0.08	125
Gut	175.2	0.67	261
Legs/wings	176.0	5.0	35
Egg casings	63.0	4.5	14
Fat body	2.7	0.08	34
Salivary glands	0.8	<0.06	
Trachea	0.9	<0.06	

[a] Measured by mAb-based ELISA (38).
[b] Bradford assay using γ globulin as standard.

is not detectable in *P. americana*. In agreement with this, Bla g 2 protein could not be measured in any extracts of *P. americana* using mAb ELISA or inhibition RIA using polyclonal Ab. These results suggested that the previously reported immunologic cross-reactivity between the two CR species must be related to allergens other than Bla g 2 (17, 26). These include Bla g 1 and Per a 1, which have been purified from both species and show antigenic cross-reactivity (27, 28).

Phylogenetically, *B. germanica* and *P. americana* belong to distantly related families, the Blattellidae and Blattidae, respectively (49). In keeping with this, another *B. germanica* allergen, Bla g 4, recently cloned in our laboratory, was only expressed in *B. germanica* (29). Conversely, a 72-kDa *P. americana* allergen (Per a 3) has been isolated, and mAb to this allergen failed to bind to *B. germanica* or *B. orientalis* extracts on immunoblotting or ELISA (50). Thus of the cloned or purified CR allergens that have been defined to date, most (three-fourths) appear to be species-specific. In the case of Bla g 2, this is unusual, since it might be expected that both *B. germanica* and *P. americana* would produce an aspartic protease. However, we have been unable to detect Bla g 2 mRNA in *P.americana*, and attempts to amplify Bla g 2 from *P. americana* genomic DNA using polymerase chain reaction have been unsuccessful (data not shown). While these negative experiments strongly suggest that *P. americana* does not produce a Bla g 2 homologue, there is a possibility that the degree of homology with a putative *P. americana* aspartic protease is too low to be detected by the probes used in our hybridization studies. How-

ever, we believe this possibility is unlikely.

The Bla g 2 cDNA is being subcloned into expression vectors to produce recombinant allergen, which will provide pure protein for diagnostic purposes; for structural and immunologic studies and, potentially, for allergen immunotherapy. The sequence information reported here is essential for identifying B cell and T cell epitopes on Bla g 2 and offers the prospect of developing T cell-based vaccines for CR allergy. Preliminary studies indicate that Bla g 2 causes T cell proliferation in CR allergic patients with asthma. Thus T cell peptides from Bla g 2 could be used to develop new forms of immunotherapy, similar to those that are currently undergoing clinical trials for cat and ragweed allergy (13). Cloning of other CR allergens has recently been reported, and, based on the prevalence of IgE reactivity with these allergens (30–70%), it seems likely the development of improved diagnostic and therapeutic reagents will need to include several of the most important allergens from either species (29, 51, 52). Further immunologic and molecular studies of Bla g 2 and other CR allergens will lead to a better understanding of CR-induced IgE responses and their role in asthma.

Acknowledgments—We are grateful to Tracy Glime and Amy Muir for excellent technical assistance. We also thank Drs. Donald Mullins and Judith Mollett for sending us cockroaches and Dr. Thomas A. E. Platts-Mills for helpful discussions. We are especially grateful to Drs. Susan Squillace, Lawrence Gelber, Robert Call, Peter Heymann, Federico Montealegre, and Lawrence Chiaramonte and staff at the Long Island College Hospital, Brooklyn, NY, for kindly providing sera from cockroach allergic patients. We thank Jacques Retief and Robert Cordaro for computer assistance, Madeline Watkins for drawing the figures, and Nancy Malone for assistance with the preparation of the manuscript.

