Iranian J Publ Health, Vol. 39, No.3, 2010, pp. 109-115

Original Article

In Silico Identification of Potential American Cockroach (*Periplaneta americana)* Allergens

**A Ahmed, K Minhas, Namood-e-Sahar, O Aftab, F Sher Khan*

Dept. of Biosciences, COMSATS Institute of Information Technology, Chak Shahzad, Park Road, Islamabad 44000, Pakistan

(Received 6 Feb 2010; accepted 19Jul 2010)

Abstract

Background: Cockroaches have been recognized as a powerful indoor allergen. Cockroach allergy can be a major factor in serious asthma and nasal allergy. Bioinformatics tools have been developed to identify potential allergens. The present study was conducted to identify potential allergens in *Periplaneta americana* (Linnaeus).

Methods: The study focused on the identification of potential allergens among the characterized proteins of *P. americana* using web-based and publicly available allergen prediction tools that follow the FAO/WHO guidelines for prediction of allergenic proteins. *P. americana* protein sequences were retrieved from UniProtKB. The sequences obtained were analyzed using AlgPred. The potential allergens obtained were further analyzed by SDAP for confirmation.

Results: Protein sequences (233 cases) of *P. americana* were obtained from UniProtKB out of which 25 were known allergens. Out of the remaining 208 proteins, 102 potential allergens were predicted by AlgPred. However, only 9 were found to be potential allergens after screening with SDAP. Arginine kinases, RNA polymerase II subunit, parcxpwnx02, peptidyl-prolyl cis-trans isomerase, hemocyanin subunit type I and type II, homologue of Sarcophaga proteinase and alpha amylase were confirmed to be potential allergens by SDAP.

Conclusion: We have identified nine potential allergens in *P. americana* that may be used as preliminary support for further laboratory experiments.

Keywords: *Allergen, Bioinformatics, Prediction,* Periplaneta americana

Introduction

Allergic disorders are among the most common chronic diseases in the developed world and an increasing problem in developing countries (1). Allergy is caused by the exaggerated and harmful response of the immune system to otherwise harmless substances known as allergens. There is overwhelming evidence that indoor domestic allergens play a key role in allergic disease. The primary arthropod allergens associated with allergic disease are house dust mites and cockroaches (2, 3).

Cockroach allergens are one of the strongest risk factors predictive of allergic sensitization and asthma morbidity (4, 5). Cockroach extracts including cast skins, egg shells, and fecal material (6) have been shown to contain several major and minor allergens (7-10). Among the 3,500 known species of cockroaches, the American cockroach (*Periplaneta americana*) is a frequently encountered cockroach in homes. Allergens with masses ranging from 6 to 120 kDa from *P. americana* have been identified by various immunochemical techniques (11) and the functional importance of some of these have been determined. Two prominent proteins of 78 and 72 kDa in Per a 3 have been reported to cause T cell proliferation in cockroach allergic patients (12). More recent data indicate that indoor insect allergens, including those of cockroaches, are potent inducers of IL-5 and eotaxin-mediated esophageal eosinophilia (13).

Despite efforts of researchers, our current knowledge about *P. americana* allergens and their cross-reactivity is still insufficient, at least, partly due to the difficulty involved in purifying cockroach allergens from extracts in significant quantities for detailed characterization. According to Universal

***Corresponding author:** E-mail: asifa_ahmed@comsats.edu.pk

Protein Database (UniprotKB) resource, to-date a total of 233 proteins have been identified in the *P. Americana* of which only 25 are known reported allergens, which suggests that many allergens may still lie unidentified.

The use of bioinformatics tools is becoming increasingly helpful for initially screening of compounds based on existing experimentally validated databases. Algpred is a web server for prediction of allergenic proteins and for mapping IgE epitopes on allergenic proteins with high accuracy (14). SDAP (Structural Database of Allergenic Proteins) is another web server containing information on allergens and can develop correlations that can be used to predict allergenicity of novel proteins and cross-reactivity between allergens (15).

The present study is focused on the identification of potential allergens using these allergen databases and bioinformatics.

