

The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment

Geoffrey A. Mueller, PhD,^{a*} Lars C. Pedersen, PhD,^{a*} Fred B. Lih, BA,^a Jill Glesner, BS,^b Andrea F. Moon, MS,^a Martin D. Chapman, PhD,^b Kenneth B. Tomer, PhD,^a Robert E. London, PhD,^a and Anna Pomés, PhD^b *Research Triangle Park, NC, and Charlottesville, Va*

Background: Sensitization to cockroach allergens is a major risk factor for asthma. The cockroach allergen Bla g 1 has multiple repeats of approximately 100 amino acids, but the fold of the protein and its biological function are unknown.

Objective: We sought to determine the structure of Bla g 1, investigate the implications for allergic disease, and standardize cockroach exposure assays.

Methods: nBla g 1 and recombinant constructs were compared by using ELISA with specific murine IgG and human IgE. The structure of Bla g 1 was determined by x-ray crystallography. Mass spectrometry and nuclear magnetic resonance spectroscopy were used to examine the ligand-binding properties of the allergen.

Results: The structure of an rBla g 1 construct with comparable IgE and IgG reactivity to the natural allergen was solved by x-ray crystallography. The Bla g 1 repeat forms a novel fold with 6 helices. Two repeats encapsulate a large and nearly spherical hydrophobic cavity, defining the basic structural unit. Lipids in the cavity varied depending on the allergen origin. Palmitic, oleic, and stearic acids were associated with nBla g 1 from cockroach frass. One unit of Bla g 1 was equivalent to 104 ng of allergen.

Conclusions: Bla g 1 has a novel fold with a capacity to bind various lipids, which suggests a digestive function associated with nonspecific transport of lipid molecules in cockroaches. Defining the basic structural unit of Bla g 1 facilitates the standardization of assays in absolute units for the assessment of

environmental allergen exposure. (*J Allergy Clin Immunol* 2013;■■■:■■■-■■■.)

Key words: Allergen, asthma, Bla g 1, cockroach, structure, ligand-binding proteins, exposure assessment

The first report on cockroach allergy dates back to 1964, when Bernton and Brown¹ described an association between sensitization to cockroach allergens and asthma morbidity in New York. Subsequent studies have confirmed this original finding, especially among lower socioeconomic populations but also in suburban middle-class homes of asthmatic children.²⁻⁴ Rates of cockroach sensitization are highest in the northeastern United States, reaching up to 81% in the Bronx, New York, with the highest allergen levels in high-rise apartments.³ A recent study showed that prenatal exposure to cockroach allergens was associated with a greater risk of sensitization.⁵ Allergic sensitization to cockroach is a risk factor for the development of asthma, as is sensitization to other indoor allergens from dust mite, mold, dog, or cat.⁶⁻⁸

A major breakthrough in understanding the role of cockroach allergy in asthma was the identification and measurement of cockroach allergens. Bla g 1 and Bla g 2 were the first 2 major allergens to be identified.⁹ Both are secreted in the cockroach digestive tract, and sensitization occurs through inhalation of fecal particles that carry the allergens and are released to the environment.¹⁰ Both cockroach allergens have been used as markers of allergen exposure. Exposure to low doses of Bla g 1 (≤ 2 U/g of dust) was determined to be a risk factor for sensitization; however, the risk factor plateaus at greater than 4 U/g of dust.^{2,11} A seminal study by Rosenstreich et al^{2,11} established 8 U/g of dust as threshold levels of “morbidity due to asthma,” as part of a National Cooperative Inner-City Asthma Study. However, the levels of Bla g 1 exposure in all these studies were expressed in arbitrary units and should be converted to absolute protein values to make comparable assessments of exposure to different allergens.

Assays to measure cockroach allergen exposure were originally developed with mAbs raised against cockroach extracts in the 1990s.¹² The cockroach extracts used for assay standardization were arbitrarily assigned a certain number of units for each allergen (Bla g 1 and Bla g 2) in a fixed volume of extract.¹³ Subsequently, the molecular cloning of Bla g 2 revealed that this globular protein is stable with a well-defined structure, and this allowed the measurement of this allergen in absolute units based on amino acid analysis of purified Bla g 2 (1 unit = approximately 80 ng).^{14,15}

However, Bla g 1 is a more complex allergen.^{10,13,16,17} Molecular cloning of Bla g 1 revealed an unusual primary structure containing repeats of approximately 100 amino acids. The gene for

From ^athe Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, and ^bIndoor Biotechnologies, Charlottesville.

*These authors contributed equally to this work.

Supported in part by Research Project no. Z01-ES102885-01 to R.E.L., Z01-ES50161 to K.B.T., and ZIA-ES102645 to L.C.P. in the Intramural Research Program of the National Institute of Environmental Health Sciences, National Institutes of Health. Research reported in this publication was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award no. R01AI077653 (to A.P. and M.D.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Use of the Advanced Photon Source, Argonne National Laboratory, was supported by the US department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-Eng-38.

Disclosure of potential conflict of interest: A. Pomés and M. D. Chapman have received a grant from the National Institute of Allergy and Infectious Disease (NIAID) and are employed by Indoor Biotechnologies, Inc. M. D. Chapman is co-owner and shareholder of Indoor Biotechnologies, Inc, and receives royalties from the University of Virginia. The rest of the authors declare that they have no relevant conflicts of interest. Received for publication January 30, 2013; revised May 28, 2013; accepted for publication June 19, 2013.

Corresponding author: Geoffrey A. Mueller, PhD, National Institute of Environmental Health Sciences, 111 T.W. Alexander Dr, MD-MR-01, Research Triangle Park, NC 27709. E-mail: Mueller3@niehs.nih.gov.

0091-6749

<http://dx.doi.org/10.1016/j.jaci.2013.06.014>

Abbreviations used

GFP:	Green fluorescent protein
GST:	Glutathione-S-transferase
MS:	Mass spectrometry
NMR:	Nuclear magnetic resonance spectroscopy
NSP:	Nitrile-specifier protein
PC:	Phosphatidylcholine
PE:	Phosphatidylethanolamine
PG:	Phosphatidylglycerol
PI:	Phosphatidylinositol
rBla g 1-EC:	Recombinant Bla g 1 expressed in <i>Escherichia coli</i>
rBla g 1-GFP:	Recombinant Bla g 1 with an N-terminal GFP
rBla g 1-PP:	Recombinant Bla g 1 expressed in <i>Pichia pastoris</i>
TEV:	Tobacco etch virus
TLR:	Toll-like receptor

Bla g 1 originated from an approximately 100 amino acid sequence that duplicated. The sequence of the second repeat diverged from the first, and this “duplex” subsequently multiplied to include many “duplexes” of the repeated motif. If we number the repeats consecutively, consecutive repeats are 26% to 29% identical, whereas comparisons among odd-numbered repeats or among even-numbered repeats show 96% to 98% identity (see Fig E1 in this article’s Online Repository at www.jacionline.org).¹⁸ The number of repeats in clones reported for group 1 cockroach allergens is between 4 and 14.^{10,19,20} In addition to variability in length due to genetic diversity, the allergen can appear to be 33 to 37, 28, 25, or 6 kDa depending on the method used for purification or visualization.^{13,16,21} The different nBla g 1 molecular forms result from cleavage by trypsin-like enzymes present in the cockroach gut.^{10,13,16} Among insects, Bla g 1-homologous proteins are found with a similar genetic structure (see Fig E1).^{18,22}

Given the complexity of this allergen, the Bla g 1 units remained arbitrary until the allergen could be better characterized. The determination of the structure of Bla g 1, as presented in this article, reveals a basic structural unit comprised of 2 consecutive amino acid repeats. Knowledge of the structure of Bla g 1 allows the standardization of assays to measure this allergen in absolute instead of relative units. In addition, the structure provides novel information about the function of this protein in the cockroach gut, the nature of the repeats, and its allergenic potential.

