SCIENTIFIC REPORTS

OPEN

Glutathione *S*-transferase (GST) of American Cockroach, *Periplaneta americana*: Classes, Isoforms, and Allergenicity

Received: 4 May 2017
Accepted: 18 December 2017
Published online: 11 January 2018

Nitat Sookrung[1,2], Onrapak Reamtong[3], Rojana Poolphol[4,5], Nitaya Indrawattana[6], Watee Seesuay[2], Nawannaporn Saelim[2,5], Pongsakorn Tantilipikorn[7], Chaweewan Bunnag[7], Wanpen Chaicumpa[2,5] & Anchalee Tungtrongchitr[2,5]

Insect glutathione *S*-transferases (GSTs) play important roles in insecticide/drug resistance and stress response. Medically, GSTs of house dust mites (*Dermatophagoides pteronyssinus* and *Blomia tropicalis*) and German cockroach (*Blattella germanica*) are human allergens. In this study, classes, isoforms and B-cell and allergenic epitopes of GST of American cockroach, *Periplaneta americana*, the predominant species in the tropics and subtropics were investigated for the first time. Enzymatically active native and recombinant *P. americana*-GSTs bound to IgE in sera of all *P. americana* allergic patients that were tested. By gel-based proteomics and multiple sequence alignments, the native GST comprises three isoforms of delta and sigma classes. All isoforms interacted with serum IgE of the cockroach allergic subjects. Molecularly, the protein contains six B-cell epitopes; two epitopes located at β1-α1 and β4-α3 regions bound to patients' serum IgE, indicating that they are allergenic. *P. americana* are ubiquitous and their GST can sensitize humans to allergic diseases; thus, the protein should be included in the allergen array for component resolved diagnosis (CRD) of allergic patients, either by skin prick test or specific IgE determination. The GST is suitable also as a target of environmental allergen detection and quantification for intervention of cockroach sensitization and allergic morbidity.

Cockroaches (CR) are ubiquitous. They are pestiferous source of human pathogens and allergens[1]. Although worldwide prevalence of the CR allergy is relatively less than the house dust mite (HDM) allergy[2,3] clinical manifestations caused by CR are usually more prolonged and severe and often require emergency-room visit, hospitalization and/or intensive care[4]. The CR allergens are prevailed in the environment especially in the infested areas[5,6]. In Thailand, the principal CR species causing human allergic sensitization and morbidity is *Periplaneta americana* (American CR)[7]. Currently, officially recognized *P. americana* allergens include Per a 1 (a protein that exists in multiple variants containing different numbers of repeated amino acid domains) which elicit 93% skin reactivity among CR allergic patients and bound to IgE in all patients' sera[8–10]; Per a 2 or aspartic protease-like protein (42 kDa) bound to IgE in sera of 63% of allergic rhinitis or asthma and rhinitis patients[11]; Per a 3 (Cr-PI), an insect hemolymph or insect storage protein related to arylphorin which caused skin reaction in 93% of CR allergic patients[12]; Per a 6 or troponin C, a 17 kDa protein that bound to IgE in sera of 54% of atopic patients[13,14]; Per a 7 or tropomyosin (a 37 kDa protein with high homology to other invertebrate and vertebrate tropomyosins) which bound to IgE in sera of 41% of atopic and 57% of CR allergic patients[15–17]; Per a 9 or arginine kinase which is a 43 kDa pan-insect protein that reacted to IgE of all CR allergic patients tested[18]; native troponin T (47 kDa)

[1]Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand. [2]Center of Research Excellence on Therapeutic Proteins and Antibody Engineering, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand. [3]Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand. [4]Graduate Program in Clinical Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand. [5]Department of Parasitology, Mahidol University, Bangkok, 10700, Thailand. [6]Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand. [7]Department of Oto-rhino-laryngology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand. Correspondence and requests for materials should be addressed to A.T. (email: anchalee.tun@mahidol.ac.th)

Subject	Sex	Age (years)	Diagnosis	Specific serum IgE to American CR extract (KAU/L)*	Skin prick test was positive to extract(s) of (mean wheal diameter in mm)**
Patient					
1	M	56	AR	2.68	American CR (6.1), Dp (4), Mosquito (3.1), House fly (3.1), and Bermuda grass (3.2)
2	F	43	AR	0.72	American CR (6.5), Dp (3.5), Sedge (3.5), Cat (3.3), Sugar cane (3.3), Mosquito (3.2), House fly (3.1), and Kapok (3)
3	F	47	AR	0.8	American CR (5), Dp (4.8), Cat (3.4), Dog (3.3), and Kapok (3)
4	M	21	AR	1.25	American CR (6.1), Dp (4.8), Cat (3.4), and House fly (3)
5	M	32	AR	2.59	American CR (6.4), Dp (5.1), Bermuda grass (4.1), Mosquito (3.5), House fly (3), and Sedge (3)
6	M	34	AR	2.33	American CR (5.1), Dp (5), Cat (3.8), Mosquito (3.2), House fly (3.1), Dog (3), and Sedge (3)
7	F	20	AR	1.71	American CR (4.1), Dp (3.5), Mosquito (3.2), Cat (3.2), Sugar cane,(3.1), and Sedge (3.1)
8	M	28	AR	<0.35	American CR (5.5), Dp (4), Dog (3.9), Cat (3.5), Mosquito (3), House fly (3), and Kapok (3)
9	F	52	AR	0.59	American CR (5.8), Dp (5), Dog (4), Bermuda grass (3.8), Mosquito (3), and House fly (3)
10	F	33	AR	2.81	American CR (6.4), Dp (4.8), Bermuda grass (4), *Cladosporium* spp. (3.1) and Cat (3)
11	M	21	AR	1.9	American CR (7.1), Dp (5), Mosquito (3), Cat (3), House fly (3), and Kapok (3)
12	M	23	AR	2.77	American CR (5), Dp (5), and Df (5)
13	F	29	AR	2.11	American CR (5.5), German CR (5.2), Dp (4.5), and Df (4.3)
14	M	21	AR	1.48	American CR (6), German CR (5.2), Dp (4.8), Df (4.5), Cat (3), and Dog (3)
15	F	22	AR	1.91	American CR (8.1), German CR (6.4), Dp (5.1), and Df (5)
Control					
1	F	25	Normal	<0.35	Negative to all allergens tested
2	F	30	Normal	<0.35	Negative to all allergens tested
3	F	35	Normal	<0.35	Negative to all allergens tested
4	M	24	Normal	<0.35	Negative to all allergens tested
5	M	30	Normal	<0.35	Negative to all allergens tested

Table 1. Background information of the *P. americana* allergic patients and normal controls. AR, Allergic rhinitis; CR, Cockroach; C, Control; Df, *Dermatophagoides farinae*; Dp, *Dermatophagoides pteronyssinus*; F, female; KAU/L, Kilo-allergy units/liter; M, male; P, patient. *Serum IgE levels were measured by using ImmunoCAP (UniCAP 250, Instrument Pharmacia Diagnostic AB, Uppsala, Sweden). Posive level was >0.35 KAU/L. **Mean wheal diameter (MWD): (the longest wheal diameter + the perpendicular wheal diameter)/2. MWD larger than 3 mm was positive.

bound to IgE in sera of 17% of *P. amaricana* allergic patients[19]; Per a 10 or serine protease, a 28 kDa protein that elicited skin reactivity in 82% of CR allergic patients[20]; and Per a 11 (alpha-amylase; 55 kDa) and Per a 12 (chitinase; 45 kDa) derived from midgut of *P. americana* which have been found to react to sera of 83 and 63.8% of allergic patients in immunoblot analysis[21].

Glutathione *S*-transferases (GSTs) is a large family of intracellular enzymes of aerobic prokaryotes and eukaryotes that the host organisms use for protection against oxidative stress, degenerative conditions, aging, and cancers and also detoxify endogenous and xenobiotic electrophiles such as drugs, herbicides, insecticides[22,23]. GSTs are classified according to their cellular localizations into three major families, i.e., cytosolic, mitochondrial/peroxisomal, and microsomal GSTs[24]. All insect GSTs are cytosolic and further subdivided into at least six classes including delta, epsilon, omega, sigma, theta, and zeta, based on phylogenetic analysis[23]. Delta and epsilon GSTs are arthropod specific[25]. Each GST subunit is approximately 21–28 kDa and may exist as either homo- or hetero-dimeric form[26].

Most research on insect GSTs focused on their role in insecticide/drug resistance[23,27]. It was reported that increased levels of GST activity contributed to increase detoxification capacity and resistance of the insects to several insecticides[28–31]. For medical aspect, GST of many organisms have been reported as human allergens including *Dermatophagoides pteronyssinus* (Der p 8)[32,33], *Bromia tropicalis* (Blo t 8)[34], *Blattella germanica* or German CR (Bla g 5)[35], *Alternaria alternata*[36], *Ascaris lumbricoides* and *Ascaris suis* (Asc s 1 and Asc s 3)[34,37]. The *B. germanica* GSTs, i.e., sigma BgGSTS1 and delta BgGSTD1 have been reported as potent human allergens[35,38]. Data on allergenicity and several other attributions of *P. americana* GSTs are lacking. Therefore in this study, *P. americana* GST classes, isoforms, allergenicity, and B cell epitopes were investigated.