Materials and Methods

Amino acid sequences of 233 proteins belonging to *P. americana* in UniprotKB database were retrieved. Out of these, 25 were known reported allergens. For the remaining 208 proteins, allergenicity test was conducted using AlgPred. The FASTA sequences of proteins were given individually to the server, which presented the results on the basis of the scanning of IgE epitopes, motif-based approach, SVM-based method using amino acid composition of protein, hybrid approach and BLAST search on allergen representative proteins ARPs (14).

The potential allergens obtained through AlgPred screening were further analyzed using SDAP (15). Searches were first carried out according to the FAO/WHO criterion for allergenicity prediction. Proteins that shared more than 35% sequence similarity with an allergen (on a segment of 80 residues) or an identity of at least six contiguous amino acids were screened out. These allergens were further analyzed using the property distance function (PD) method. A protein that passed the initial step and gave a PD score of less than 10 was considered to be potential allergen.

Results

Following screening of the 208 *P. americana* proteins with AlgPred (excluding very short peptides which the AlgPred was unable to screen), a total of 102 proteins were considered potentially allergenic. Further analysis of these 102 proteins with any of the two FAO/WHO criterion of SDAP indicated that 93 proteins did not fulfill both criteria (Table 1). According to the PD scores obtained 9 proteins were predicted to be potential allergens (Table 2). Alpha-amylase (Fragment), parcxpwnx02, homologue of sarcophaga 26, 29 kDa proteinase, RNA polymerase II largest subunit (fragment), and hemocyanin subunit type I and II were predicted to be potentially significant allergens as they had a PD score of less than 10, whereas the arginine kinases and peptidyl-prolyl cis-trans isomerase (fragment) passed all the screening steps and showed to have PD scores of less then 3 which indicated their being highly significant potential allergens.

Of all the predicted allergens, Parcxpwnx02 was positive with IgE mapping searches, and was found to contain an IgE epitope starting from the position 307 of the protein with the sequence LANSWNYDWGDNGY.

Table 1: Proteins showing potential allergenicity with only AlgPred analysis

No.	Protein name	Accession Number
1.	Diuretic Hormone	**P41538**
2.	Corazonin	**P11496**
3.	Bursicon (Fragments)	**P84118**
4.	Pyrokinin-6	**P82693**
5.	Peptide Hormone 4	**P82697**
6.	Peptide Hormone 3	**P82696**
7.	Peptide Hormone 2	**P82695**

Table 1: Continued…

8.	Peptide Hormone 1	**P82694**
9.	Hemolymph Lipopolysacharide	**P26305**
10.	Trehalase Inhibitor	**P19986**
11.	Troponin T	**Q9XZ71**
12.	Sulfakinin 1	**P36885**
13.	Periviscerokinin-2	**P81555**
14.	Periviscerokinin 2.2	**P84422**
15.	Ubiquitin	**A1E2I6**
16.	Rab5-GTP binding protein	**B6ZLL4**
17.	Notch protein (Fragment)	**B6EBG3**
18.	Delta protein	**B6EBG2**
19.	Odorant-binding protein	**B6E9U9**
20.	Odorant-binding protein (Fragment)	**B6E9U8**
21.	Odorant-binding protein	**B6E9U7**
22.	Cxpwmw03	**Q2LIY7**
23.	Cxpwmw01	**Q1PS51**
24.	MRNA, clone: 1. (Fragment)	**P92056**
25.	MRNA, clone: 2. (Fragment)	**P92055**
26.	MRNA, , clone: 3. (Fragment)	**P92054**
27.	MRNA, , clone: 4. (Fragment)	**P92053**
28.	MRNA, , clone: 5. (Fragment)	**P92052**
29.	Lectin-related protein (Fragment)	**P92050**
30.	Lectin-related protein	**P92049**
31.	Lectin-related protein (Fragment)	**P92048**
32.	Lectin-related protein (Fragment)	**P92047**
33.	26-kDa lectin (Fragment)	**O76155**
34.	Rsp60	**O76153**
35.	P10	**O17447**
36.	Elongation factor 1-alpha (Fragment)	**O02460**
37.	DNA-directed RNA polymerase (Fragment)	**O02459**
38.	Dynamin (Fragment)	**D0UTF4**
39.	Methionine aminopeptidase (Fragment)	**D0UTA5**
40.	Putative uncharacterized protein (Fragment)	**D0UT69**
41.	AMP deaminase (Fragment)	**D0UT23**
42.	Pyrimidine biosynthesis (Fragment)	**D0USX9**
43.	Proteasome subunit (Fragment)	**D0USM1**
44.	F-box protein (Fragment)	**D0US62**
45.	Glu-+ pro-tRNA synthetase (Fragment)	**D0URW2**
46.	Glycogen synthase (Fragment)	**D0UR99**
47.	Gelsolin (Fragment)	**D0UR43**
48.	ATP synthase (Fragment)	**D0UQZ3**
49.	Acetylglucosaminyl-transferase (Fragment)	**D0UQT6**
50.	Clathrin heavy chain (Fragment)	**D0UQ69**
51.	Clathrin heavy chain (Fragment)	**D0UQ18**
52.	Glucosamine phosphate isomerase (Fragment)	**D0UPW5**
53.	GTP-binding protein (Fragment)	**D0UPR4**
54.	Syntaxin (Fragment)	**D0UPE9**
55.	Spliceosome-associated protein (Fragment)	**D0UPA1**
56.	Signal recognition particle (Fragment)	**D0UP45**
57.	Pre-mRNA splicing factor (Fragment)	**D0UNV4**
58.	Alpha-spectrin (Fragment)	**D0UND9**
59.	Alpha-spectrin (Fragment)	**D0UNA0**
60.	Alpha-spectrin (Fragment)	**D0UN65**
61.	Acetyl-CoA carboxylase (Fragment)	**D0UN29**
62.	domain binding protein (Fragment)	**D0UMV2**