REFERENCES

1. Chua, K. Y., Stewart, G. A., Thomas, W. R., Simpson, R. J., Dilworth, R. J., Plozza, T. M., and Turner, J. (1988) *J. Exp. Med.* **167,** 175–182
2. Rafnar, T., Ghosh, B., Metzler, W. J., Huang, S.-K., Perry, M. P., Mueller, L., and Marsh, D. G. (1992) *J. Biol. Chem.* **267,** 21119–21123
3. Singh, M. B., Hough, T., Theerakulpusut, P., Avjioglu, A., Davies, S., Smith, P. M., Taylor, P., Simpson, R. J., Ward, L., McCluskey, J., Puy, R., and Knox, B. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1384–1388
4. Morgenstern, J. P., Griffith, I. J., Brauer, A. W., Rogers, B. L., Bond, J. F., Chapman, M. D., and Kuo, M.-c. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 9690–9694
5. Valenta, R., Duchene, M., Pettenburger, K., Sillaber, C., Valent, P., Bettelheim, P., Breitenbach, M., Rumpold, H., Kraft, D., and Scheiner, O. (1991) *Science* **253,** 557–560
6. Arruda, L. K., Mann, B. J., and Chapman, M. D. (1992) *J. Immunol.* **149,** 3354–3359
7. Rafnar, T. I., Griffith, I., Kuo, M., Bond, J. F., Rogers, B. L., and Klapper, D. G. (1991) *J. Biol. Chem.* **266,** 1229–1236
8. van Neerven, R. J. J., van t'Hof, W., Ringrose, J. H., Jansen, H. M., Aalberse, R. C., Wierenga, E. A., and Kapsenberg, M. L. (1993) *J. Immunol.* **151,** 2326–2335
9. O'Hehir, R. E., Verhoef, A., Panagiotopoulou, E., Keswani, S., Hayball, J. D., Thomas, W. R., and Lamb, J. R. (1993) *J. Allergy Clin. Immunol.* **92,** 105–113
10. Yssel H., Johnson, K. E., Schneider, P. V., Wideman, J., Terr, A., Kastelein, R., and de Vries, J. E. (1992) *J. Immunol.* **148,** 738–745
11. van't Hof, W., Driedijk, P. C., van den Berg, M., Beck-Sickinger, A. G., Jung, G., and Aalberse, R. C. (1991) *Mol. Immunol.* **28,** 1225–1232
12. Ebner, C., Szepfalusi, Z., Ferreira, F., Jilek, A., Valenta, R., Parronchi, P., Maggi, E., Romagnani, S., Scheiner, O., and Kraft, D. (1993) *J. Immunol.* **150,** 1047–1054
13. Norman, P. S., Ohman, J. L., Long, A. A., Creticos, P. S., Gefter, M. L., Shaked, Z., Wood, R. A., Eggleston, P. A., Lichtenstein, L. M., Jones, N. H., and Nicodemus, C. F. (1995) *J. Allergy Clin. Immunol.* **95,** 259 (abstr.)
14. Bernton, H. S., and Brown, H. (1964) *J. Allergy* **35,** 506–513
15. Bernton, H. S., McMahon, T. F., and Brown, H. (1972) *Br. J. Dis. Chest* **66,** 61–66
16. Kang, B. (1976) *J. Allergy Clin. Immunol.* **58,** 357–365
17. Twarog, F. J., Picone, F. J., Strunk, R. S., So, J., and Colten, H. R. (1977) *J. Allergy Clin. Immunol.* **59,** 154–160
18. Kang, B., Vellody, D., Homburger, H., and Yunginger, J. W. (1979) *J. Allergy Clin. Immunol.* **63,** 80–86
19. Hulett, A. C., and Dockhorn, R. J. (1979) *Ann. Allergy* **42,** 160–165
20. Chapman, M. D. (1993) *Clin. Exp. Allergy* **23,** 459–461
21. Call, R. S., Smith, T. F., Morris, E., Chapman, M. D., and Platts-Mills, T. A. E. (1992) *J. Pediatrics* **121,** 862–866
22. Gelber, L. E., Seltzer, L., Bouzoukis, J. K., Pollart, S. M., Chapman, M. D., and Platts-Mills, T. A. E. (1993) *Am. Rev. Respir. Dis.* **147,** 573–578
23. Wu, C. H., and Lan, J. L. (1988) *J. Allergy Clin. Immunol.* **82,** 727–735
24. Helm, R. M., Bandele, E. O., Swanson, M. C., Campbell, A. R., and Wynn, S. R. (1988) *Int. Arch. Allergy Appl. Immunol.* **87,** 230–238
25. Kang, B. C., Chang, J. L., and Johnson, J. (1989) *Ann. Allergy* **63,** 207–212.26