Materials and Methods

Serum samples. This study was approved by Siriraj Ethical Committee (COA no. SI268/2008), Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok. All methods were performed in accordance with the

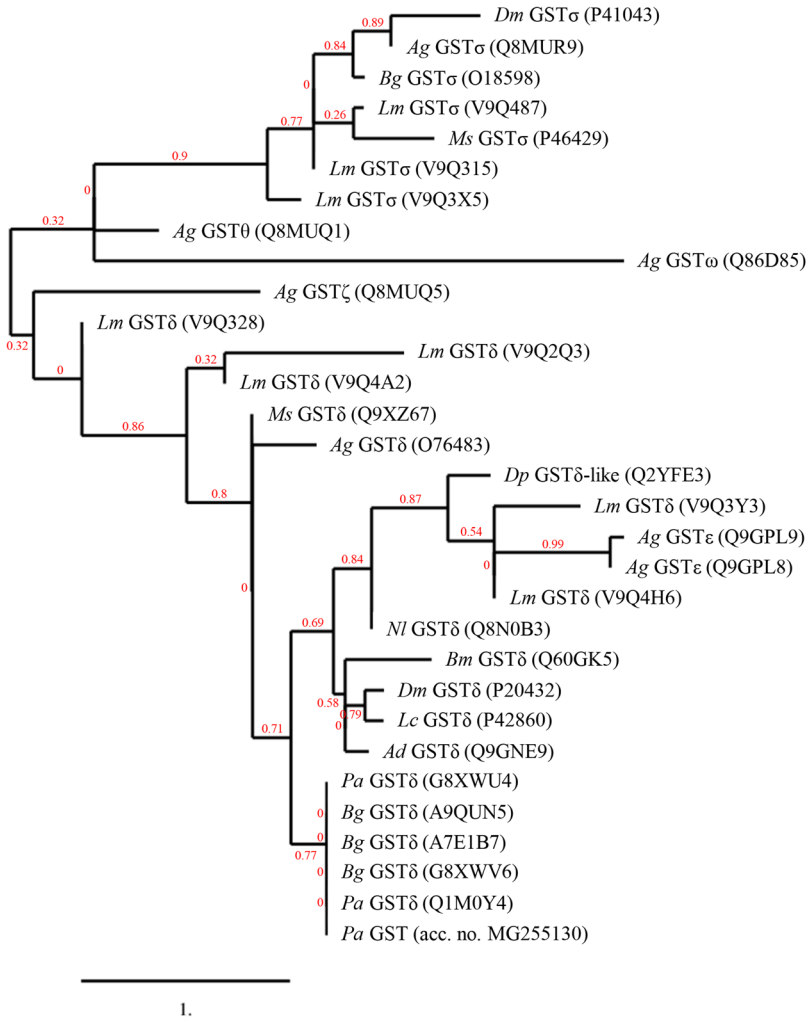

Figure 1. Phylogram of glutathione-S-transferases of *P. americana* and other insects. Ad, *Anopheles dirus*; Ag, *Anopheles gambiae*; Bg, *Blattella germanica*; Bm, *Bombyx mori*; Dm, *Drosophila melanogaster*; Dp, *Dermatophagoides pteronyssinus*; Lc, *Lucilia cuprina*; Lm, *Locusta migratoria*; Ms, *Manduca sexta*; Nl, *Nilaparvata lugens*; Pa, *Periplaneta americana*.

relevant guidelines and regulations by International Ethical Guidelines for Health-related Research Involving Humans. Informed consent was obtained from each subject. Serum samples were isolated from clotted blood aliquots collected from 15 patients who visited Allergy Clinic, Department of Oto-Rhino-Laryngology, Siriraj Hospital, Bangkok. All patients were multiply positive by skin prick test (SPT) to crude *P. americana* extract and other allergens, but positivity to the *P. americana* extract was more pronounced. Serum specific IgE levels to American CR extract were measured by using ImmunoCAP (UniCAP 250, Instrument Pharmacia Diagnostic AB, Uppsala, Sweden). Serum samples of five subjects who were negative by SPT, IgE ImmunoCAP and IgE-binding ELISA to the *P. americana* and other extracts served as non-allergic (normal) controls. A pool of sera of 10 *P. americana* allergic patients was prepared by mixing 0.5 ml of individual samples.

Preparation of recombinant GST (rGST) of *P. Americana*. Adult cockroaches were caught from houses in Bangkok. They were identified entomologically and only the *P. americana* were kept frozen at −80 °C until use. Frozen CR was ground to fine pieces in liquid nitrogen and total RNA was isolated from the powder (100 mg) by using TRIzol reagent (Invitrogen, CA, USA). After checking RNA integrity by agarose gel electrophoresis, cDNA was synthesized from the RNA and used as a PCR template for amplification of full-length GST coding sequence (*gst*). The PCR primers were designed from GenBank database (accession number AY792949; UniProt ID Q1M0Y4); forward: 5′-CCG GAT ATC ATG ACC ATC GAC TTC TAC-3′; reverse: 5′-CGA AAG CTT TCA CTT CTT GGC GAG GTT-3′. The PCR reaction mixture was: 1 µl cDNA, 10 µM each of the forward and reverse primers, 2.5 µl of 10× buffer, 3 µl of 25 mM $MgCl_2$, 2 µl of 2.5 mM dNTP, 0.2 µl of 5 units/µl DNA polymerase (Fermentas, Lithuania), and 40.3 µl ultrapure distilled water (UDW). The thermal cycles were: initial denaturation at 94 °C, 5 minutes; 30 cycles of 94 °C for 30 seconds, 50 °C for 30 seconds, and 72 °C for 40 seconds; and final extension at 72 °C for 7 minutes. The PCR product was verified by DNA sequencing before cloning into pKRX-T (Gentaur, Belgium), subcloned into pET20b$^+$(Novagen, Merck, Germany) expression vector, and

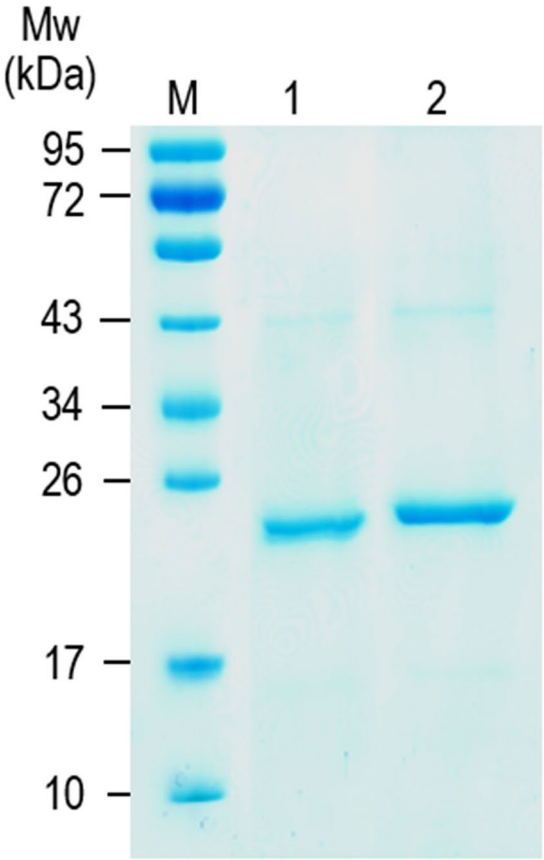

Figure 2. SDS-PAGE-separated patterns of *P. americana* native and recombinant GSTs after staining with Coomassie Brilliant Blue G-250 dye. Lane 1, nGST (~23 kDa) eluted from glutathione affinity column; lane 2, rGST (~25 kDa) purified from *gst*-pET23b^{+}-transformed *E. coli* lysate M, protein molecular weight (MW) marker. Numbers at the left are relative molecular masses of proteins.

Proteins	Orthologous protein	Accession no.	Number of matched peptides	Protein score	Matched peptide sequence (average peptide score; APS)
nGST	Glutathione *S*-transferase class delta variant 1 (*Periplaneta americana*)	Gi 359326557	5	249	LYFDIGTLYHR (102)
					FGEYYYPIYFAK (52)
					TIDFYYLPGSAPCR (34)
					VTNLMAGEHLTPEFLK (29)
					AILSYLADQYGKDDSLYPK (34)
rGST	Glutathione *S*-transferase class delta variant 1 (*Periplaneta americana*)	Gi 359326557	7	356	AIGVDLNLK (50)
					FKEMCDNLAK (45)
					AILSYLADQYGK (52)
					FGEYYYPIYFAK (64)
					VTNLMAGEHLTPEFLK (48)
					AILSYLADQYGKDDSLYPK (62)
					MNPQHTIPTLNDNGFCLWESR (35)

Table 2. LC-MS/MS Mascot results of peptides generated from in-gel tryptic digestion of *P. americana* nGST and rGST after searching against the Swiss-Prot database.

put into BL21 (DE3) *E. coli*. A transformed *E. coli* colony was grown in isopropyl β-D-1-thiogalactopyranoside (IPTG) (Affimetrix, OH, USA) conditioned-Luria-Bertani (LB) broth (Himedia, India). The recombinant protein was purified from the bacterial lysate by using HISTrap FF affinity chromatography (GE Healthcare Lifesciences) and verified by SDS-PAGE and Coomassie Brilliant Blue G-250 (CBB) staining and LC-MS/MS. Nucleotide and deduced amino acid sequences of the *P. americana* rGST was subjected to phylogenetic analysis together with GSTs of other closely related insects to determine percent identity and GST class.