METHODS**Constructs**

nBla g 1 exists in multiple molecular forms, and cloning of the gene revealed repeated amino acid sequences.^{10,16} The largest repeated unit, containing a “duplex” from Bla g 1.0101 consisting of 2 consecutive repeats, was selected for expression in *Escherichia coli* and crystallization (based on previous observations that proteolysis of the protein generated a similarly sized molecule) to determine the 3-dimensional structure of Bla g 1.¹⁷ Green fluorescent protein (GFP) was fused to the N-terminus of the allergen to stabilize the construct and facilitate crystallization; the construct was named rBla g 1 with an N-terminal GFP (rBla g 1-GFP; see Fig E2 and details in the Methods section in this article’s Online Repository at www.jacionline.org).^{23,24}

For standardization purposes, 3 Bla g 1 preparations were tested, containing Bla g 1 molecules of different lengths. Two were expressed by methanol induction in *Pichia pastoris*, as previously described, and were named recombinant Bla g 1 expressed in *P pastoris* (rBla g 1-PP).²⁵ These

2 preparations contained different molecular forms, resulting from expression of the same Bla g 1.0101 clone containing 2 duplexes (accession no. AF072219), under different conditions (presence or absence of antibiotic zeocin for lots 34074 and 33045, respectively). In addition, an isolated rBla g 1 construct of repeats 1 and 2 (rBla g 1 expressed in *E coli* [rBla g 1-EC]) was obtained by means of cleavage with tobacco etch virus (TEV) protease from a fusion with glutathione-S-transferase (GST) expressed in *E coli* (see details in the Methods section in this article’s Online Repository). The reference standard containing nBla g 1 was a standard prepared from a *Blattella germanica* frass extract and used for ELISA that contains 10 U/mL Bla g 1 (lot no. 32023; INDOOR Biotechnologies, Charlottesville, Va).¹³ The protein content of the 3 Bla g 1 preparations was quantified by using amino acid analysis. Briefly, samples were hydrolyzed (6N HCl, 110°C, 24 hours), and the resulting amino acids were separated on a strong cation exchange column detected with a secondary reaction with ninhydrin and quantified against a known standard run in the same sequence.

For nuclear magnetic resonance spectroscopic (NMR) and mass spectrometric (MS) studies, the Bla g 1 constructs rBla g 1-EC and rBla g 1-PP (lot 34074) were used. nBla g 1 was purified from cockroach frass (debris and feces produced by the cockroach) by using the anti-Bla g 1 mAb 10A6 and a similar protocol to that described previously for purifying nBla g 2 from frass.^{10,25}

Antibody binding to nBla g 1 and rBla g 1 constructs

IgE antibody binding to natural and recombinant Bla g 1-GFP used for crystallization was compared by using ELISA with microtiter plates that were coated with anti-Bla g 1 mAb 10A6, as described previously.²⁵ After incubation with the cockroach extract or the recombinant allergen, sera from patients with cockroach allergy (n = 15) were added, and bound IgE was detected by using biotinylated goat anti-human IgE. Details regarding the sera are shown in the Methods section in this article’s Online Repository. Control sera were from a nonallergic patient and 2 patients with mite allergy. An IgE standard curve was obtained with anti-Der p 2 mAb, nDer p 2, and the chimeric antibody 2B12-IgE, with a dynamic range of 0.5 to 250 ng IgE/mL.²⁶

For standardization purposes, a polyclonal rabbit anti-Bla g 1 antibody was used for detection instead of IgE in an ELISA, as previously described.¹³ Dose-response curves (n = 4 per Bla g 1 preparation) were performed to compare 3 rBla g 1 preparations with the standard containing nBla g 1. Curves were fitted by using MATLAB, and the equivalence between relative and absolute units at the half maximal effective concentration was determined.

Structural and biochemical characterization

Details of the crystallography, MS analysis, and NMR procedures are presented in the Methods section in this article’s Online Repository.

RESULTS**rBla g 1 is comparable with natural allergen**

The rBla g 1 construct used for crystallization was compared with the natural allergen for human IgE antibody binding. There was an excellent quantitative correlation between IgE antibody binding to nBla g 1 and rBla g 1-GFP by using sera from patients with cockroach allergy (n = 15, $r = 0.96$, $P < .001$, Fig 1). These results indicated that the recombinant and natural allergens were equally recognized by IgE antibodies and that the GFP molecule used in the rBla g 1-GFP construct for crystallography did not interfere with IgE antibody binding.

Bla g 1 has a novel 3-dimensional structure

The structure of rBla g 1-GFP was determined by using x-ray crystallography (PDB code: 4JRB, see Fig E3 and Table E1 in this

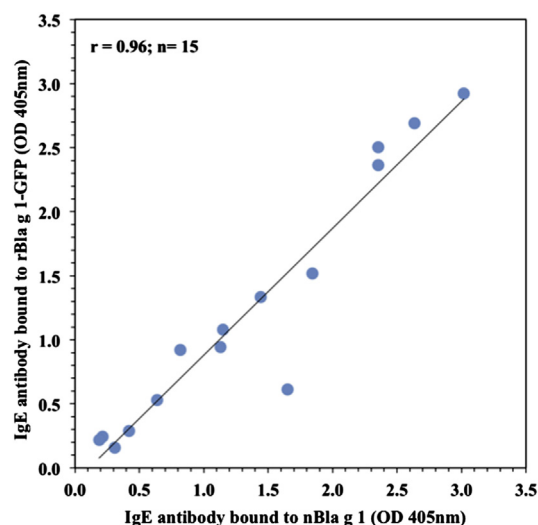


FIG 1. Correlation between IgE antibody binding to rBla g 1-GFP and nBla g 1 with sera from patients with cockroach allergy, as measured by using ELISA.

article's Online Repository at www.jacionline.org). The structure of each 100 amino acid repeat of Bla g 1 is comprised of 6 short helices. The first 5 helices of each repeat form a planar pentagon, whereas the sixth helix is positioned above and across the pentagon face (Fig 2). Two consecutive pentagonal structures come together to form a duplex in which the interacting helices from each repeat run in opposite directions. The structures of the 2 repeats are remarkably similar (C α RMSD of 1.4 Å over the first 5 helices) given only 26% identity of the amino acid sequences. Attempts to identify a similar fold in the protein database were unsuccessful by using the program DALI.²⁷ The top 40 protein structures from the DALI search had low significance scores (5-7), but each was manually examined for similarities, and none matched Bla g 1, indicating this is likely a new fold family.

Defining the allergenic unit

Based on the structure and the data in Fig 1, a duplex of 2 repeated sequences appears to define the minimum structural unit. The recombinant allergens expressed in *E coli* (rBla g 1-EC) and *P pastoris* (rBla g 1-PP, lots 34075 and 33045) were compared for murine/rabbit IgG antibody binding to the ELISA standard containing nBla g 1 in extract. Note that rBla g 1-PP has 2 of the structural units defined in Fig 2, rBla g 1-EC has 1 structural unit, and nBla g 1 has been reported as differently sized molecular forms.^{10,12,16,21} Parallel and overlapping dose-response curves were obtained for the 3 allergen preparations by using ELISA, indicating equivalent antibody binding on a gram per milliliter basis (Fig 3, A). The equivalence between absolute and relative units of nBla g 1 in the standard extract was determined to be as follows: 1 unit equals 104 ± 1 ng ($n = 4$ per preparation; 3 preparations). Fig 3, B, shows an SDS-PAGE analysis of the samples used for standardization. The antibodies equivalently recognized the different preparations despite seemingly different fragmentation patterns. MS analysis showed that the fragmentation patterns of rBla g 1-PP resulted from partial proteolysis between and within the duplexes in the unstructured loops between the repeats (see Fig E4 and Table E2 in this article's Online Repository at www.jacionline.org).

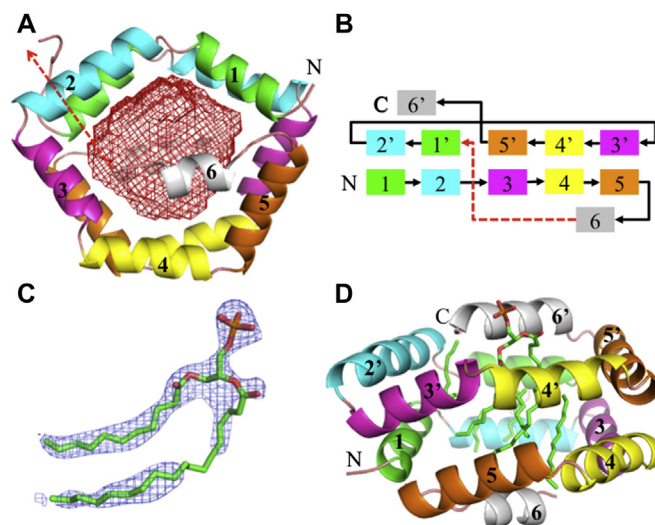


FIG 2. Bla g 1 structure. **A**, The "top view" shows the first 5 helices forming a pentagon. The dashed red arrow indicates residues with missing electron density between helix 6 and 1'. The hydrophobic cavity is indicated by red mesh. **B**, Topologic diagram showing the interactions of the helices. **C**, Electron density in the cavity was modeled with phosphatidic acid and short acyl chains, which are rendered as sticks in Fig 2, D. **D**, The "side view" shows how the helices interact between the repeats.