Table 1: Countinued…

63.	ATP synthase (Fragment)	**D0UM59**
64.	polymerase subunit 2 (Fragment)	**D0ULU7**
65.	RNA helicase (Fragment)	**D0ULN9**
66.	Protein kinase (Fragment)	**D0ULI4**
67.	Histone deacetylase (Fragment)	**D0ULD3**
68.	Gln-tRNA synthetase (Fragment)	**D0UL28**
69.	Arg methyltransferase (Fragment)	**D0UK69**
70.	Regenectin	**Q9Y098**
71.	NADP-dependent isocitrate dehydrogenase (Fragment)	**Q9XY39**
72.	Putative transcription factor	**Q9U0S0**
73.	10 kDa LEG regeneration protein (Fragment)	**Q9TWV5**
74.	Beta-1,4-glucanase 1 (Fragment)	**Q9NCF3**
75.	Beta-1,4-glucanase 2 (Fragment)	**Q9NCF2**
76.	40S ribosomal protein S12 (Fragment)	**Q8MTJ6**
77.	Rab11 (Fragment)	**Q8MTJ5**
78.	Vitellogenin receptor	**Q8MP02**
79.	Large conductance calcium activated potassium channel pSlo spliceform 1B (Fragment)	**Q8I9V0**
80.	Large conductance calcium activated potassium channel pSlo spliceform 4C (Fragment)	**Q8I9U6**
81.	Large conductance calcium activated potassium channel pSlo spliceform 5B (Fragment)	**Q8I9U5**
82.	Ryanodine receptor pRyR (Fragment)	**Q86LC1**
83.	Elongation factor-2 (Fragment)	**Q6JU91**
84.	RNA polymerase II largest subunit (Fragment)	**Q6JU05**
85.	TRPgamma cation channel	**Q5YJT9**
86.	Parcxpwfx01	**Q5MBV8**
87.	Parcxpwfx02	**Q5MBV7**
88.	Parcxpwnx01	**Q5MBV6**
89.	Parcxpwnx03	**Q5G5C3**
90.	Parcxpwnx04	**Q5G5C1**
91.	Adipokinetic hormone preproprotein	**Q5EY02**
92.	Putative uncharacterized protein (Fragment)	**Q5EY00**
93.	Rieske Fe-S protein (Fragment)	**Q5EXZ9**

Table 2: Potential Allergens Predicted by AlgPred and SDAP in *Periplaneta Americana*

