Molecular Immunology 60 (2014) 86–94

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

The major cockroach allergen Bla g 4 binds tyramine and octopamine

Lesa R. Offermann [a], Siew Leong Chan [b], Tomasz Osinski [c], Yih Wan Tan [b], Fook Tim Chew [b], J. Sivaraman [b], Yu-Keung Mok [b,***], Wladek Minor [c,**], Maksymilian Chruszcz [a,*]

[a] Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208, USA
[b] Department of Biological Sciences, National University of Singapore, Singapore
[c] Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, USA

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form 27 March 2014
Accepted 31 March 2014
Available online 24 April 2014

Keywords:
Cockroach allergen
Bla g 4
Per a 4
Tyramine
Octopamine
Allergy

ABSTRACT

Bla g 4 is a male cockroach specific protein and is one of the major allergens produced by *Blattella germanica* (German cockroach). This protein belongs to the lipocalin family that comprises a set of proteins that characteristically bind small hydrophobic molecules and play a role in a number of processes such as: retinoid and pheromone transport, prostaglandin synthesis and mammalian immune response. Using NMR and isothermal titration calorimetry we demonstrated that Bla g 4 binds tyramine and octopamine in solution. In addition, crystal structure analysis of the complex revealed details of tyramine binding. As tyramine and octopamine play important roles in invertebrates, and are counterparts to vertebrate adrenergic transmitters, we speculate that these molecules are physiological ligands for Bla g 4. The nature of binding these ligands to Bla g 4 sheds light on the possible biological function of the protein. In addition, we performed a large-scale analysis of Bla g 4 and Per a 4 (an allergen from American cockroach) homologs to get insights into the function of these proteins. This analysis together with a structural comparison of Blag 4 and Per a 4 suggests that these proteins may play different roles and most likely bind different ligands.

Accession numbers: The atomic coordinates and the structure factors have been deposited to the Protein Data Band under accession codes: 4N7C for native Bla g 4 and 4N7D for the Se-Met Bla g 4 structure.

1. Introduction

According to the National Center for Health Statistics, more than 25 million Americans (8.4% of the US population) had asthma in 2010 and the prevalence of the disease continues to increase. Moreover, asthma is more prevalent among children, women, African–Americans, Native Americans, and those with family incomes below the poverty line. While the disease can be triggered by many different environmental factors, exposure to cockroach allergens has been identified in numerous studies as a prominent risk factor for asthma (Call et al., 1992; Cohn et al., 2006; Eggleston et al., 1998; Gelber et al., 1993; Matsui et al., 2003). Initial studies investigating the role of cockroach allergies in asthma focused on inner city populations, but subsequent evidence suggests that the problem is more widespread. Surveys have estimated that measurable concentrations of cockroach allergens are present in a majority of all US households, and that the rate of sensitization to cockroach allergens among suburban middle-class children with asthma is much larger than previously suspected (Cohn et al., 2006; Matsui et al., 2003).

Abbreviations: Se-Met, selenomethionine; PDB, Protein Data Bank; rmsd, root mean square derivative; HSQC NMR, heteronuclear single quantum correlation nuclear magnetic resonance; Tdp1, human tyrosyl-DNA phosphodiesterase; PNMT, human phenylethanolamine *N*-methyltransfrase; ITC, isothermal titration calorimetry; IgE, immunoglobulin E; PSSM, position-specific scoring matrix; e-value, expectation value; MAD, multi-wavelength anomalous diffraction; MR, molecular replacement.

* Corresponding author at: Department of Chemistry and Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC 29208, USA. Tel.: +1-803-777-7399; fax: +1-803-777-9521.
** Corresponding author at: Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA. Tel.: +1-434-243-6865; fax: +1-434-982-1616.
*** Corresponding author at: Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore. Tel.: +65-6516-2967; fax: +65-6779-2486.
E-mail addresses: dbsmokh@nus.edu.sg (Y.-K. Mok), wladek@iwonka.med.virginia.edu (W. Minor), chruszcz@mailbox.sc.edu (M. Chruszcz).

http://dx.doi.org/10.1016/j.molimm.2014.03.016
0161-5890/

Bernton and Brown (1964) were the first to link cockroaches with allergic disease. Since then multiple cockroach-allergen specific proteins have been identified (Arruda et al., 1995a,b; Arruda et al., 1997; Helm et al., 1996; Santos et al., 1999). Cockroach allergens are produced mainly by two species: *Blattella germanica* (German cockroach) and *Periplaneta americana* (American cockroach). The German cockroach is most commonly found in Europe and the US, while the American cockroach is predominantly found in South America and some regions in Asia. There are 14 cockroach allergens officially registered on the WHO/IUIS list of Allergen Nomenclature (www.allergen.org), eight from *B. germanica* and six from *P. americana*. Among all patients sensitive to *B. germanica* allergens, 95% are sensitized to one (or more) of four allergens: Bla g 1, Bla g 2, Bla g 4, and Bla g 5. Of these allergen proteins, Bla g 1, Bla g 2 and Bla g 4 have been structurally characterized. Bla g 1 is suggested to have a digestive function and represent a novel fold with the capacity to bind lipids (Mueller et al., 2013). Bla g 2 is an atypical aspartic protease (Gustchina et al., 2005), while Bla g 4 is a lipocalin (Arruda et al., 1995a; Tan et al., 2009).