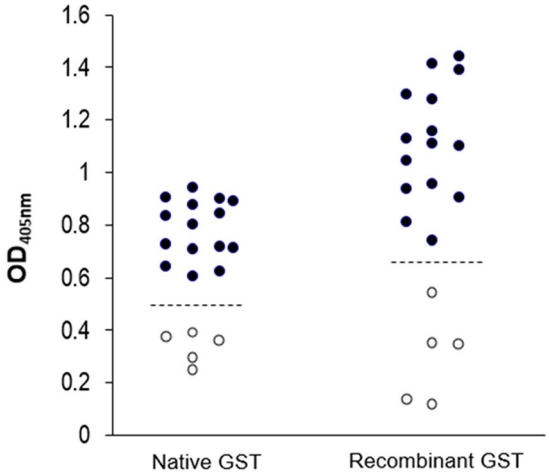

Figure 3. Scattered plots of indirect ELISA OD_{405nm} for determining IgE binding activities (allergenicity) of the *P. americana* native and recombinant GST proteins. Both proteins gave significant ELISA signals to all sera of the 15 ACR allergic patients (●) above the 5 normal controls (○). Dotted lines are cut-off OD_{405nm} between the positive and the negative IgE-ELISA which was arbitrarily set at mean + 2 SD of the normal controls.

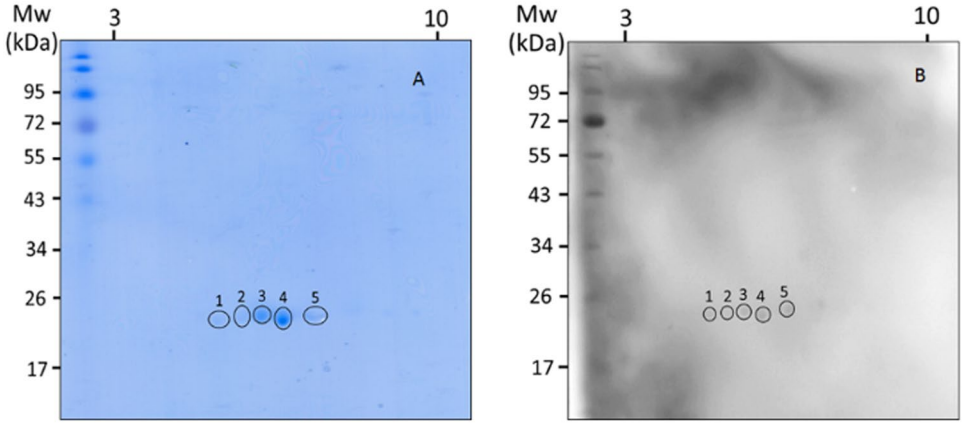

Figure 4. Isoforms of *P. americana* nGST and their IgE reactivity. (**A**) Native GST of *P. americana* was separated by 2DE at non-linear pH 3–10 and stained by Coomassie Brilliant Blue G-250 dye. Circles indicate protein spots that were subjected to LC-MS/MS for identification. The results revealed that spots 1, 2, and 4 are delta GST while spots 3 and 5 are two different sigma GST variants (see also Table 3). Thus, the *P. americana* nGST comprises of three isoforms. (**B**) All GST isoforms reacted to IgE in a serum pool of *P. americana* allergic subjects, indicating that they are human allergens.

Preparation of native *P. americana* GST (nGST). Five ml of binding buffer (phosphate buffered saline, pH 7.3) were added to dissolve the CR powder (15 mg). The preparation was sonicated (Sartorius LABSONIC® P sonicator, Germany) in ice-bath at 20 kHz, 2 minutes pulse-on, 3 minutes pulse-off for a total of 15 minutes and then centrifuged at 10,000 *g* at 4 °C for 15 minutes. The clear supernatant was collected, filtered through a sterile 0.45 µm filter, and protein content was quantified by Bradford's method (Bio-Rad, CA, USA) using bovine serum albumin (BSA) standard curve for calibration. The preparation was loaded onto a GSTrap FF affinity column (GE) that had been equilibrated with the binding buffer. The column was washed thoroughly with the binding buffer to eliminate unbound materials. The column-bound protein was eluted in 1 ml-fractions with a total of 10 ml elution buffer [50 mM Tris-HCl, 20 mM reduced glutathione (Affimetrix), pH 8.0, 5 mM DTT] at a flow rate of 1–2 ml/min. Fractions containing the eluted protein were verified by SDS-PAGE and CBB staining, pooled, dialyzed against distilled water, and lyophilized. The protein was verified by LC-MS/MS.

SDS-PAGE and Western blot analysis (WB). SDS-PAGE and WB were performed as described previously[17]. A 4% stacking and 12.5% separating polyacrylamide gels were used in the SDS-PAGE which was performed in a Mini-PROTEAN® 3 Cell (Bio-Rad). Separated proteins in the gels were either stained or electro-blotted onto a nitrocellulose membrane (NC) for WB. For WB, empty sites on the blotted NC were blocked with 3% BSA in PBS before placing the membrane in a solution of mouse anti-6× His tag (Abcam, UK).

Gel plug no.	Access. No.	Protein	Protein score	Mass (Da)	p*I*	No. of peptide	m/z	Ion score	Sequence	Possible phosphorylation sites	Possible O-glycosylation
1	Gi 359326557	GST class delta (*P. americana*)	368	24606	6.4	6	424.5294	31	FKEMCDNLAK	11, 42, 56, 72, 78, 83, 85, 100, 105, 114, 145, 166, 176, and 195	116
							466.5781	87	LYFDIGTLYHR		
							780.8726	62	FGEYYYPIYFAK		
							830.3934	65	TIDFYYLPGSAPCR		
							720.6989	44	AILSYLDQYGKDDSLYPK		
							849.3951	79	MNPQHTIPLNDNGFCLWESR		
2	Gi 359326557	GST class delta (*P. americana*)	217	24606	6.4	4	780.8727	48	FGEYYYPIYFAK	11, 42, 56, 72, 78, 83, 85, 100, 105, 114, 145, 166, 176, and 195	116
							830.3931	93	TIDFYYLPGSAPCR		
							720.697	55	AILSYLDQYGKDDSLYPK		
							849.3917	20	MNPQHTIPLNDNGFCLWESR		
3	Gi 565341529	GST class sigma (*Locusta migratoria)*	96	23459	6.19	2	434.2514	32	LTYFPVK	5, 8, 24, 42, 46, 65, 97, 105, 110, 114, 151, 163, 176, and 202	N/A
							579.8286	64	YKLTYFPVK		
4	Gi 359326557	GST class delta (*P. americana*)	100	24606	6.4	3	780.8705	42	FGEYYYPIYFAK	11, 42, 56, 72, 78, 83, 85, 100, 105, 114, 145, 166, 176, and 195	116
							830.3964	37	TIDFYYLPGSAPCR		
							720.7007	22	AILSYLDQYGKDDSLYPK		
5	Gi 359326585	GST class sigma (*B. germanica)*	80	23377	6.84	2	434.2532	31	LTYFPVK	5, 8, 24, 46, 52, 62, 65, 69, 71, 79, 95, 97, 126, 131, 144, and 151	N/A
							486.2729	48	TPVLEIDGK		

Table 3. Orthologous proteins of database that contained peptides matched with peptides of proteins in gel plugs 1–5 of 2DE-separated *P. americana*-native GST.

P. americana GST peptide no.	Amino acid sequence and residue numbers	Location on the *P. americana*-GST molecule
1	1MTIDFYYLPGSAPCRSVLLA20	β1-α1
2	47MNPQHTIPTLNDNGFCLWESRA68	α2-β3-β4
3	61GFCLWESRAILSYLADQYGK80	β 4-α3
4	72SYLADQYGKDDSLYPKDAKKRALVD96	α3-α4
5	116PIYFAKQAADPEKMKKLEEAFE137	α5-α6
6	187KCKKIVPGYEELNHSGCLKF206	α8-α9

Table 4. Sequences of *P. americana* synthetic peptides 1–6 which encompassed the 6 predicted *P. americana*-GST B-cell epitopes. These peptides were used for determining B-cell and allergenic (IgE-binding) epitopes.