No.	**Protein name**	**Accession Number**	**AlgPred Analysis**	**SDAP Analysis**		
				FAO/WHO Criteria		**PD score**
				Stretches of 6 contiguous amino acid identical to an allergen*	**% identity with an allergen over a window of 80 a.as**	
1.	Arginine kinase [Periplaneta americana]	D3JUE7	Predicted allergen by *SVM-based method* and *BLAST* approach	Present	93.75% with *Bomb m 1.0101* from a.a number 263 to 342	< 3
2.	Alpha-amylase (Fragment)	D2YVM9	Predicted allergen by *SVM-based method*	Present	61.25% with *Blo t 4.0101* from a.a number 35 to 114	< 10
3.	Homologue of Sarcophaga 26,29kDa proteinase	Q9U914	Predicted allergen by SVM-based method	Present	50.00% with Act c 1 from a.a number 466 to 545	> 3 and < 10
4.	Peptidyl-prolyl cis-trans isomerase (Fragment)	Q9U8K2	Predicted allergen by SVM-based method and BLAST approach	Present	77.50% with Mala s 6	< 3

Table 2: Countinued…

5.	Parcxpwnx02	Q5MBV5	Predicted allergen by IgE mapping, BLAST and Hybrid approach	Present	42.50% with Act c 1 from a.a number 257 to 336	> 3 to < 10
6.	RNA polymerase II largest subunit (Fragment)	Q5EXZ8	Predicted allergen by *SVM-based method*	Present	41.25% with *Der f 15* from a.a number 16 to 95	> 3 and < 10
7.	Hemocyanin subunit type II	B9W4N8	Predicted allergen by *SVM-based method*	Present	48.75% with *Per a 3.0201* from a.a number 112 to 191	> 3 and < 10
8.	Hemocyanin subunit type I	B9W4N7	Predicted allergen by *SVM-based method*, *BLAST* and *hybrid approach*	Present	65.00% with *Per a 3.0201* from a.a number 100 to 179	< 3
9.	Arginine kinase	A1KY39	Predicted allergen by *SVM-based method*	Present	93.75% with *Plo i 1* from a.a number 257 to 336	< 3

*present – indicates that there are stretches of 6 contiguous amino acid identical to a known allergen

Discussion

In silico protein analysis is a well-established technique for assessment of allergenicity and immunological cross-reactivity (16, 17).

We screened 208 *P. americana* proteins using AlgPred and SDAP bioinformatics tools. Highly significant prediction score (PD<3) were obtained for two arginine kinases (Accession nos. D3JUE7 and A1KY39). Interestingly, another arginine kinase in *P. americana* (Accession no. B1A7S7) has been documented to be an important allergen in the Thai population (18) that correlates well with our highly significant scores in our study for this enzyme. Arginine kinase isomers have been reported in *Caenorhabditis elegans* and it has been suggested that tissue restricted expression of isoforms in this family evolved early (19). In addition, highly significant homology (93.75%) was seen between the arginine kinase of *P. americana* and that of *Plodia interpunctella* (Indian-meal moth) that also acts as a powerful allergen (16). It is most likely that the two arginine kinases reported in this study are allergens.

The protein, peptidyl-prolyl cis-trans isomerase (accession no. Q9U8K2) also showed significant prediction score (PD<3). Peptidyl-prolyl cis-trans isomerase belongs to the cyclophilin-type PPIase family and shows significant homology (77.5%) with a cyclophilin allergen from the yeast *Malassezia sympodialis* (Mala s 6) (20). Cyclophilins constitute a family of proteins involved in many important cellular functions. They have also been identified as a pan-allergen family able to elicit IgE-mediated hypersensitivity reactions (21, 22).

Homologue of Sarcophaga 26 and 29 kDa proteinase (accession no. Q9U914) and parcxpwnx02 (accession no. Q5MBV5) were predicted as potential allergens due to their peptidase property. Both these proteins showed significant identity, 50% and 42.5% respectively, with allergen Act c 1 (or actinidin), a cysteine protease and also a major allergen in kiwi fruit. This 30 kDa acidic protein is present in kiwi fruit in several isoforms that differ in the *PI* value (23). Homologue of Sarcophaga 26 and 29kDa proteinase appears to eliminate foreign proteins in this insect and is conserved in a wide variety of insects and participates in their defense mechanism (24). German cockroach proteases have been known to play a role as allergens, participating in cleavage of matrix metalloproteinase (MMP-9) thereby remodeling airway passage (25).