Lipid binding of Bla g 1

Another defining feature of the Bla g 1 molecule is that the duplex of repeats 1 and 2 enclose a large spherical cavity that is lined exclusively with hydrophobic residues. The cavity is approximately 3750 Å^3 (Fig 2). For comparison, in a survey of almost 600 enzymes, the mean active site volume was 1072 Å^3 .²⁸ Within the cavity exists electron density consistent with long-chain fatty acids. This fragmented density was modeled with 6 carbon chains of variable length, as well as 1 generic phospholipid.

Bla g 1 molecules from various sources were analyzed for lipid content by using MS and NMR to confirm the ligand content in the cavity. In the *E coli*-expressed Bla g 1 constructs (rBla g 1-EC), phosphatidylethanolamine (PE) and phosphatidylglycerol (PG) head groups with various lengths of saturated and unsaturated diacyl chains were found in abundance by using MS (see Fig E5 in this article's Online Repository at www.jacionline.org). On the basis of total ion count, the Bla g 1-EC sample contained 60% PG and 40% PE, whereas the *E coli* lysate contained 36% PG and 64% PE. rBla g 1-EC had fewer of the cyclopropyl acyl chains than are commonly found in *E coli* membranes. Fig 4, A, shows the ^{31}P NMR spectra of the phospholipids derived from the insoluble fraction of *E coli* lysate, from rBla g 1-EC, and from GST as a negative control (see the Methods section in this article's Online Repository). The ^{31}P NMR spectrum of rBla g 1-EC showed a stronger peak intensity at the shift of PG versus PE, which is the opposite of the normal distribution in *E coli* (Fig 4, B). The MS and NMR data indicate that Bla g 1 has a relatively greater affinity for the anionic PG relative to the other lipid species present in *E coli*, although it is not so selective that it binds to only this type of lipid. As a negative control, the GST sample did not show any significant phospholipid peaks (Fig 4, C). In summary, the data show that Bla g 1 has an affinity for saturated or nearly saturated acyl chains and preferentially binds PG over PE from *E coli*.

A similar MS and NMR analysis of rBla g 1-PP showed that rBla g 1-PP retains phosphatidylinositol (PI), phosphatidylserine,

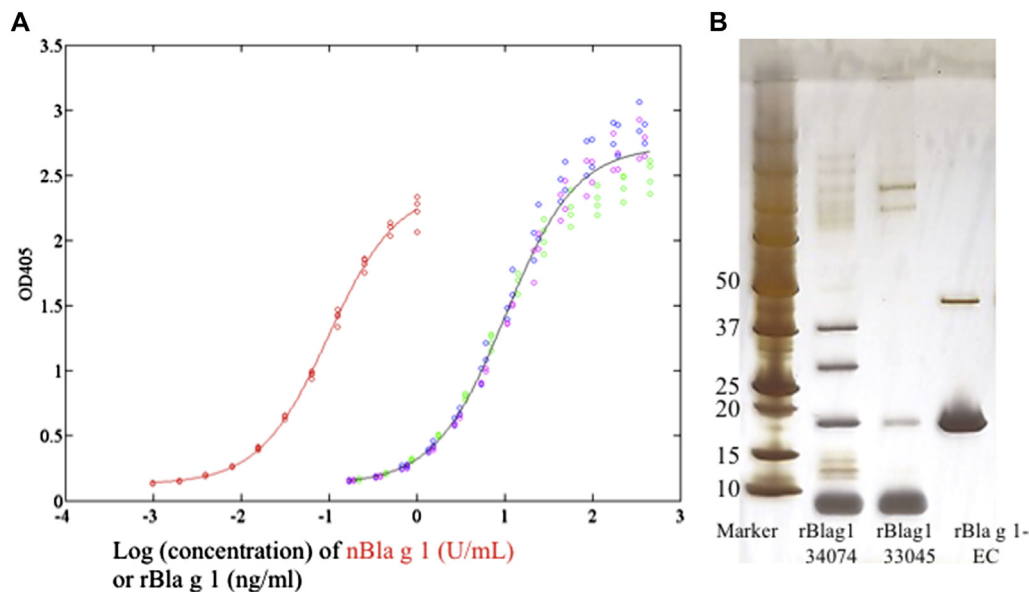


FIG 3. Standardization of nBla g 1 units. **A**, Binding curves with rabbit polyclonal antibodies to rBla g 1-EC (green circles, $n = 4$), rBla g 1-PP lot 34074 (blue circles, $n = 4$), rBla g 1-PP lot 33045 (magenta circles, $n = 4$), and nBla g 1 from the standard cockroach extract (lot no. 32023; red circles, $n = 4$). Fits to the curves to determine the half maximal effective concentration are shown for nBla g 1 (red line) and the recombinant constructs, which were fit as a pooled data set (black line, $n = 12$). **B**, Silver-stained SDS-PAGE of protein samples used in standardization. Note that the concentration loaded on the gel was approximately 1000-fold greater than the highest data point in the ELISA. A small amount of uncleaved (GST) Bla g 1 is apparent in lane 4 at 50 kDa.

and phosphatidylcholine (PC). An NMR comparison of the relative abundance of phospholipids in an extract of *P. pastoris* (Fig 4, D) and the yeast-expressed allergen (Fig 4, E) showed that rBla g 1-PP preferentially retains PC and PI and does not bind phosphatidic acid or PE. nBla g 1 purified from cockroach frass was analyzed by using MS to determine whether phospholipid binding was relevant in the cockroach. Surprisingly, no phospholipids were detected in nBla g 1. Instead, the excreted allergen contains primarily the nonphosphorylated fatty acids palmitate (C16:0), stearate (C18:0), and oleate (C18:1; data not shown).

Comparisons with Per a 1 and *Pieris rapae* nitrile-specifier protein

Bla g 1 is the first structure to be solved in a family of proteins identified in cockroaches (Per a 1, 70% to 72% identity), butterflies (*P. rapae* nitrile-specifier protein [NSP], 20% to 30% identity), and other insects.²⁹ On the basis of the sequence alignments in Fig E1, residues on the molecular surface of Bla g 1 were categorized according to their similarity with Per a 1 and *P. rapae* NSP (Fig 5). This analysis predicts that the Per a 1 and Bla g 1 surfaces will be nearly identical (Fig 5, A and B). By using the Bla g 1 structure, calculations of the solvent-accessible area showed that 87% of the surface of cockroach group 1 allergens is likely to be identical or similar (Fig 5, C). The most notable area of sequence differences forming a contiguous surface is located in the segment from helix 4 to helix 6 (Fig 5, B). In contrast, a similar comparison between Bla g 1 and a distantly related protein from *P. rapae* showed that only 39% of the surface area was predicted to be identical or similar (Fig 5, D). The sequence differences between Bla g 1 and the other proteins were not limited solely

to the external surfaces but were equally distributed among buried and surface-exposed residues in Bla g 1.

DISCUSSION

In this study we present an unusual example of how the determination of the molecular structure of an allergen, Bla g 1, allows progress in the assessment of allergen exposure associated with respiratory symptoms. Over the years, measurement of environmental exposure to cockroach allergens has been essential to risk assessment for inner-city asthma studies.^{2-4,11,30} Allergen measurements have also been important for integrative pest management to reduce allergen exposure and improve asthma.³¹⁻³³ Regarding diagnosis, a high variability in Bla g 1 and Bla g 2 content has been found in commercial cockroach skin test extracts, with an up to 200-fold difference in Bla g 1 levels (4.7-1085 U/mL, $n = 6$ commercial extracts).³⁴ However, in all previous studies the levels of allergen exposure were expressed in relative units.