Anti-mouse immunoglobulin-alkaline phosphatase (AP) conjugate (Dako Cytomation, Denmark) and BCIP/NBT substrate (KPL, MD, USA) were used to reveal the 6× His-tagged-rGST band.

Protein identification by LC-MS/MS. Native and recombinant GSTs were verified by LC-MS/MS as described previously[39]. The generated ion spectra of the peptides from tryptic-digested GSTs were interpreted by using the Turbo SEQUEST algorithm in the BioWorks™ 3.1SR1 software package (Thermo Fisher Scientific) and the nr.fasta database. The protein search parameter was performed as described previously[39], which included mass tolerance of 1.25 amu, a fragment mass tolerance of ±0.4 amu, methionine (M) oxidation, and threonine (T) or serine (S) phosphorylation. The identified peptides were further evaluated using charge state *versus* cross-correlation numbers (X_{corr}). The criteria for a positive identification of the peptides were $X_{corr} > 1.5$ for singly charged ions, $X_{cor} > 2.0$ for doubly charged ions, and Xcorr > 2.5 for triply charged ions. A delta correlation (ΔCn) of >0.08 was used as a cut-off for peptide acceptance.

Determination of enzymatic activity of GST preparations. Enzymatic activities of the nGST and the rGST were determined by using glutathione *S*-transferase assay kit (Sigma-Aldrich., MO, USA) which the 1-chloro-2,4-dinitrobenzene (CDNB) was used as the enzymatic substrate. GST standard (0.25 mg/ml) was provided with the test kit. Active GST catalyzes the conjugation of L-glutathione to CDNB *via* a thiol group of the glutathione. The reaction product, GS-DNB conjugate, could be detected at absorbance 340 nm (A_{340nm}). The amount of the product is directly proportional to the GST activity in the sample. In a reaction mixture, 2 µl (0.5 µg) of native, recombinant, and standard GST were mixed individually with 1 ml of substrate solution containing 980 µl Dulbecco's PBS, 10 µl of 200 mM reduced L-glutathione, and 10 µl of 100 mM CDNB in a 1-ml cuvette. Absorbance 340 nm of the reaction mixture was determined by spectrometer against blank (1 ml non-enzymatic conjugation substrate solution alone) using BioMate™3 series spectrophotometer (Thermo Fisher Scientific). A kinetic program was set for every 30 seconds over a period of 5 minutes after a lag time of 1 minute. Specific catalytic activities of both nGST and rGST (µmol of GST/ml/minute) were calculated.

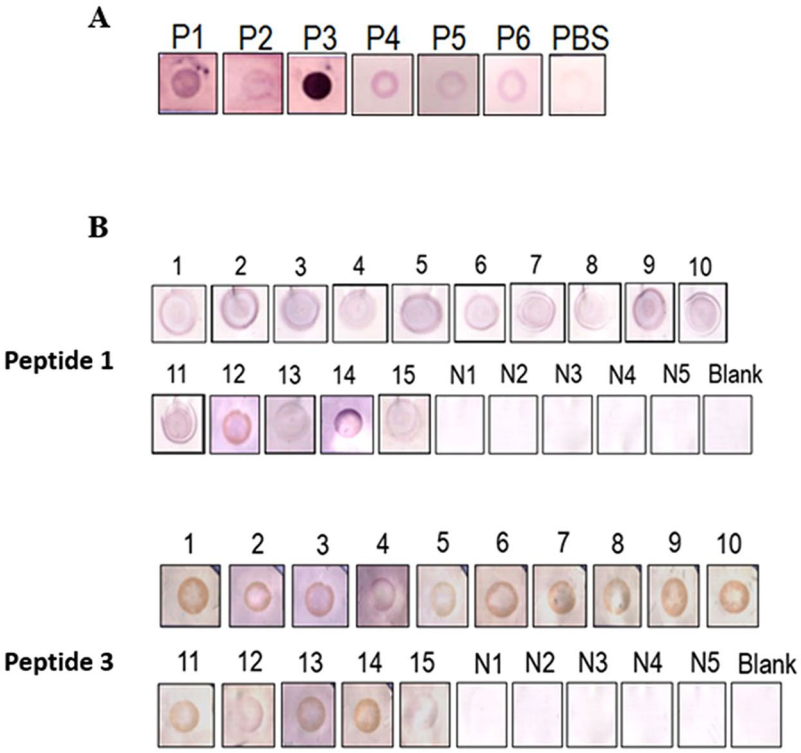

Figure 5. Results of dot-ELISAs for determining binding activities of the synthetic peptides that contained predicted B-cell epitopes of the *P. americana*-GST. (**A**) All peptides were bound by antibodies in serum pool of *P. americana* sensitized subjects, indicating that the six peptides contain B cell epitopes. PBS was used as a negative antigen control. (**B**) Peptides 1 and 3 were bound by IgE in sera of all 15*P. americana* allergic Thai patients (1–15), indicating that these two peptides contained allergenic epitopes. Sera of non-allergic controls (N1–N5) were used as negative IgE control in (**B**). Peptides 2, 4, 5, and 6 were not bound by the serum IgE (data not shown). Blank, peptide probed with PBS.

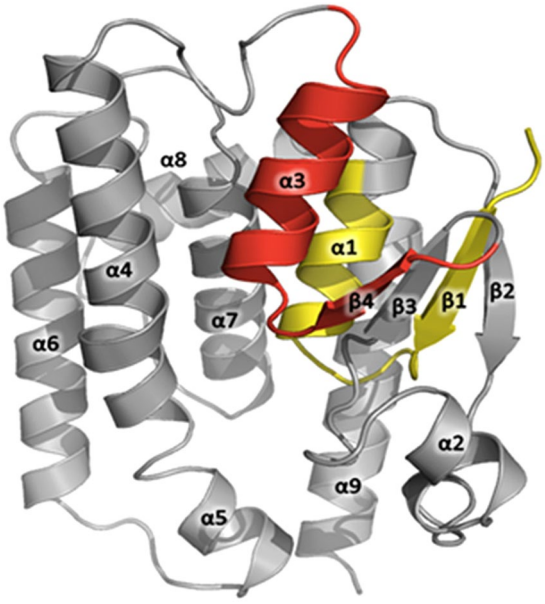

Figure 6. Homology model of the *P. americana*-GST three dimensional structure (grey) and locations of peptides 1 (yellow) and 3 (red) that were bound by IgE in sera of the 15 *P. americana* allergic patients, implying that these regions contain the GST allergenic epitopes. Peptide 1 is located on β1-α1 while peptide 3 is on β4-α3.

Allergenicities of nGST and rGST. Allergenicities (IgE binding frequencies) of the nGST and the rGST were determined by IgE-ELISA. The assay was performed as described previously[17]. The nGST and rGST (5 μg/ml carbonate-bicarbonate buffer, pH 9.6) were added to separate wells (100 μl/well) of a microtiter plate (Costar, MA, USA) and kept at 37 °C until dried. All GST-coated wells were washed and blocked with 200 μl of a blocking solution (1% BSA in PBS) before adding with serial two-fold dilutions of individual sera and the plate was incubated for 3 hours. Wells added with only the serum diluent served as blank. All wells were washed and added with 100 μl of mouse anti-human IgE-biotin conjugate (Southern Biotech, AL, USA; diluted 1:1,000 in PBS-T). Streptavidin-horseradish peroxidase (HRP) conjugate (Dako Cytomation) and ABTS substrate solution (KPL) were used for color development. Absorbance at 405 nm (OD_{405nm}) of the content in each well was determined (ELISA reader, Multiscan*EX*, Labsystem, Helsinki, Finland) against the blank. Cut-off OD_{405nm} between positive and negative IgE-ELISA was arbitrarily set at ≥mean OD_{405nm} of non-allergic sera + 2 standard deviations (SD).

Identification of nGST isoforms and their IgE reactivity. Isoforms of nGST were determined by a gel-based proteomics. Purified nGST was subjected to 2DE as described previously[39]. For the first dimensional electrophoresis, 7 cm-IPG strips and 0.5% pH 3–10 IPG buffer (GE Healthcare) were used. The electrophoresed-IPG strips were then subjected to 12% SDS-PAGE and proteins in the gel were stained by CBB. Gel pieces containing proteins of ~21 kDa were excised from the stained gel and subjected to in-gel tryptic digestion and LC-MS/MS, respectively. Protein orthologues were identified by comparing the peptide sequences of the *P. americana*-GST generated from the mass spectrometry with the Arthropoda/Insecta database sequences.

The nGST isoforms were checked for their reactivity to IgE in the pool of CR allergic patients' sera by 2DE IgE-immunoblotting. The 2DE-separated nGST was electro-transblotted onto an NC, blocked with BSA, and the blot was allowed to react with the CR allergic patients' serum pool. After keeping at 4 °C overnight, the NC was washed with TBS-T before placing in a solution of appropriately diluted mouse anti-human IgE-biotin conjugate (Southern Biotech) and kept at 25 °C on a rotating platform for 3 hours. Spots of the nGST isoforms bound by the specific serum IgE were revealed by using streptavidin-AP conjugate (Dako Cytomation) and BCIP/NBT substrate (KPL).