Hemocyanin subunit type I and II proteins show 48.75 and 65% identity respectively, with the known cockroach allergen of Per a 3-family. *Per a 3* belongs to the most potent allergens (26).

This result also becomes noteworthy if we consider that certain other proteins with hemocyanin domains are allergens (27, 28).

Another protein RNA polymerase II (accession no. Q5EXZ8) has shown 41.25% identity with a Der f 15 that is a major canine high molecular weight allergen (29). Alpha-amylase (fragment) (accession no. D2YVM9) was also found to be a potential allergen. Fungal α-amylase is a known dust allergen that is commonly found in bakery, in particularly wheat or flour, products (30). There is a possibility that the α-amylase protein of *P. americana* is also a dust allergen associated with the cockroach species. Cockroaches are found in flour and it is highly possible that the dust allergen, α-amylase, is transferred from the flour to the cockroaches.

In conclusion, *in silico* studies are valuable tools for predicting potential proteins should be given priority in allergen research. We have identified 9 proteins of the *P. americana* that are potential allergens and warrant further studies in this area.

Ethical considerations

Ethical issues (Including plagiarism, Informed Consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc) have been completely observed by the authors.

Acknowledgements

The authors declare that they have no conflict of interests.

References

1. Chapman MD (1998). In *Encyclopedia of Immunology* (Roitt IM and Delves PJ eds) 2nd Ed., 1: 64–69, Academic Press, London.
2. Richardson G, Eick S, Jones R (2005). How is the indoor environment related to asthma? literature review. *J Adv Nurs*, 52: 328–39.
3. Wu CH, Lee MF (2005). Molecular Characteristics of Cockroach Allergens. *Cell Mol Immunol*, 2:177-180.
4. Arshad SH (2010). Does Exposure to Indoor Allergens Contribute to the Development of Asthma and Allergy? *Curr Allergy Asthma Rep*, 10:49–55. DOI 10.1007/s 11882-009-0082-6.
5. Arruda LK, Chapman MD (2001). The role of cockroach allergens in asthma. *Curr Opin Pulm Med*, 7:14–19.
6. Peden D, Reed CE (2010). Environmental and occupational allergies. *J Allergy Clin Immunol*, 125: S150-60.
7. Asturias JA, Bayón NG, Arilla MC, Martínez A, Palacios R, Gascón FS, Martínez J (1999). Molecular Characterization of Cross-Reactive Allergen Periplaneta americana Allergen 7, a American Cockroach Tropomyosin. *J Immunol*, 162: 4342-4348.
8. Stankus RP, Hornet WE, Lehrer SB (1990). Identification and characterization of important cockroach allergens. *J Allergy Clin Immunol*, 86:781-6.
9. Jomori T, Natori S (1991). Molecular cloning of cDNA for lipopolysaccharide-binding protein from the hemolymph of American cockroach, *Periplaneta americana*. *J Biol Chem*, 266:13318-13323.
10. Yang CY, Wu JD, Wu CH (2000). Sequence analysis of the first complete cDNA clone encoding an American cockroach Per a 1 allergen. *Biochim Biophys Acta*, 1517: 153-58.
11. Helm RM (1999). Cockroach and other inhalant insect allergens. In: Lockey RF and Bukantz SC, editors. Allergens and allergen immunotherapy. 2nd ed. New York: Marcel Dekker, Inc., 203-223.
12. Jeng KCG, Liu MT, Wu CH, Wong D, Lan JL (1996). American cockroach Cr-PI allergen induces lymphocyte proliferation and cytokine production in atopic patients. *Clin Exp Allergy*, 26:349-356.
13. Rayapudi M, Mavi P, Zhu X, Pandey AK, Abonia JP, Rothenberg ME, Mishra A (2010). Indoor insect allergens are potent inducers of experimental eosinophilic esopha-