A process of assay standardization in absolute units is presented for Bla g 1. This usually requires the comparison of the Bla g 1 standard used in the environmental assays with a known amount of purified allergen by using a reliable protein quantification technique. nBla g 1 is a very unusual allergen that exists in multiple molecular forms resulting from trypsin cleavage of the protein (see the Results section in this article's Online Repository at www.jacionline.org).¹⁰ A molecule that breaks down into fragments would be typically difficult or impossible to standardize because of the consequent abrogation of antibody epitopes. The importance of the new structural motif for the quantitation of Bla g 1 is that it provides a rationale to understand why the molecule preserves antibody epitopes in solution and yet appears fragmented on a denaturing gel. The 3-dimensional

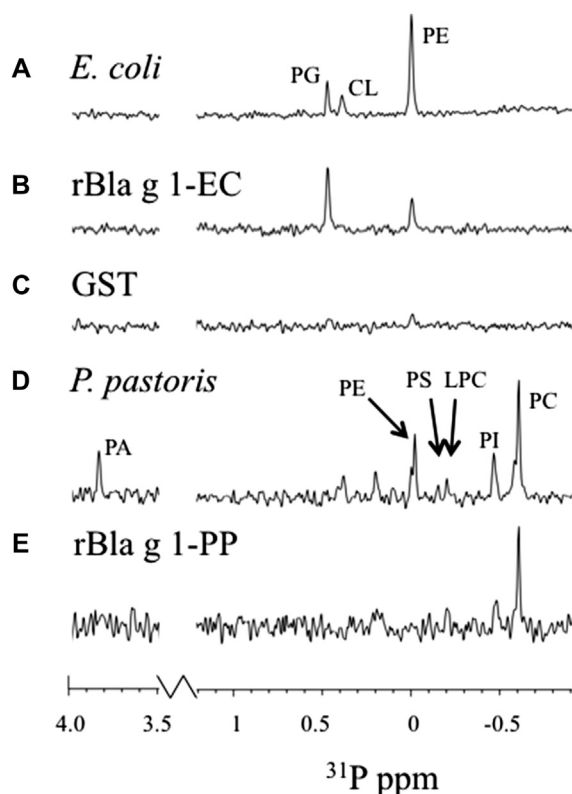


FIG 4. Comparison of the phospholipid content with ^{31}P NMR of *E coli* extract (positive control; **A**), rBla g 1-EC (**B**), GST expressed in *E coli* (negative control; **C**), *P pastoris* extract (**D**), and rBla g 1-PP (**E**). The ^{31}P NMR peaks were assigned based on literature values, relative abundance, and the MS data in this article. CL, Cardiolipin; LPC, lysophosphatidylcholine; PS, phosphatidylserine.

structure of repeats 1 and 2 determined by using x-ray crystallography is novel, and the basic structural unit has the shape of a capsule or bead with a large internal cavity that accommodates lipid ligands. This suggests that nBla g 1 might resemble beads on a string attached by flexible linkers (see Fig E4). The multiple molecular forms of Bla g 1 observed by means of SDS-PAGE resemble fragments of a broken necklace, each with different numbers of beads or half beads because the long flexible linker between the repeats is also vulnerable to proteolysis (see Fig E4). The crystallographic and antibody-binding results suggest that the capsule structure is likely maintained in solution, even with limited proteolysis. This scenario is very different from a globular molecule that degrades into different fragments with a consequent loss of epitopes, which makes standardization impossible.

We showed, using ELISA dose-response curves, that the behavior of different Bla g 1 preparations was very similar regarding antibody binding and comparable with the natural allergen, despite a different molecular fragment composition. The basic structural unit showed a good correlation with the natural allergen regarding IgE antibody binding, as also reported for the *Pichia*-expressed allergen.¹⁷ In addition, the 4 bands of the *Pichia*-expressed Bla g 1 were shown to bind IgE antibodies,¹⁷ which is consistent with the fact that allergic sensitization occurs by means of inhalation of fragmented nBla g 1.¹⁰ In summary, a similar fragmentation pattern occurs in natural and recombinant allergens, leading to molecules that are relevant for patient IgE and mAb binding.

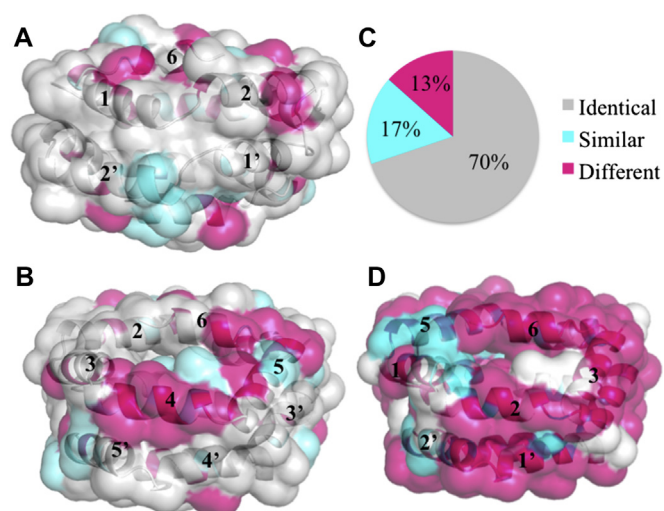


FIG 5. Surface residue comparison among Bla g 1, Per a 1, and *P rapae* NSP. **A** and **B**, The solvent-accessible surface of Bla g 1 in 2 different orientations colored by residue identity (white), similarity (cyan), and differences (hot pink) with Per a 1. Visible helices are labeled by number beneath the semitransparent surface. **C**, A pie chart of the total solvent-accessible surface categorized similarly. **D**, A similar comparison with the distantly related *P rapae* NSP.

One arbitrary unit of Bla g 1 was calculated to be equivalent to 104 ng as absolute units. Bla g 2, with a similar equivalence of 80 ng per unit, has been reported to be very potent, inducing IgE sensitization at exposure levels that are 10- to 100-fold lower than those of other common indoor allergens, such as dust mite and cat.³⁵ The threshold levels for risk of sensitization and asthma morbidity (2 and 8 U/g of dust, respectively) would be equivalent to 208 and 832 ng/g of dust for Bla g 1. These values were comparable with levels of 160 and 640 ng/g of dust for Bla g 2 but lower than values reported for dust mite and cat, confirming the potency of cockroach allergens. In summary, these results indicate that Bla g 1 and Bla g 2 have a similar allergenic potential, which is high compared with other indoor allergens, and allow direct comparison of allergen exposure measurements for different allergens.⁶

The structure of Bla g 1 is the first to be described for this group of insect proteins. Electron density in the cavity of Bla g 1 resembled phospholipids, and this in turn led us to investigate the internal ligands in Bla g 1 using MS and NMR. Both methods indicated that phospholipids with various head groups could be retained by Bla g 1. MS analysis primarily identified saturated or monounsaturated acyl chains, as well as a few cyclopropyl acyl chains. Bla g 1.01 appears partially selective for certain phospholipid head groups, including PC, PI, and PG. Promiscuous binding of lipids would be consistent with the idea of Bla g 1 being important for general nutrient uptake in cockroach: (1) the allergen is expressed in the midgut and excreted in fecal particles^{10,36}; (2) the expression of Bla g 1 increases with food intake³⁷; (3) knockdown of Bla g 1 expression by using RNA interference blocked nutrient absorption³⁸; (4) homologous proteins in mosquitoes (ANG12 and AEG12) are upregulated in the midgut after a blood meal³⁹; and (5) AEG12 was localized by means of immunogold staining to microvilli in the gut.⁴⁰ On the basis of our results, it is conceivable that Bla g 1 might be involved in the uptake of dietary lipids. The binding of certain lipids (palmitate, stearate, and oleate) and not others to Bla g

1 from cockroach frass can be interpreted as a selective removal and excretion of these lipids into the feces to facilitate retention and absorption of other, more nutritive compounds. Alternatively, homologous butterfly proteins allow the larvae to excrete toxic nitriles as an adaptation against the major chemical defense of their host plants.^{22,41} In either case, the structure and data shown here suggest this protein fold could accommodate a variety of hydrophobic ligands for various physiologic purposes.

A common feature of many allergens is the ability to bind hydrophobic ligands, which are potent stimulators of the innate immune system.⁴² The lipids that were found by using MS in the cockroach frass are fatty acids that are known to activate Toll-like receptor (TLR) 4 and TLR2.⁴³ Low-level stimulation of the TLR4 pathway by LPS was shown to drive the immune response toward T_H2,^{44,45} and TLR2 stimulation in the presence of allergen enhances T_H2 responses.^{46,47} Lipid adjuvants, like those found here associated with Bla g 1, can strongly influence the innate immune response and subsequently skew the adaptive response toward allergy. However, this remains to be demonstrated for nBla g 1 and fatty acids.