26. Stankus, R. P., Horner, W. E., and Lehrer, S. B. (1990) *J. Allergy Clin. Immunol.* **86,** 781–786
27. Schou, C., Lind, P., Fernandez-Caldas, E., Lockey, R. F., and Lowenstein, H. (1990) *J. Allergy Clin. Immunol.* **86,** 935–946
28. Pollart, S. M., Vailes, L. V., Mullins, D. E., Hayden, M. L., Platts-Mills, T. A. E., Sutherland, W. M., and Chapman, M. D. (1991) *J. Allergy Clin. Immunol.* **87,** 511–521
29. Arruda, L. K., Vailes, L. D., Benjamin, D. C., and Chapman, M. D. (1995) *Int. Arch. Allergy Appl. Immunol.*, in press
30. LeGendre, N., and Matsudaira, P. (1989) in *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Matsudaira, P. T., ed) pp. 49–69, Academic Press, Inc., San Diego
31. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299
32. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
33. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2444–2448
34. Pollart, S. M., Chapman, M. D., Fiocco, G. P., Rose, G., and Platts-Mills, T. A. E. (1989) *J. Allergy Clin. Immunol.* **83,** 875–882
35. Chapman, M. D., and Platts-Mills, T. A. E. (1980) *J. Immunol.* **125,** 587–592
36. Russell, P. J., Wagner, S., Rodland, K. D., Feinbaum, R. L., Russel, J. P., Bret-Harte, M. S., Free, S. J., and Metzenberg, R. L. (1984) *Mol. Gen. Genet.* **196,** 275–282
37. Bell, W. J. (1981) *The Laboratory Cockroach*, Chapman and Hall Ltd., London
38. Pollart, S. M., Smith, T. F., Morris, E., Gelber, L. E., Platts-Mills, T. A. E., and Chapman, M. D. (1991) *J. Allergy Clin. Immunol.* **87,** 505–510
39. Cho, W.-L., and Raikhel, A. S. (1992) *J. Biol. Chem.* **267,** 21823–21829
40. Faust, P. L., Kornfeld, S., and Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4910–4914
41. Imai, T., Miyazaki, H., Hirose, S., Hori, H., Hayashi, T., Kageyama, R., Ohkubo, H., Nakanishi, S., and Murakami, K. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 7405–7409
42. Sogawa, K., Fujii-Kuriyama, Y., Mizukami, Y., Ichihara, Y., and Takahashi, K. (1983) *J. Biol. Chem.* **258,** 5306–5311
43. Foltmann, B., Pedersen, V. B., Kauffman, D., and Wybrandt, G. (1979) *J. Biol. Chem.* **254,** 8447–8456
44. Szecsi, P. B. (1992) *Scand. J. Clin. Lab. Invest.* **52,** (Suppl. 210), 5–22
45. Azuma, T., Pals, G., Mohandas, T. K., Couvreur, J. M., and Taggart, R. T. (1989) *J. Biol. Chem.* **264,** 16748–16753
46. Geisow, M. J., and Smyth, D. G. (1980) in *The Enzymology of Post-translational Modification of Proteins* (Freedman, R. B., and Hawkins, H. C., eds) pp. 259–287, Academic Press Inc., New York
47. Zwick, H., Popp, W., Sertl, K., Rauscher, H., and Wanke, T. (1991) *J. Allergy Clin. Immunol.* **87,** 626–630
48. Lehrer, S. B., Horner, E., Menon, P., and Stankus, R. P. (1991) *J. Allergy Clin. Immunol.* **87,** 574–580
49. McKittrick, F. A. (1964) *Evolutionary Studies of Cockroaches*, Memoir 389, Cornell University Agricultural Experiment Station, New York State College of Agriculture, Ithaca, NY
50. Wu, C.-H., Chiang, B. T., Fann, M. C., and Lan, J. L. (1990) *Clin. Exp. Allergy* **20,** 675–681
51. Wu, C.-H., Lee, M.-F., and Liao, S.-C. (1995) *J. Allergy Clin. Immunol.*, in press
52. Helm, R. M., Cockrell, G., Stanley, J. S., Brenner, R. J., Burks, A. W., and Bannon, G. A. (1995) *J. Allergy Clin. Immunol.* **95,** 158 (abstr.)