Identification of *P. americana*-GST B cell and allergenic epitopes. Linear B-cell epitopes of *P. americana*-GST were predicted by using BepiPred 1.0 server. The amino acid sequence of *P. americana*-GST was submitted to the server. Three methods available at the server, i.e., BCPred[40], AAP[41], and FBCPred[42] were used for the epitope prediction. Specific threshold and the epitope length were set at 85% and 20–25 residues, respectively. All predicted peptides obtained from individual methods were aligned with the *P. americana*-GST sequence to obtain consensus sequences (predicted linear B-cell epitopes). Peptides containing the potential linear B-cell epitopes of *P. americana*-GST were synthesized. Moreover, in the case that two of the predicted epitope sequences were close to each other (peptides 2 and 4 of this study), the overlapped peptide between the two sequences was also synthesized (peptide 3 of this study). Binding of the synthetic peptides to antibodies in a pool of *P. americana* sensitized human sera (for determining B-cell epitopes) and to IgE in individual allergic patients' sera (for determining allergenic epitopes) were determined by dot-ELISAs.

Dot-ELISAs. For detecting binding of the synthetic peptides to antibodies in the pool of the CR allergic patients' sera, individual synthetic peptides (1 μg) were dotted onto one cm-NC squares. PBS was used as negative peptide control. The NC pieces were blocked with 1% BSA in PBS-T and then placed in the allergic patients' serum pool. After keeping at 25 °C for 3 hours, the membranes were washed and allowed to react with goat-anti-human IgG-AP conjugate and BCIP/NBT substrate, respectively, with appropriate incubation and washing with PBS-T between the steps. The enzymatic reaction was stopped by rinsing the membranes with distilled water. Color appeared at the peptide-dotted spots indicated that the peptide contained B-cell epitope.

For determining the IgE reactivity of the synthetic peptides, individual peptides were dotted separately onto NC squares (1 μg/dot) and let air-dried. They were blocked with 1% BSA in PBST and probed with individual serum samples (diluted 1:4 in PBST). Serum samples of normal subjects were included in the experiments. After incubating and washing, all NC squares were reacted sequentially with biotin-labeled-goat-anti-human IgE, streptavidin-AP conjugate, and BCIP/NBT substrate, respectively. The peptides that gave the typical colored spots indicated that they contained allergenic (IgE binding) epitopes.

Locations of the IgE-binding (allergenic) epitopes on the *P. americana* GST three dimensional (3D) structural model. Deduced amino acid sequence of the cloned *P. americana* GST was submitted to I-TASSER server[43]. The templates used for GST 3D modeling (selected by the server) were PDB ID: 3WYW, 3VK9, 3AY8 and 4PNF. The allergenic peptides were mapped on the 3D modeled structure of the *P. americana*-GST.

Results

CR allergic patients. Dermographic data of the 15 *P. americana* allergic patients and 5 normal subjects as well as clinical diagnosis and results of the skin prick tests and the specific IgE levels among the patients are shown in Table 1.

***P. americana* nGST and rGST.** Gene sequence coding for full-length *P. americana*-GST (648 bp) was cloned into pKRX-T cloning vector and subcloned into pET20b$^+$ protein expression vector. Supplementary Fig. 1 illustrates nucleotide and deduced amino acid sequences of the *P. americana*-GST of this study (accession number MG255130). Fig. 1 illustrates a phylogenetic tree of deduced amino acid sequence (216 residues) of the cloned *P. americana*-GST and GSTs of homologous and heterologous insects. Alignment (Clustal Omega Multiple

Sequence Alignment) and percent identity of amino acid sequence of *P. americana*-rGST of this study (accession number MG255130) with GSTs of other insects are shown in Supplementary Fig. 2. The *P. americana*-rGST was purified from the *gst*-pET20b^{+}-transformed *E. coli* (Fig. 2, lane 2) and the protein was verified by the mass spectrometry (Table 2).

The nGST prepared from the frozen *P. americana* whole body powder and purified by using GST trap column revealed a protein band at ∼23 kDa (Fig. 2, lane 1). LC-MS/MS verified that the protein band was the nGST (Table 2).

Catalytic activities of nGST and rGST. Enzymatic activities of the nGST and the rGST were 36.65 and 2.93 μmol/ml/min, respectively. The nGST had a much higher enzymatic activity than the rGST.

Serum IgE reactivities of *P. americana* nGST and rGST. The mean + 2 SD of IgE-ELISA OD_{405nm} of the non-allergic control sera tested against the nGST and rGST were 0.460 and 0.649, respectively. These values were used arbitrarily as cut-off levels between positive and negative IgE-ELISA and it was found that all of the 15 allergic patients' sera gave positive IgE-ELISA results to both nGST and rGST proteins (Fig. 3).

***P. americana* nGST isoforms.** The 2 DE-pattern of purified *P. americana* nGST stained by CBB is shown in Fig. 4A. There are 5 protein spots at ∼21 kDa. The gel pieces containing the 5 protein spots were subjected to LC-MS/MS. The results (Table 3) revealed that tryptic peptides generated from spot nos. 1, 2, and 4 matched with peptides of *P. americana* delta variant 1 (accession number Gi 359326557), which the molecular mass was 24606 Da and the pI was 6.44. This protein has possibility of 14 phosphorylation sites and a possible O-glycosylation at residue 116. Peptides of spot no. 3 matched with sigma GST of *Locusta migratoria* (migratory locust) (accession number Gi 565341529) with molecular masses of 23459 Da and pI 6.19. This protein also has 14 possible phosphorylation sites; albeit most sites are different from the delta GST of spots 1, 2, and 4. Peptides of spot no. 5 matched with peptides of *B. germanica* sigma GST (accession number Gi 359326585) which has a molecular mass of 23377 Da and pI 6.84 with possible 16 phosporylation sites. Thus, the nGST of *P. americana* of this study has three isoforms which belong to delta (one isoform) and sigma classes (two isoforms). By the 2 DE-IgE immunoblotting, all three isoforms of the nGST bound to IgE in a serum pool of the *P. americana* allergic patients (Fig. 4B).

Linear B-cell epitopes and allergenic epitopes of *P. americana*-GST. Linear B-cell epitopes of *P. americana*-GST predicted by using BepiPred 1.0 server are shown in Supplementary Fig. 3. The BCPred, AAP, and FBCPred methods predicted 2, 3, and 4 epitopic sequences, respectively. All of the predicted peptides were aligned with the *P. americana*-GST sequence. They were found to match with five regions (marked in red in the Supplementary Fig. 3) of the GST including β1-α1, α2-β3-β 4, α3-α4, α5-α6, and α8-α9. The consensus peptides of these locations were synthesized (Pep 1, 2, 4, 5, and 6; marked in blue in the Supplementary Fig. 3 and peptides 1, 2, 4, 5, and 6 in Table 4). Because the matched regions α2-β3-β4 and α3-α4 were close to each other, an overlapped peptide of the two regions was also synthesized (Pep 3 in the Supplementary Fig. 3 and peptide 3 in Table 4).

All synthetic peptides gave positive binding to antibodies in a pool of 10 *P. americana* allergic subjects (Fig. 5A), verifying the computerized results that the peptides contained B-cell epitopes. Sera of all *P. americana* allergic patients gave positive IgE-dot-ELISA to peptide 1 (1MTIDFYYLPGSAPCRSVLLA20) located between β1 and α1 and peptide 3 (61GFCLWESRAILSYLADQYGK80) located between β4 and α3 of the *P. americana*-GST molecule (Fig. 5B) and did not give positive IgE binding with peptides 2, 5, and 6 (data not shown).

Figure 6 illustrates locations of the two IgE-binding (allergenic) epitopes on the 3D modeled structure of the *P. americana*-GST.

Discussion

American cockroach, *P. americana*, is a predominant species and a major source of indoor allergen causing type 1 hypersensitivity in atopic inhabitants of subtropics (such as Taiwan) and tropics (such as Brazil, Malaysia, Singapore, and Thailand), especially in crowded and unhygienic urban cities where their infestation is enhanced[1,44]. Several *P. americana*-derived proteins have been recognized as important (major) allergens as they sensitized >50% of cockroach allergic subjects. Glutathione *S*-transferases (GSTs) are indispensable enzymes which insects use to protect themselves against oxidative damage and stress and insecticide toxicity[30]. Most interest on the insect GSTs has been focused on their role in insecticide and drug resistance while other attributions of this protein family seem to be neglected. In this study, three *P. americana*-GST isoforms belonging to delta and sigma classes were identified. The roles of the protein in causing human allergy, i.e., allergenicity (specific serum IgE-binding frequency) as well as B-cell epitopes and allergenic epitopes were investigated. To our knowledge, this is the first report on the allergenic attribution of the American cockroach, *P. americana*-GST. The protein that was cloned from the GenBank database (accession number AY792949; UniProt ID Q1M0Y4) has been designated Per a 5 (Per a 5.0101) by the IUIS Allergen Nomenclature. The other isoforms reported in this study were based on the tryptic digested peptides generated from the 2DE protein spots that matched with the orthologous proteins of the database. Because the complete amino acid sequences of these isoforms were not available, they were not submitted to the IUIS Allergen Nomenclature.