There is currently very little literature on specific epitopes of group 1 cockroach allergens, except for 2 small-scale studies with peptides of Per a 1.^{48,49} On the basis of the approximately 70% sequence identity between Per a 1 and Bla g 1, the homologous allergen from the American cockroach is expected to have a similar fold and similar ligand specificity and is unlikely to be a membrane-spanning protein, as was previously suggested.⁵⁰ Looking at the unique fold of Bla g 1, there are very few long stretches of amino acids that are significantly buried because there is not a canonical protein hydrophobic core. The hollow spherical shape exposes many more residues as potential IgE epitopes compared with a globular protein of similar mass, and hence almost the entire sequence is potentially antigenic. The structure presented here should be very useful for future analysis of the antigenic surface and the design of Bla g 1 molecules for safer and more effective immunotherapeutics. It also provides insight into the digestive role of Bla g 1, with the potential adjuvant effects of lipids in allergenicity. The Bla g 1 structure facilitated measurement of the conversion factor from the previous arbitrary units to absolute units for the standardization of assays to assess allergen environmental exposure and allergen content in extracts used for diagnosis. These measurements are essential for understanding allergen levels associated with risk of sensitization and asthma morbidity and for reducing allergen exposure and managing disease.

We thank Jason Williams, PhD, for assistance with MS.

Key messages

- Bla g 1, like many allergens, binds hydrophobic ligands. It is thought that these ligands can contribute to allergenicity as adjuvants.
- The structurally defined basic allergenic unit provides absolute standardization for cockroach exposure assessment for asthmatic patients.

REFERENCES

- Bernton HS, Brown H. Insect allergy—preliminary studies of cockroach. *J Allergy* 1964;35:506-13.
- Rosenstreich DL, Eggleston P, Kattan M, Baker D, Slavin RG, Gergen P, et al. The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. *N Engl J Med* 1997;336:1356-63.
- Gruchalla RS, Pongracic J, Plaut M, Evans R, Visness CM, Walter M, et al. Inner City Asthma Study: relationships among sensitivity, allergen exposure, and asthma morbidity. *J Allergy Clin Immunol* 2005;115:478-85.
- Matsui EC, Wood RA, Rand C, Kanchanaraks S, Swartz L, Curtin-Brosnan J, et al. Cockroach allergen exposure and sensitization in suburban middle-class children with asthma. *J Allergy Clin Immunol* 2003;112:87-92.
- Perzanowski MS, Chew GL, Divjan A, Jung KH, Ridder R, Tang DL, et al. Early-life cockroach allergen and polycyclic aromatic hydrocarbon exposures predict cockroach sensitization among inner-city children. *J Allergy Clin Immunol* 2013;131:886-93.
- Platts-Mills TA, Vervloet D, Thomas WR, Aalberse RC, Chapman MD. Indoor allergens and asthma: report of the Third International Workshop. *J Allergy Clin Immunol* 1997;100:S2-24.
- Salo PM, Arbes SJ, Crockett PW, Thorne PS, Cohn RD, Zeldin DC. Exposure to multiple indoor allergens in US homes and its relationship to asthma. *J Allergy Clin Immunol* 2008;121:678-84.
- Matsui EC, Sampson HA, Bahnson HT, Gruchalla RS, Pongracic JA, Teach SJ, et al. Allergen-specific IgE as a biomarker of exposure plus sensitization in inner-city adolescents with asthma. *Allergy* 2010;65:1414-22.
- Arruda LK, Vailes LD, Ferriani VPL, Santos BR, Pomés A, Chapman MD. Cockroach allergens and asthma. *J Allergy Clin Immunol* 2001;107:419-28.
- Pomés A, Melen E, Vailes LD, Retief JD, Arruda LK, Chapman MD. Novel allergen structures with tandem amino acid repeats derived from German and American cockroach. *J Biol Chem* 1998;273:30801-7.
- Eggleston PA, Rosenstreich D, Lynn M, Gergen P, Baker D, Kattan M, et al. Relationship of indoor allergen exposure to skin test sensitivity in inner-city children with asthma. *J Allergy Clin Immunol* 1998;102:563-70.
- Pollart SM, Smith TF, Morris EC, Gelber LE, Platts-Mills TAE, Chapman MD. Environmental exposure to cockroach allergens—analysis with monoclonal antibody-based enzyme immunoassays. *J Allergy Clin Immunol* 1991;87:505-10.
- Pollart SM, Mullins DE, Vailes LD, Hayden ML, Platts-Mills TAE, Sutherland WM, et al. Identification, quantitation, and purification of cockroach allergens using monoclonal-antibodies. *J Allergy Clin Immunol* 1991;87:511-21.
- Arruda LK, Vailes LD, Mann BJ, Shannon J, Fox JW, Vedvick TS, et al. Molecular-cloning of a major cockroach (*Blattella germanica*) allergen, Bla-G-2—sequence homology to the aspartic proteases. *J Biol Chem* 1995;270:19563-8.
- Gustchina A, Li M, Wunschmann S, Chapman MD, Pomés A, Wlodawer A. Crystal structure of cockroach allergen Bla g 2, an unusual zinc binding aspartic protease with a novel mode of self-inhibition. *J Mol Biol* 2005;348:433-44.
- Schou C, Lind P, Fernandezcaldas E, Lockey RF, Lowenstein H. Identification and purification of an important cross-reactive allergen from American (*Periplaneta americana*) and German (*Blattella germanica*) cockroach. *J Allergy Clin Immunol* 1990;86:935-46.
- Pomés A, Vailes LD, Helm RM, Chapman MD. IgE reactivity of tandem repeats derived from cockroach allergen, Bla g 1. *Eur J Biochem* 2002;269:3086-92.
- Pomés A, Wunschmann S, Hindley J, Vailes LD, Chapman MD. Cockroach allergens: function, structure and allergenicity. *Protein Pept Lett* 2007;14:960-9.
- Yang CY, Wu JD, Wu CH. Sequence analysis of the first complete cDNA clone encoding an American cockroach Per a 1 allergen. *Biochim Biophys Acta* 2000;1517:153-8.
- Helm R, Cockrell G, Stanley JS, Brenner RJ, Burks W, Bannon GA. Isolation and characterization of a clone encoding a major allergen (Bla g Bd90K) involved in IgE-mediated cockroach hypersensitivity. *J Allergy Clin Immunol* 1996;98:172-80.
- Twarog FJ, Picone FJ, Strunk RS, So J, Colten HR. Immediate hypersensitivity to cockroach— isolation and purification of major antigens. *J Allergy Clin Immunol* 1977;59:154-60.
- Fischer HM, Wheat CW, Heckel DG, Vogel H. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Mol Biol Evol* 2008;25:809-20.
- Mueller GA, Edwards LL, Aloor JJ, Fessler MB, Glesner J, Pomés A, et al. The structure of the dust mite allergen Der p 7 reveals similarities to innate immune proteins. *J Allergy Clin Immunol* 2010;125:909-17.
- Mueller GA, Gosavi RA, Pomés A, Wunschmann S, Moon AF, London RE, et al. Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity. *Allergy* 2011;66:878-85.
- Pomés A, Chapman MD, Vailes LD, Blundell TL, Dhanaraj V. Cockroach allergen Bla g 2—structure, function, and implications for allergic sensitization. *Am J Respir Crit Care Med* 2002;165:391-7.
- Schuerman J, Perdok GJ, Lourens TE, Parren PWHI, Chapman MD, Aalberse RC. Production of a mouse/human chimeric IgE monoclonal antibody to the house dust mite allergen Der p 2 and its use for the absolute quantification of allergen-specific IgE. *J Allergy Clin Immunol* 1997;99:545-50.