gitis in Epub ahead of print-doi: 10.1189/jlb.0110025.
14. Saha S, Raghava GPS (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*, 34: 202-209. www.imtech.res.in/raghava/algpred/.
15. Ivanciuc O, Schein CH, Braun W (2003). SDAP: Database and Computational Tools for Allergenic Proteins. *Nucleic Acids Res*, 31:359-362. http://fermi.utmb.edu/SDAP/.
16. Binder MV, Mahler B, Hayek WR, Sperr M, Scholler S, Prozell G, et al. (2001). Molecular and immunological characterization of arginine kinase from the Indianmeal moth, Plodia interpunctella, a novel cross-reactive invertebrate pan-allergen. *J Immunol*, 167: 5470-77.
17. Flückiger S, Fijten H, Whitley P, Blaser K, Crameri R (2002). Cyclophilins, a new family of cross-reactive allergens. *Eur J Immunol*, 32: 10-17.
18. Sookrung N, Chaicumpa W, Tungtrongchitr A, Vichyanond P, Bunnag C, Ramasoota P, e al. (2006). Periplaneta americana arginine kinase as a major cockroach allergen among Thai patients with major cockroach allergies. *Environ Health Perspect*, 114: 875-80.
19. Andrus VM, Thomas D, Meulia T, Snider M, Fraga D (2008). Characterization of the arginine kinase family of *Caenorhabditis elegans*. *The FASEB Journal*, 22: 1004.
20. Glaser AG, Limacher A, Uckiger FL, Scheynius K, Scapozza L, Crameri R (2006). Analysis of the cross-reactivity and of the 1.5 A° crystal structure of the Malassezia sympodialis Mala s 6 allergen, a member of the cyclophilin pan-allergen family. *Biochem*, 396, 41–49.
21. Cadot PD, Proost P, Van Damme J, Engelborghs Y, Stevens EA, Ceuppens JL (2000). Purification and characterization of an 18-kd allergen of birch (Betula verrucosa) pollen: identification as a cyclophilin. *J Allergy Clin Immunol*, 105: 286–291.
22. Fujita C, Moriyama T, Ogawa T (2001) Identification of cyclophilin as an IgE-binding protein from carrots. *Int Arch Allergy Immunol*, 125: 44–50.
23. Pastorello EA, Conti A, Pravettoni V, Farioli L, Rivolta F, Ansaloni R, et al. (1998). Identification of actinidin as the major allergen of kiwi fruit. *J Allergy Clin Immunol*, 101:531-537.
24. Fujimoto YA, Kobayashi S, Murata S, Natori S (1999). Two subunits of the insect 26/29-kDa proteinase are probably derived from a common precursor protein. *J Biochem (Tokyo)*, 125:566-73.
25. Hughes VS, Page K (2007). German cockroach grass proteases cleave pro-matrix metalloproteinase 9. *Exp Lung Res*, 33: 135-50.
26. Mindykowski B, Jaenicke E, Tenzer S, Cirak S, Schweikardt T, Schild H, Decker H (2010). Cockroach allergens Per a 3 are oligomers. *Dev Comp Immunol*, 34:722-33. Epub 2010 Jan 31.
27. Guilloux L, Vuitton DA, Delbourg M, Lagier A, Adessi B, Marchand CR, Ville G (2007). Cross-reactivity between terrestrial snails (Helix species) and house-dust mite (Dermatophagoides pteronyssinus) II: in vitro study. *Allergy*, 53:151–158.
28. Wu CHH, Lee MF, Liao SC (1996). Sequencing analysis of cDNA clones encoding the American cockroach Cr-PI allergens: homology with insect hemolymph proteins. *J Biol Chem*, 271:17937-17943.
29. McCall C, Hunter S, Stedman K, Weber E, Hillier A, Bozic C, Rivoire B, Olivry T (2001). Characterization and cloning of a major high molecular weight house dust mite allergen (Der f 15) for dogs. *Vet Immunol Immunopathol*, 78:231-47.
30. Baatjies R, Meijster T, Lopata A, Sander I, Raulf-Heimsoth M, Heederik D, Jeebhay M (2010). Exposure to flour dust in South African supermarket bakeries: modeling of baseline measurements of an intervention study. *Ann Occup Hyg*, 54:309-18. Epub 2010 Mar 3.