Enzymatically active native and recombinant *P. americana*-GSTs were produced. On the equal weight of both proteins (0.5 μg), the native protein was about 12.5 times more active than the recombinant counterpart. The difference may be because the nGST contains several isoforms belonging to different GST classes whereas the recombinant one is produced from only one cDNA sequence. The multiple isoforms in the nGST might confer

additive or synergistic enzymatic activity. Besides, the rGST produced from the transformed BL21 (DE3) *E. coli* contains additional 6× His tag (useful for subsequent protein detection and purification) but lacks putative post-translational modifications such as glycosylation and disulfide bridge formation which might impact on the protein folding compared to the native state and hence the less enzymatic activity. After subjecting the nGST and rGST to SDS-PAGE and protein staining, the molecular masses of the native and recombinant proteins were ∼23 and ∼25 kDa, respectively. The larger size of the latter should be from the plasmid franking regions and the 6× His tag. The rGST produced in this study (accession number MG255130) contained 216 amino acids which are similar in residue number to the cloned 24614 Da *B. germanica* GSTD1[45]. The *P. americana*-rGST has 99.5 and 98.1% sequence identity to the previously reported *P. americana*-GSTs (UniProt ID G8XWU4 and Q1M0Y4) and 32.9–82.4% identity to the delta GSTs of other organisms and less so to the other GST classes. Thus, the recombinant *P. americana*-GST (accession number MG255130) should belong to the delta class.

Both native and recombinant *P. americana* GSTs reacted with IgE in sera of all cockroach allergic patients when tested by indirect ELISA, indicating that the GST is a novel and important (major) allergen of the *P. americana*. The ELISA using rGST as antigen gave higher background signal than when the nGST was used. This could be due to the contamination of residual proteins of the *E. coli* used as the rGST expression host in the rGST preparation which reacted to the anti-*E. coli* that exists naturally in the human sera.

Data on the allergenic repertoire of a protein are useful for understanding the patients' allergenic response, cross-reacting allergenic determinants among allergens, as well as for properly designing of an engineered therapeutic allergen vaccine/diagnostic material, particularly for personalized immunotherapy and component resolved diagnosis (CRD)[46,47]. Several methods have been used for gaining information on B-cell and allergenic epitopes of an allergen. These include the use of overlapping synthetic peptides or fragments of recombinant allergens[47,48]; mimotope mapping[49]; peptide microarray immunoassay[50,51]; X-ray crystallography and nuclear magnetic resonance techniques[47,52–55], computerized prediction[56–58] and specific monoclonal antibody binding and IgE competition assay[58–60]. In this study, a combination of *in silico* methods and synthetic peptide based-immunoassays was used for predicting and determining B-cell and allergenic epitopes of the *P. americana*-GST. For the *in silico* prediction of B cell epitopes, three different methods of the BepiPred 1.0 server, i.e., BCPred, AAP, and FBCPred were used. All methods gave a conformed prediction of a peptide located at α2-β3-β4; the AAP and PBCPred methods predicted another peptide at α3-α4; the BCPred and the FBCPred methods predicted another peptide at α5-α6; and only the AAP method predicted two more peptides at β1-α1 and α8-α9. The different results so-obtained suggest that several *in silico* methods should be used for increasing the possibility of finding the potential B-cell epitopes of a particular protein. From the *in silico* prediction, consensus peptides encompassed the potential B-cell epitopes were synthesized and used in the dot-ELISAs for detecting peptides bound by antibodies (B-cell epitopes) and IgE (allergenic epitopes) in sera of *P. americana* sensitized subjects. By using the dot-ELISAs, all six synthetic peptides that contained predicted GST B-cell epitopes reacted with antibodies in a serum pool of *P. americana* exposed-subjects, implying that the six peptides were part or contained GST B-cell epitopes. Among them were two IgE-binding epitopes located at the β1-α1 and β4-α4 regions of the GST molecule.

In summary, this study provides an insight into characteristics and medically important role of the *P. americana*-GST beyond the previously established physiologic roles of the protein in the host defense against toxic substances and stress conditions. Both native and recombinant GSTs of the *P. americana* were bound by IgE in sera of cockroach allergic subjects and thus the GST is a novel and major *P. americana* allergen. Because cockroaches are ubiquitous, attention should be paid on reducing the insect derived-allergenic proteins from human environment for intervention of allergic sensitization of the naives and clinical aggravation of the sensitized subjects.

References

1. Arruda, L. K. *et al.* Cockroach allergens and asthma. *J. Allergy Clin. Immunol.* **107**, 419–428 (2001).
2. Huss, K. *et al.* House dust mite and cockroach exposure are strong risk factors for positive allergy skin test responses in the Childhood Asthma Management Program. *J. Allergy Clin. Immunol.* **107**, 48–54 (2001).
3. Bunjean, K. *et al.* Prevalence of allergic rhinitis and types of sensitized allergen in adult at Wat Intaram community, Hua Raeu, Phra Nakhon Si Ayutthaya District, Phra Nakhon Si Ayutthaya Province, Thailand. *J. Med. Assoc. Thai* **95**, S63–68 (2012).
4. Yong, T. S. & Jeong, K. Y. Household arthropod allergens in Korea. *Korean J. Parasitol.* **47**, S143–153 (2009).
5. Richman, P. G., Khan, H. A., Turkeltaub, P. C., Malveaux, F. J. & Baer, H. The important sources of German cockroach allergens as determined by RAST analyses. *J. Allergy Clin. Immunol.* **73**, 590–595 (1984).
6. Menon, P., Menon, V., Hilman, B., Stankus, R. & Lehrer, S. B. Skin test reactivity to whole body and fecal extracts of American (*Periplaneta americana*) and German (*Blatella germanica*) cockroaches in atopic asthmatics. *Ann. Allergy* **67**, 573–577 (1991).
7. Tungtrongchitr, A. *et al.* The levels of cockroach allergen in relation to cockroach species and allergic diseases in Thai patients. *Asian Pac. J. Allergy Immunol.* **22**, 115–121 (2004).
8. Schou, C., Lind, P., Fernandez-Caldas, E., Lockey, R. F. & Løwenstein, H. Identification and purification of an important cross-reactive allergen from American (*Periplaneta americana*) and German (*Blattella germanica*) cockroach. *J. Allergy Clin. Immunol.* **86**, 935–946 (1990).
9. Diraphat, P. *et al.* Recombinant American cockroach component, Per a 1, reactive to IgE of allergic Thai patients. *Asian Pac. J. Allergy Immunol.* **21**, 11–20 (2003).
10. Wu, C. H., Lee, M. F. & Wang, N. M. Expression of the American cockroach Per a 1 allergen in mammalian cells. *Allergy* **55**, 1179–1183 (2000).
11. Lee, M. F., Song, P. P., Hwang, G. Y., Lin, S. J. & Chen, Y. H. Sensitization to Per a 2 of the American cockroach correlates with more clinical severity among airway allergic patients in Taiwan. *Ann. Allergy Asthma Immunol.* **108**, 243–248 (2012).
12. Wu, C. H., Lee, M. F., Liao, S. C. & Luo, S. F. Sequencing analysis of cDNA clones encoding the American cockroach Cr-PI allergens. Homology with insect hemolymph proteins. *J. Biol. Chem.* **271**, 17937–17943 (1996).
13. Hindley, J. *et al.* Bla g 6: a troponin C allergen from *Blattella germanica* with IgE binding calcium dependence. *J. Allergy Clin. Immunol.* **117**, 1389–1395 (2006).