27. Holm L, Sander C. The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res* 1996;24:206-9.
28. Jimenez-Morales D, Liang J, Eisenberg B. Ionizable side chains at catalytic active sites of enzymes. *Eur Biophys J* 2012;41:449-60.
29. Dostalova A, Votykka J, Favreau AJ, Barbican KD, Volf P, Valenzuela JG, et al. The midgut transcriptome of *Phlebotomus (Larroussius) perniciosus*, a vector of *Leishmania infantum*: comparison of sugar fed and blood fed sand flies. *BMC Genomics* 2011;12:223.
30. deBlay F, Sanchez J, Hedelin G, PerezInfante A, Verot A, Chapman M, et al. Dust and airborne exposure to allergens derived from cockroach (*Blattella germanica*) in low-cost public housing in Strasbourg (France). *J Allergy Clin Immunol* 1997; 99:107-12.
31. Arbes SJ, Sever M, Mehta J, Gore JC, Schal C, Vaughn B, et al. Abatement of cockroach allergens (Bla g 1 and Bla g 2) in low-income, urban housing: month 12 continuation results. *J Allergy Clin Immunol* 2004;113:109-14.
32. Gore JC, Schal C. Cockroach allergen biology and mitigation in the indoor environment. *Annu Rev Entomol* 2007;52:439-63.
33. Morgan WJ, Crain EF, Gruchalla RS, O'Connor GT, Kattan M, Evans RI, et al. Results of a home-based environmental intervention among urban children with asthma. *N Engl J Med* 2004;351:1068-80.
34. Patterson ML, Slater JE. Characterization and comparison of commercially available German and American cockroach allergen extracts. *Clin Exp Allergy* 2002; 32:721-7.
35. Sporik R, Squillace SP, Ingram JM, Rakes G, Honsinger RW, Platts-Mills TAE. Mite, cat, and cockroach exposure, allergen sensitisation, and asthma in children: a case-control study of three schools. *Thorax* 1999;54:675-80.
36. Gore JC, Schal C. Gene expression and tissue distribution of the major human allergen Bla g 1 in the German cockroach, *Blattella germanica* L. (Dictyoptera: Blattellidae). *J Med Entomol* 2004;41:953-60.
37. Gore JC, Schal C. Expression, production and excretion of Bla g 1, a major human allergen, in relation to food intake in the German cockroach, *Blattella germanica*. *Med Vet Entomol* 2005;19:127-34.
38. Suazo A, Gore C, Schal C. RNA interference-mediated knock-down of Bla g 1 in the German cockroach, *Blattella germanica* L., implicates this allergen-encoding gene in digestion and nutrient absorption. *Insect Mol Biol* 2009;18:727-36.
39. Nolan T, Petris E, Muller HM, Cronin A, Catteruccia F, Crisanti A. Analysis of two novel midgut-specific promoters driving transgene expression in *Anopheles stephensi* mosquitoes. *PLoS One* 2011;6:e16471.
40. Shao L, Devenport M, Fujioka H, Ghosh A, Jacobs-Lorena M. Identification and characterization of a novel peritrophic matrix protein, Ae-Aper50, and the microvillar membrane protein, AEG12, from the mosquito, *Aedes aegypti*. *Insect Biochem Mol Biol* 2005;35:947-59.
41. Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, Mitchell-Olds T, et al. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proc Natl Acad Sci U S A* 2004;101:4859-64.
42. Karp CL. Guilt by intimate association: what makes an allergen an allergen? *J Allergy Clin Immunol* 2010;125:955-60.
43. Huang SR, Rutkowski JM, Snodgrass RG, Ono-Moore KD, Schneider DA, Newman JW, et al. Saturated fatty acids activate TLR-mediated proinflammatory signaling pathways. *J Lipid Res* 2012;53:2002-13.
44. Eisenbarth SC, Piggott DA, Huleatt JW, Visintin I, Herrick CA, Bottomly K. Lipopolysaccharide-enhanced, toll-like receptor 4-dependent T helper cell type 2 responses to inhaled antigen. *J Exp Med* 2002;196:1645-51.
45. Piggott DA, Eisenbarth SC, Xu L, Constant SL, Huleatt JW, Herrick CA, et al. MyD88-dependent induction of allergic Th2 responses to intranasal antigen. *J Clin Invest* 2005;115:459-67.
46. Chisholm D, Libet L, Hayashi T, Horner AA. Airway peptidoglycan and immunostimulatory DNA exposures have divergent effects on the development of airway allergen hypersensitivities. *J Allergy Clin Immunol* 2004;113:448-54.
47. Redecke V, Hacker H, Datta SK, Fermin A, Pitha PM, Broide DH, et al. Cutting edge: activation of Toll-like receptor 2 induces a Th2 immune response and promotes experimental asthma. *J Immunol* 2004;172:2739-43.
48. Tsai WJ, Liu CH, Chen ST, Yang CY. Identification of the antigenic determinants of the American cockroach allergen Per a 1 by error-prone PCR. *J Immunol Methods* 2003;276:163-74.
49. Wu CH, Lee MF, Yang JS, Tseng CY. IgE-binding epitopes of the American cockroach Per a 1 allergen. *Mol Immunol* 2002;39:459-64.
50. Diraphat P, Sookrun N, Chaicumpa W, Pumhirun P, Vichyanond P, Tapchaisri P, et al. Recombinant American cockroach component, Per a 1, reactive to IgE of allergic Thai patients. *Asian Pac J Allergy Immunol* 2003;21:11-20.

METHODS

Sera from patients with cockroach allergy

Sera from patients with cockroach allergy were from previous emergency department asthma studies carried out at the University of Virginia.^{E1,E2} Sera from patients with cockroach allergy were also kindly provided by Dr Robert Wood (Johns Hopkins University, Baltimore, Md) as part of a collaborative study with the Inner-City Asthma Consortium (National Institute of Allergy and Infectious diseases protocol no. ICAC-18). All studies were approved by the Human Investigation Committees of the University of Virginia or Johns Hopkins University. Total IgE antibody levels in the sera from University of Virginia patients ranged from 6 to 640 ng/mL, with a median of 63 ng/mL, as measured by using RAST.^{E1,E2} The ranges of allergen-specific IgE levels were 4 to 318 ng/mL (median, 106 ng/mL) for Bla g 1 and 0.4 to 65 ng/mL (median, 17 ng/mL) for Bla g 2, as measured by using an mAb-based RIA.^{E3} Finally, 5 sera were obtained at Johns Hopkins University from Baltimore-area patients with a history of allergic rhinitis, asthma, and sensitivity to cockroach. The 5 sera had total IgE levels ranging from 35 to 1852 kU/L and cockroach-specific IgE levels ranging from 0.59 to greater than 100 kU/L; Bla g 1-specific IgE had a median level of 0.85 kU/L (range, 0.35–29.6 ng/mL) and Bla g 2-specific IgE had a median level of greater than 0.35 for 2 of the sera (0.99 and 9.03 kU/L), as measured by using ImmunoCAP (Phadia US, Portage, Mich).^{E4}

Constructs and purification

The PET28a-yEGFP expression vector was kindly provided by Dr Kawasaki.^{E5} This vector uses a yeast-optimized gene of GFP with the specific mutations S65G and S74A to enhance fluorescence.^{E6} We modified the vector by removing the C-terminal Thr of GFP and inserting a multicloning site identical to that used in a maltose binding protein fusion crystallization system immediately downstream of the GFP (Fig E2).^{E7} The gene for Blag1 1.0101 (GenBank accession no. AF072219) repeats 1 and 2 (residues 34–216) was synthesized and inserted by using the *NotI* and *BamHI* restriction sites, thus resulting in a protein linker region of 3 alanine residues. In addition, the thrombin cleavage site between the N-terminal 6xHis-tag and the GFP of the original vector was replaced by a TEV cleavage site. This vector was transformed into BL21 (DE3) Codon Plus RIL cells. Cells containing the 6xHis-GFP-Bla g 1 vector were grown in 1 L of Luria broth (LB) in Fernbach flasks at 37°C. When the OD reached 0.8, the temperature was decreased to 18°C, expression was induced by addition of Isopropyl β-D-1-thiogalactopyranoside to 0.4 mmol/L, and expression proceeded overnight. Cells were pelleted and resuspended in 25 mmol/L Tris (pH 7.5), 10 mmol/L imidazole, and 500 mmol/L NaCl, followed by sonication. The lysate was pelleted by means of centrifugation, and the soluble fraction was added to Ni-NTA resin in batches. Resin was washed and loaded into a column with sonication buffer. 6xHis-GFP-Bla g 1 was eluted from the resin with a linear imidazole gradient from 10 to 250 mmol/L imidazole. Purified protein was concentrated, and then the buffer was exchanged to 25 mmol/L Tris (pH 7.5), 250 mmol/L NaCl, and 2.5 mmol/L imidazole. TEV protease was added to remove the 6xHis-tag, and the sample was incubated overnight at 4°C. The sample was concentrated and loaded onto a Superdex-200 column equilibrated in 25 mmol/L Tris (pH 7.5) and 250 mmol/L NaCl. The peak containing the rBla g 1-GFP sample was then added to Ni-NTA in batch to remove uncleaved protein. The rBla g 1-GFP protein sample was collected, and the buffer was exchanged to the final conditions of 25 mmol/L Tris (pH 7.5), 125 mmol/L NaCl, and 2.5 mmol/L imidazole. The protein was concentrated to 21 mg/mL.