The lipocalin family comprises a set of proteins that characteristically bind small hydrophobic molecules and play a role in a number of processes, such as retinoid and pheromone transport, prostaglandin synthesis and mammalian immune response. Lipocalins are typified by their ability to bind small molecules; however, no ligands that bind Bla g 4 had been previously identified and the function of the protein is unknown. Bla g 4 is produced in the conglobate gland and apical utricles of the male reproductive system of *B. germanica* (Fan et al., 2005). Bla g 4 is later transferred along with spermatophore to the female's genital tract during copulation. The fate of Bla g 4 in females is unknown, but it has been demonstrated that the immunoreactivity of Bla g 4 disappears 24 h after mating (Fan et al., 2005; Gore and Schal, 2007).

In this paper, we demonstrate that Bla g 4 specifically binds two biogenic amines, tyramine and octopamine. Additionally, we identify the specific binding mode of tyramine with Bla g 4 by X-ray diffraction. Tyramine and octopamine play important roles in invertebrates, and are counterparts to vertebrate adrenergic transmitters. The nature of the binding of Bla g 4 with these ligands sheds light on the possible biological function of the protein.

2. Materials and methods

2.1. Sequence analysis

Sequences were obtained by running PSI-BLAST (Altschul et al., 1997) against the Uniprot database (Uniprot version: 2013_2) (UniProt, 2012) using the sequences of Bla g 4 and Per a 4 (a close homologous protein to Bla g 4 from *P. americana*) as the queries. As the first step, position-specific scoring matrix (PSSM) profiles were created by performing searches with Bla g 4 and Per a 4 as queries with an expectation value (*e*-value) of 10^{-5} for three cycles. As the second step, PSSM profiles were used to perform searches against the same database as for the first step with an *e*-value of 10^{-3} until convergence was achieved. Protein structures, homologous to Bla g 4 and Per a 4, obtained for structure analysis were added to the sequence dataset. Identification of a particular sequence PFAM (Punta et al., 2012) membership was achieved by creating a BLAST database from the PFAMs used as source (PF00061—Lipocalin, PF08212—Lipocalin-like, PG03973—Triabin; PFAM database version: 26.0) to create two (AF015—Lipocalin, AF119—Triabin family) AllFam (Radauer et al., 2008) allergen families (AllFam database version: 2011–09-12). Sequences obtained from searches against UniProt were subjected to CD-HIT (Fu et al., 2012), where sequences with 80% identity or higher were removed. The created dataset was merged with results from searches against PFAM and pdbaa databases, ultimately returning 1561 non-redundant protein sequences belonging to Lipocalin, Lipocalin-like and Triabin PFAM family and sequences obtained for structural analysis. CLANS (Frickey and Lupas, 2004) was used to create 2D visualization of sequences pairwise similarity by using the Fruchterman–Reingold graph layout algorithm. Clustering was performed with an *e*-value of 10^{-6} until convergence was achieved. Allergens found in identified clusters were aligned with MAFFT (Katoh and Standley, 2013) with the L-INS-I option and later adjusted manually in Jalview (Waterhouse et al., 2009) according to the 2D projection of the structural alignment of representative allergens found in the sequence dataset—Bla g 4, Per a 4, Can f 2 and Equ c 1 (PDB IDs: 3EBK, 3EBW, 3L4R and 1EW3) created in Swiss-PdbViewer (Guex and Peitsch, 1997).

2.2. Evolutionary analysis

Sequences from AllFam families AF015 and AF119 were mapped on the dataset used for clustering, then were aligned by MAFFT (Katoh and Standley, 2013) with the L-INS-I option to increase accuracy. The obtained sequence alignment was subjected to MEGA5 (Tamura et al., 2011). Phylogeny reconstruction was performed using the maximum-likelihood estimation with WAG (Whelan and Goldman, 2001) amino acid substitution model with gamma-distributed rates among patterns. The bootstrap method with 1000 replications was used to test branch probabilities.

2.3. Structure analysis

Representation of protein structures, homologous to Bla g 4 and Per a 4, were obtained by performing a PSI-BLAST search against the pdbaa. NCBI BLAST database (as of January 2013), was used to create a structural alignment in the VMD (Humphrey et al., 1996) program. Homology between protein structures was measured by using the Q_H algorithm (O'Donoghue and Luthey-Schulten, 2003). Calculated Q_H values for given residues (Q_{RES}) were then applied to the Bla g 4 structure, instead of B-factor values, and displayed in Pymol (DeLano, 2002).

2.4. Structure determination

Protein production, crystallization, and data collection have been described previously (Tan et al., 2009). Here we present the reinterpretation of diffraction data using a new methodology included in the HKL-3000 package (Minor et al., 2006). During these studies, we reinvestigated data obtained from both Se-Met labeled and native Bla g 4 crystals. The Se-Met and native structures were determined using the multi-wavelength anomalous diffraction (MAD) technique and Molecular Replacement (MR), respectively, by HKL-3000 coupled with SHELXD/C/E (Sheldrick, 2008), MLPHARE (Otwinowski, 1991), DM (Cowtan and Main, 1993), ARP/wARP (Perrakis et al., 1999), MOLREP (Vagin and Teplyakov, 1997), SOLVE/RESOLVE (Terwilliger, 2004), and selected programs from the CCP4 package (Winn et al., 2011). Both the Se-Met derivative and the native crystal structures were re-examined in the $P4_12_12$ space group. Models were later refined with REFMAC (Murshudov et al., 2011) and COOT (Emsley and Cowtan, 2004), TLS groups were determined with the TLSMD server (Painter and Merritt, 2006), and structure validation was performed using MOLPROBITY (Chen and Arendall, 2010) and ADIT (Yang et al., 2004). Structures and structure factors were deposited to the PDB (Berman et al., 2000) with accession code 4N7D and 4N7C for Se-Met derivative and native Bla g 4, respectively. Refinement statistics are summarized in Table 1.