14. Pomes, A. & Arruda, A. K. Investigating cockroach allergens: aiming to improve diagnosis and treatment of cockroach allergic patients. *Methods* **66**, 75–85 (2014).
15. Witteman, A. M., Akkerdaas, J. H., van Leeuwen, J., van der Zee, J. S. & Aalberse, R. C. Identification of a cross-reactive allergen (presumably tropomyosin) in shrimp, mite and insects. *Int. Arch. Allergy Immunol.* **105**, 56–61 (1994).
16. Asturias, J. A. *et al*. Molecular characterization of American cockroach tropomyosin (*Periplaneta americana* allergen 7), a cross-reactive allergen. *J. Immunol.* **162**, 4342–4348 (1999).
17. Sookrung, N. *et al.* Allergenicity of native/recombinant tropomyosin, per a 7, of American cockroach (CR), *Periplaneta americana*, among CR allergic Thais. *Asian Pac. J. Allergy Immunol.* **27**, 9–17 (2009).
18. Sookrung, N. *et al*. *Periplaneta americana* arginine kinase as a major cockroach allergen among Thai patients with major cockroach allergies. *Environ. Health Perspect.* **114**, 875–880 (2006).
19. Khantisitthiporn, O. *et al*. Native troponin-T of the American cockroach (CR), *Periplaneta americana*, binds to IgE in sera of CR allergic Thais. *Asian Pac. J. Allergy Immunol.* **25**, 189–197 (2007).
20. Sudha, V. T., Arora, N., Gaur, S. N., Pasha, S. & Singh, B. P. Identification of a serine protease as a major allergen (Per a 10) of *Periplaneta americana*. *Allergy* **63**, 768–776 (2008).
21. Fang, Y. *et al*. Two new types of allergens from the cockroach. *Periplaneta americana. Allergy.* **70**, 1674–1678 (2015).
22. Hayes, J. D. & Pulford, D. J. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem Mol. Biol.* **30**, 445–600 (1995).
23. Enayati, A. A., Ranson, H. & Hemingway, J. Insect glutathione transferase. *Insect Mol. Biol.* **14**, 3–8 (2005).
24. Birben, E., Sahiner, U. M. & Sackesen, C. Erzurum. S.& Kalayci, O. Oxidative stress and antioxidant defense. W*orld Allergy*. *Organ. J.* **5**, 9–19 (2012).
25. Ding, Y., Ortelli, F., Rossiter, L. C., Hemingway, J. & Ranson, H. The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles. *BMC. Genomics* **4**, 35, https://doi.org/10.1186/1471-2164-4-35 (2003).
26. Hayes, J. D., Flanagan, J. U. & Jowsey, I. R. Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* **45**, 51–58 (2005).
27. Chen, L. *et al*. Structure of an insect delta-class glutathione S-transferase from a DDT-resistant strain of the malaria vector *Anopheles gambiae*. *Acta. Crystallogr. D. Biol. Crystallogr.* **59**, 2211–2117 (2003).
28. Fournier, D., Bride, J. M., Poirie, M., Bergé, J. B. & Plapp, F. W. Jr. Insect glutathione S-transferases. Biochemical characteristics of the major forms from houseflies susceptible and resistant to insecticides. *J. Biol. Chem.* **267**, 1840–1845 (1992).
29. Grant, D. F. & Hammock, B. D. Genetic and molecular evidence for a trans-acting regulatory locus controlling glutathione S-transferase-2 expression in *Aedes aegypti*. *Mol. Gen. Genet.* **234**, 169–176 (1992).
30. Kostaropoulos, I. *et al*. Glutathione S-transferase in the defense against pyrethroids in insects. *Insect Biochem. Mol. Biol.* **31**, 313–319 (2001).
31. Vontas, J. G., Small, G. J. & Hemingway, J. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. *Biochem. J.* **357**, 65–72 (2001).
32. O'Neill, G. M., Donovan, G. R. & Baldo, B. A. Cloning and characterization of a major mite allergen of the house dust mite *Dermatophagoides pteronyssinus*, homologous with glutathione S-transferase. *Biochim. Biophys. Acta.* **1219**, 521–528 (1994).
33. Huang, C. H. *et al*. Characterization of glutathione S-transferase from dust mite, Der p 8 and its immunoglobulin E cross-reactivity with cockroach glutathione S-transferase. *Clin. Exp. Allergy* **36**, 369–376 (2006).
34. Acevedo, N. *et al*. Proteomic and immunochemical characterization of glutathione transferase as a new allergen of the nematode *Ascaris lumbricoides*. *PLoS One* **8**, e78353, https://doi.org/10.1371/journal.pone.0078353 (2013).
35. Arruda, L. K., Vailes, L. D., Platts-Mills, T. A., Hayden, M. L. & Chapman, M. D. Induction of IgE antibody responses by glutathione S-transferase from the German cockroach (*Blattella germanica*). *J. Biol. Chem.* **272**, 20907–20912 (1997).
36. Shankar, J., Singh, B. P., Gaur, S. N. & Arora, N. Recombinant glutathione-s-transferase a major allergen from *Alternaria alternata* for clinical use in allergy patients. *Mol. Immunol.* **43**, 1927–1932 (2006).
37. Acevedo, N. *et al*. Proteomic and immunochemical characterization of glutathione transferase as a new allergen of the nematode *Ascaris lumbricoides*. *PLoS One.* **8**, e78353 (2013).
38. Arlian, L. G. Arthropod allergens and human health. *Ann. Rev. Entomo.* **47**, 395–433 (2002).
39. Choopong, J. *et al*. Proteome, allergenome, and novel allergens of house dust mite. *Dermatophagoides farinae*. *J. Proteome Res.* **15**, 422–430 (2016).
40. El-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* **21**, 243–255 (2008).
41. Chen, J., Liu, H., Yang, J. & Chou, K. C. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* **33**, 423–428 (2007).
42. El-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting flexible length linear B-cell epitopes. *Comput. Syst. Bioinformatics Conf.* **7**, 121–132 (2008).
43. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
44. Sookrung, N. & Chaicumpa, W. A revisit to cockroach allergens. *Asian Pac. J. Allergy Immunol.* **28**, 95–106 (2010).
45. Ma, B. & Chang, F. N. Purification and cloning of a delta class glutathione S-transferase displaying high peroxidase activity isolated from the German cockroach *Blattella germanica*. *FEBS. J.* **274**, 1793–1803 (2007).
46. Valenta, R. *et al*. The recombinant allergen-based concept of component-resolved diagnostics and immunotherapy (CRD and CRIT). *Clin. Exp. Allergy* **29**, 896–904 (1999).
47. Pomes, A. Relevant B cell epitopes in allergic disease. *Int. Arch. Allergy Immunol.* **152**, 1–11 (2009).
48. van Milligen, F. J., van 't Hof, W., van den Berg, M. & Aalberse, R. C. IgE epitopes on the cat (*Felis domesticus*) major allergen Fel d I: a study with overlapping synthetic peptides. *J. Allergy Clin. Immunol.* **93**, 34–43 (1994).
49. Pacios, L. F. *et al*. Mimotope mapping as a complementary strategy to define allergen IgE-epitopes: peach Pru p 3 allergen as a model. *Mol. Immunol.* **45**, 2269–2276 (2008).
50. Vereda, A. *et al*. Identification of IgE sequential epitopes of lentil (Len c 1) by means of peptide microarray immunoassay. *J. Allergy Clin. Immunol.* **126**, 596–601 (2010).
51. Otsu, K., Guo, R. & Dreski, S. C. Epitope analysis of Ara h 2 and Ara h 6: characteristic patterns of IgE-binding fingerprints among individuals with similar clinical histories. *Clin. Exp. Allergy* **45**, 472–484 (2015).
52. Amit, A. G., Mariuzza, R. A., Phillips, S. E. & Poljak, R. J. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* **233**, 747–753 (1986).
53. Sheriff, S. *et al*. Three-dimensional structure of an antibody-antigen complex. *Proc. Natl. Acad. Sci. USA* **84**, 8075–8079 (1987).
54. Padlan, E. A. *et al*. Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. USA* **86**, 5938–5942 (1989).
55. Ichikawa, S. *et al*. NMR study on the major mite allergen Der f 2: its refined tertiary structure, epitopes for monoclonal antibodies and characteristics shared by ML protein group members. *J. Biochem.* **137**, 255–263 (2005).
56. Dall' Antonia, F., Gieras, A., Devanaboyina, S. C., Valenta, R. & Keller, W. Prediction of IgE-binding epitopes by means of allergen surface comparison and correlation to cross-reactivity. *J. Allergy Clin. Immunol.* **128**, 872–879 (2011).
57. Dall' Antonia, F., Pavkov-Keller, T., Zangger, K. & Keller, W. Structure of allergens and structure based epitope predictions. *Methods* **66**, 3–21 (2014).

58. Tasaniyananda, N. *et al.* A novel IgE-binding epitope of cat major allergen, Fel d 1. *Biochem. Biophys. Res. Commun.* **470**, 593–598 (2016).
59. Olson, J. R. & Klapper, D. G. Two major human allergenic sites on ragweed pollen allergen antigen E identified by using monoclonal antibodies. *J. Immunol.* **136**, 2109–2115 (1986).
60. Chapman, M. D., Heymann, P. W. & Platts-Mills, T. A. E. Epitope mapping of two major inhalant allergens, Der p I and Der f I, from mites of the genus *Dermatophagoides*. *J. Immunol.* **139**, 1479–1484 (1987).

Acknowledgements

The work was co-supported by the National Research University (NRU) Project of the Office of Commission on Higher Education, Ministry of Education, the Thailand Research Fund (RSA5780025), Faculty of Medicine Siriraj Hospital, Mahidol University, and the NSTDA Chair Professor grant (P-1450624) of the National Science and Technology Development Agency (NSTDA) funded by the Crown Property Bureau of Thailand.