Crystallography

Crystals of rBla g 1-GFP were grown by using the sitting drop vapor diffusion method by mixing 1 μL of 21 mg/mL protein with 1 μL of the reservoir solution. For the 2.8-Å resolution data set, the reservoir solution consisted of 100 mmol/L Tris (pH 7.5), 10% vol/vol propanol, and 20% wt/vol polyethylene glycol (PEG) 4K. For the 2.4-Å resolution data set, the previous reservoir was diluted by 30% with water. For data collection, both crystals were transferred to a cryoprotective solution consisting of 90 mmol/L Tris (pH 7.5), 9% vol/vol propanol, 18% wt/vol PEG 4K, and 10% ethylene glycol and flash-frozen in liquid nitrogen. Both data sets were collected at the Advanced

Photon Source by using the SER-CAT 22 beamline, and data were processed with HKL2000.^{E8} A model of GFP (PDB code 3AI4^{E1}) was used for molecular replacement in Phenix against the 2.8-Å data set.^{E9} Initial model building of the Bla g 1 repeat was carried out in Coot, followed by successive rounds of model building and refinement in Coot and Phenix.^{E10} This model was used for further refinement against the 2.4-Å resolution data set to obtain the final model. Table E1 shows crystallographic statistics.

MS

MS analyses of the lipids were performed on an API 3000 triple quadrupole mass spectrometer (AB Sciex, Farmingham, Mass) equipped with an electrospray source operated in both positive and negative ion mode. Lipids were extracted with 2:1 chloroform/methanol and reconstituted in ethanol for analysis.^{E11} Chromatography was performed on an Agilent Zorbax XDB-C18 column (2.1 mm ID × 50 mm; Agilent, Santa Clara, Calif) with an ammonium acetate/methanol gradient at a flow rate of 250 μL/min. Phospholipid standards were obtained from Indofine Chemical Company (Hillsborough, NJ). Positive and negative tandem MS experiments were used to determine head groups and fatty acid chains of the phospholipids.

Molecular weights of breakdown products of rBla g 1 were determined by means of electrospray ionization MS with a Micromass Q-ToF micro mass spectrometer and flow injection from a pressurized bomb. Peptide coverage maps of the different species of Bla g 1 observed by using gel electrophoresis were generated by means of in-gel digestion with either trypsin or chymotrypsin, followed by MALDI-TOF MS and MS/MS with an AB Sciex 4800 Plus MALDI TOF/TOF Analyzer.

NMR

Both rBla g 1-EC and rBla g 1-PP NMR samples were prepared by diluting the purified proteins 1:1 with 10% cholate and 100 mmol/L Tris (pH 8.0). The *E coli* lipid extract was prepared with a 1-mL sample of *E coli* grown to saturation in LB, sonicated with six 0.5-second pulses, and centrifuged to remove the soluble fraction. The insoluble fraction was resuspended in 0.25 mL of water, 0.25 mL of PBS, and 0.5 mL of 10% cholate and 100 mmol/L Tris (pH 8.0). *P pastoris* lipid extract was prepared with a 1-mL sample of *P pastoris* grown overnight, pelleted, and resuspended in 0.5 mL of PBS and 0.5 mL of glass beads (Sigma, St Louis, Mo). The sample was vortexed for 5 minutes, and the supernatant was extracted with multiple washes of the beads with PBS. The insoluble fraction was pelleted and resuspended in the same NMR buffer as the *E coli* extract. ³¹P NMR was performed on a 14T magnet with a broadband probe tuned to the ³¹P resonant frequency of 243 MHz and used broadband ¹H decoupling.

RESULTS

MS analysis of rBla g 1-PP

The gel in Fig 3, B, indicated that the rBla g 1-PP (lot 34074) breaks down at multiple sites. We hypothesized that these fragments correspond to partial proteolysis of the protein, with cleavage sites in unstructured loops between the “beads” and the 2 unstructured loops located between the repeats (Fig E4). Electrospray MS of the protein solution found intact masses around 10 kDa that correspond uniquely to each of the 4 possible fragments, assuming the cut sites are in these loops (Table E2). MS and MS/MS analysis of tryptic and chymotryptic digests of the 9-kDa band confirmed unique peptides from each of the expected fragments (data not shown). A similar analysis of the 30-kDa band confirmed the presence of unique peptides from pentagonal structures 2, 3, and 4. A peptide comprised of those 3 subdomains would be approximately 30 kDa.

nBla g 1 was reported to contain fragments resulting from trypsin cleavage of the protein at the beginning of the 100 amino acid repeats and close to the cleavage sites found for rBla g 1-PP.^{E12,E13} In other words, both Bla g 1-PP and nBla g 1 appear to show similar fragmented patterns on denaturing SDS-PAGE.

REFERENCES

- E1. Gelber LE, Seltzer LH, Bouzoukis JK, Pollart SM, Chapman MD, Platts-Mills TAE. Sensitization and exposure to indoor allergens as risk-factors for asthma among patients presenting to hospital. *Am Rev Respir Dis* 1993;147:573-8.
- E2. Pollart SM, Chapman MD, Fiocco GP, Rose G, Platts-Mills TAE. Epidemiology of acute asthma—IgE antibodies to common inhalant allergens as a risk factor for emergency room visits. *J Allergy Clin Immunol* 1989;83:875-82.
- E3. Arruda LK, Vailes LD, Mann BJ, Shannon J, Fox JW, Vedvick TS, et al. Molecular cloning of a major cockroach (*Blattella germanica*) allergen, Bla g 2—sequence homology to the aspartic proteases. *J Biol Chem* 1995;270:19563-8.
- E4. Oseroff C, Sidney J, Tripple V, Grey H, Wood R, Broide DH, et al. Analysis of T cell responses to the major allergens from German cockroach: epitope specificity and relationship to IgE production. *J Immunol* 2012;189:679-88.
- E5. Suzuki N, Hiraki M, Yamada Y, Matsugaki N, Igarashi N, Kato R, et al. Crystallization of small proteins assisted by green fluorescent protein. *Acta Crystallogr D Biol Crystallogr* 2010;66:1059-66.
- E6. Cormack BP, Bertram G, Egerton M, Gow NAR, Falkow S, Brown AJP. Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in *Candida albicans*. *Microbiology* 1997;143:303-11.
- E7. Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci* 2010;19:901-13.
- E8. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology, Macromolecular Crystallography, Part A* 1997;276:307-26.
- E9. Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. Automated structure solution with the PHENIX suite. *Methods Mol Biol* 2008;426:419-35.
- E10. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126-32.
- E11. Folch J, Lees M, Stanley GHS. A simple method for the isolation and purification of total lipids from animal tissues. *J Biol Chem* 1957;226:497-509.
- E12. Pomés A, Melen E, Vailes LD, Retief JD, Arruda LK, Chapman MD. Novel allergen structures with tandem amino acid repeats derived from German and American cockroach. *J Biol Chem* 1998;273:30801-7.
- E13. Pomés A, Vailes LD, Helm RM, Chapman MD. IgE reactivity of tandem repeats derived from cockroach allergen, Bla g 1. *Eur J Biochem* 2002;269:3086-92.

FIG E1. Sequence alignments. **A**, The alignment of the Bla g 1.0101 repeats is colored according to the helix number, which was determined from the crystal structure. The coloring is equivalent to that of Fig 2. **B**, Bla g 1 repeats 1 and 2 are aligned with Bla g 1.02, Per a 1.0101, and ANG12 from *Anopheles gambiae*; AEG12 from *Aedes aegypti*; a similar protein from *Tenebrio molitor*; and an NSP from *P. rapae*.

```

AGGGTACAATTCCCTCTAGCAATAATTTGTTAACTTTAAGAGGAGATATACCATGGGCAGCAGCCATCATCATCATCACAGC
                                     RBS                               NcoI           6xHis
                                     *                               MetGlySerSerHisHisHisHisHisHisSer

AGCGGCCGAAAACCTGTATTTTCAG  GGCTCGCATATGTCTAAAGGTGAAGAATTATTCACTGGCGTTGTCCCAATTTGGTTGAATTA...
TEV                               NdeI                               yEGFP
SerGlyGluAsnLeuTyrPheGln  *GlySerHisMetSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeu...

                               NotI   NheI  EcoRI  BamHI  XbaI  SalI   PstI  HindIII
...GTTACTGCTGCTGGTATTGCGGCCGCAGCTAGCGAATTCGGATCCTCTAGAGTCGACCTGCAGGCAAGCTTGATCGTGACTGACTGA
...ValThrAlaAlaGlyIleAlaAlaAlaAlaSerGluPheGlySerSerLysValAspLeuGlnAlaSerLeu

```

FIG E2. The multicloning site (MCS) of pEGFPX. The original vector was the pET28a-yEGFP kindly provided by Dr Kawasaki^{E1} and was modified in the following ways: the thrombin cleavage site downstream of the 6xHis tag (blue) was replaced by a TEV protease site (purple, cleavage site marked by purple asterisk), and the pET28a MCS was replaced by a sequence of the MCS from the pMALX vector.^{E3} The terminal threonine residue of the EGFP was deleted to decrease the length of the linker between the EGFP (orange) and the protein of interest. The *XbaI* restriction site (gray) upstream of the ribosome-binding site was deactivated (A to C substitution, marked by gray asterisk). Finally, 3 TGA stop codons (red) were added downstream of the MCS in each reading frame.

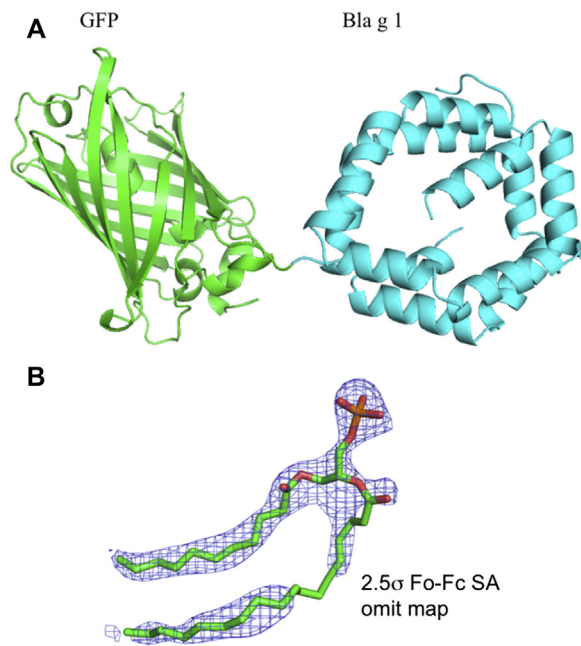


FIG E3. Additional structure figures. **A**, Structure of the GFP-Bla g 1 asymmetric unit in the crystal structure. The 2 moieties do not make significant contact. **B**, The 2.5 σ Fo-Fc SA omit map for the proposed phosphatidic acid molecule found in the central cavity of Bla g 1.

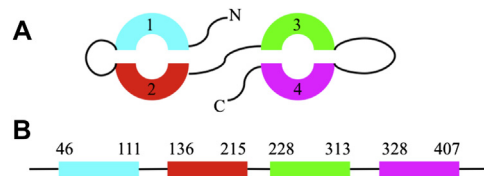


FIG E4. Model for the fragmentation pattern of rBla g 1-PP in Fig 3. **A**, rBla g 1-PP is composed of 2 beads (1 + 2 and 3 + 4) with unstructured loops (*black lines*) between the repeats. **B**, Residue numbering corresponding to the structured and unstructured elements.

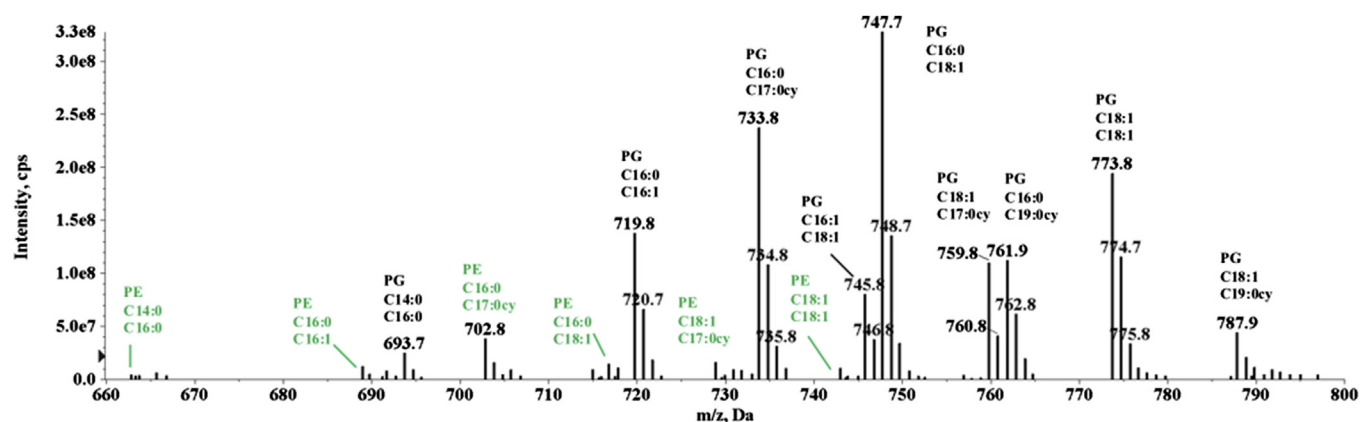


FIG E5. MS data of lipids in *E. coli* expressed Bla g 1. The mass/charge ratio is plotted against the signal intensity, with annotations of lipids identified from tandem MS: PE lipids (green) and PG lipids (black). The mass spectrum is the sum of 20 scans using negative ion-flow injection.

TABLE E1. Crystallographic data statistics

Data set	rBla g 1–GFP (high)	rBla g 1–GFP (low)
Unit cell	a = 91.77 Å, b = 141.92 Å, c = 93.47 Å; $\alpha = \beta = \gamma = 90^\circ$	a = 92.67 Å, b = 141.62 Å, c = 93.53 Å; $\alpha = \gamma = 90^\circ$, $\beta = 90.4^\circ$
Space group	C222 ₁	C2
Resolution (Å)	50.0–2.4	50–2.8
No. of observations	165,823	103,638
Unique reflections	23,505	29,002
Rsym (% [last shell])*	8.9 (45.8)	6.8 (25.1)
I/ σ I (last shell)	14.9 (2.5)	11.6 (2.9)
Mosaicity range	0.69–0.93	0.27–0.58
Completeness (% [last shell])	99.3 (94.0)	97.8 (79.4)
Refinement statistics		
Rcryst (%)†	19.8	
Rfree (%)‡	24.6	
No. of waters	100	
Overall mean B (Å ²) average for:	58.2	
GFP	51.4	
Bla g 1	67.6	
Ligands	75.3	
Water	51.8	
Root-mean-square deviation from ideal values		
Bond length (Å)	0.007	
Bond angle (°)	1.1	
Dihedral angle (°)	16.8	
Ramachandran statistics§		
Residues in:		
Favored (98%) regions (%)	99.2	
Allowed (>99.8%) regions (%)	100	

*Rsym = $\sum (|I_i - \langle I \rangle|) / \sum \langle I_i \rangle$, where I_i is the intensity of the i th observation and $\langle I \rangle$ is the mean intensity of the reflection.

†Rcryst = $\sum ||F_o| - |F_c|| / \sum |F_o|$, as calculated from the working data set.

‡Rfree was calculated from 5% of data randomly chosen not to be included in refinement.

§Ramachandran results were determined by using MolProbity (<http://molprobity.biochem.duke.edu/>).

TABLE E2. MS analysis of rBla g 1-PP

Mass (Da)	Repeat(s)	Residues (Bla g 1.0101)
9,811.4	1 or 4	23-112 or 318-403
9,967.8	1	38-128
10,099.0	4	322-411
10,237.1	2	108-198
10,284.2	2	133-225
10,374.0	1, 3, or 4	38-131, 221-315, or 321-412
10,505.0	2 or 3	129-221 or 218-313
10,666.3	3	224-320
10,965.0	1 or 4	12-111 or 316-412
11,260.8	1	19-121