Author Contributions

A.T.T., N.S.R. and W.C.C. contributed to the design of the experiments, result analysis, and manuscript writing. N.S.R., R.P.P., N.I.D., O.R.T. and N.S.L. performed the experiments. W.S.S. predicted the GST B cell epitopes and the locations of the IgE-binding epitopes on the 3D structure of the GST computerized model. P.T.T. and C.B.N. performed skin prick test, checked serum specific IgE levels and collected blood samples. N.S.R., W.S.S. and W.C.C. created Figures.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-18759-z.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017

Glutathione S-transferase (GST) of American Cockroach, Periplaneta americana: Classes, Isoforms, and Allergenicity

Nitat Sookrung [1,2], Onrapak Reamtong [3], Rojana Poolphol [4,5], Nitaya Indrawattana [6], Watee Seesuay [2], Nawannaporn Saelim [2,5], Pongsakorn Tantilipikorn [7], Chaweewan Bunnag [7], Wanpen Chaicumpa [2,5], Anchalee Tungtrongchitr [2,5,*]

[1]Department of Research and Development, [2]Center of Research Excellence on Therapeutic Proteins and Antibody Engineering, [4] Graduate Program in Clinical Pathology, [5]Department of Parasitology, and [7] Department of Oto-rhino-laryngology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand [3] Department of Molecular Tropical Medicine and Genetics and [6]Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand

```
1    atgaccatcgacttctactacctgcccggcagcgcaccatgccggtcagctctcctggct
      M  T  I  D  F  Y  Y  L  P  G  S  A  P  C  R  S  A  L  L  A    20
61   gccaaggccatcggtgtggatctgaacctcaaagtgacgaacctcatggctggcgaacat
      A  K  A  I  G  V  D  L  N  L  K  V  T  N  L  M  A  G  E  H    40
121  ctcacgcctgagtttcttaagatgaatcctcagcatacgattccagccctaaacaacaac
      L  T  P  E  F  L  K  M  N  P  Q  H  T  I  P  A  L  N  N  N    60
181  gttagctatttgtgggaaagccgagccattctaagtaagctggctaaccggtatggcaag
      V  S  Y  L  W  E  S  R  A  I  L  S  K  L  A  N  R  Y  G  K    80
241  gacgactcgctgtaccgcagtacggcaaaaagcgagctcttgtgggacagagactgtact
      D  D  S  L  Y  R  S  T  A  K  S  E  L  L  W  D  R  D  C  T   100
301  tcgatattgttcccctgtaccattgatacggagaattcttttatgcaatttcatttagca
      S  I  L  F  P  C  T  I  D  T  E  N  S  F  M  Q  F  H  L  A   120
361  aagcaacctgtagatcatgaaaagataaagcaactagagaagcccttcgagtttttgaat
      K  Q  P  V  D  H  E  K  I  K  Q  L  E  K  P  F  E  F  L  N   140
421  aagttcctggaatcgcaacagtttgtggcagaaaataagctcaccattgcggacctggca
      K  F  L  E  S  Q  Q  F  V  A  E  N  K  L  T  I  A  D  L  A   160
481  attgtctcctctgtttccactgctgacatcataggctttgatgtaagcaaatactcaaac
      I  V  S  S  V  S  T  A  D  I  I  G  F  D  V  S  K  Y  S  N   180
541  gtcgccaaatggttcgagaaatgcaagaagattgttccaggctatgaggaactgaatcac
      V  A  K  W  F  E  K  C  K  K  I  V  P  G  Y  E  E  L  N  H   200
601  tccggatgcttgaagttcaaggagatgtgcgataacctcgccaagaagtga
      S  G  C  L  K  F  K  E  M  C  D  N  L  A  K  K  -            216
```

Supplementary Figure 1. Nucleotide and deduced amino acid sequences of GST of American cockroach, *P. americana,* of this study (accession number MG255130).

```
GST|Accession no.MG255130   ----------------------------------------------MTIDFYYLPGSA
GSTδ|V9Q2Q3                 ----------------------------------------------MPSIILYGNELSP
GSTδ|V9Q3Y3                 ----------------------------------------------MAPPTLYDNPISP
GSTδ|V9Q4H6                 ----------------------------------------------MSPIILYNHDGSP
GSTδ|V9Q328                 ----------------------------------------------MAGLKLYSVSDSP
GSTδ|V9Q4A2                 ----------------------------------------------MPKLRLYHAEASP
GSTδ|O76483                 ---------------------------------------------MTPVLYYLPPSP
GSTδ|Q9XZ67                 -------------------------------------------MPSEPIKFYYLAPSP
GSTδ|Q60GK5                 ----------------------------------------------MTIDLYYVPGSA
GSTδ|Q8N0B3                 ----------------------------------------------MPIDLYYVPGSA
GSTδ|A9QUN5                 ----------------------------------------------MTIDFYYLPGSA
GSTδ|A7E1B7                 ----------------------------------------------MTIDFYYLPGSA
GSTδ|G8XWV6                 ----------------------------------------------MTIDFYYLPGSA
GSTδ|Q1M0Y4                 ----------------------------------------------MTIDFYYLPGSA
GSTδ|G8XWU4                 ----------------------------------------------MTIDFYYLPGSA
GSTδ|Q9GNE9                 ---------------------------------------------MDFYYLPGSA
GSTδ|P20432                 ----------------------------------------------MVDFYYLPGSS
GSTδ|P42860                 ---------------------------------------------MDFYYLPGSA
GSTδ-like|Q2YFE3            ----------------------------------------------MSKPIFYYHPFSG
GSTε|Q9GPL9                 ----------------------------------------------MPKPVLYTVHLSP
GSTε|Q9GPL8                 ----------------------------------------------MSNLVLYTLHLSP
GSTσ|P46429                 ----------------------------------------------MPKVVFHYFGAKG
GSTσ|V9Q3X5                 ---------------------------------------------MAPKYKLLYFPFRV
GSTσ|P41043                 MADEAQAPPAEGAPPAEGEAPPPAEGAEGAVEG-GEAAPPAEPAEPIKHSYTLFYFNVKA
GSTσ|Q8MUR9                 -------------------------------------------------------FNVKA
GSTσ|O18598                 ----------------------------------------------MAPSYKLTYCPVKA
GSTσ|V9Q487                 ----------------------------------------------MAAKYKLTYFPIMA
GSTσ|V9Q315                 ----------------------------------------------MAPKYKLTYFPVKA
GSTω|Q86D85                 ----------------------------MSNGKHLAKGSSPPSLPDDGKLRLYSMRFCP
GSTθ|Q8MUQ1                 ---------------------------------------------MSKNLKYYYDLMSQ
GSTζ|Q8MUQ5                 -------------------------------------MANVDILPESQPILYSYWRSS


GST|Accession no.MG255130   PCRSVLLAAKAIGVDLNLKVTNLM--AGEHLTPEFLKMNPQHTIPTLND---NGFCLWES
GSTδ|V9Q2Q3                 PSRAAKMIAEKLGVDVDFKRTYPI--KGECKKPEYLKINPMHTIPTIID---GPFTLSDS
GSTδ|V9Q3Y3                 PCRLVRLVAGVIGVDLKVVNVKDA--GIDMKSPEMLKKNPQHTVPTLED---NGLCLAES
GSTδ|V9Q4H6                 PCCLVRLVAGVLGLELKKVDISDI--ENGMRNPEMLKKNPQHTIPTIED---DGLVITES
GSTδ|V9Q328                 PSTAVKMALEALALEYTNVEVDFA--AGEHLSDDFSKKNPQREIPCLDD---NGFFLSES
GSTδ|V9Q4A2                 PCRTVRMVARTLGLELQLVDVDLL--QNEHLKPQFLQMNPQHQVPLLDD---NGFYLPES
GSTδ|O76483                 PCRSVLLLAKMIGVELELKALNVM--EGEQLKPDFVELNPQHCIPTLDD---HGLVLWES
GSTδ|Q9XZ67                 PCRAVMMAARALDLELDLIPTNIM--DSDHKTPEYLKMNPQHTIPTMDD---SGFILWES
GSTδ|Q60GK5                 PCRAVLLTAKALNLNLNLKLVDLH--HGEQLKPEYLKLNPQHTVPTLVD---DGLSIWES
GSTδ|Q8N0B3                 PCRNVLLAAKAVGVDLNLKLTDLK--SGQHLTPEFIKLNPQHNVPTLDD---NGFVLNES
GSTδ|A9QUN5                 PCRSVLLAAKAFGVNLNLKVTNLM--AGEHLTPEFLKMNPQHTIPTLND---NGFCLWES
GSTδ|A7E1B7                 PCRSVLLAAKAFGVDLNLKVTNLM--AGEHLTPEFLKMNPQHTIPTLND---NGFCLWES
GSTδ|G8XWV6                 PCRSVLLAAKAFGVDLNLKVTNLM--AGEHLTPEFLKMNPQHTIPTLND---NGFCLWES
GSTδ|Q1M0Y4                 PCRSVLLAAKAIGVDLNLKVTNLM--AGEHLTPEFLKMNPQHTIPTLND---KGFCLWES
GSTδ|G8XWU4                 PCRSVLLAAKAIGVDLNLKVTNLM--AGEHLTPEFLKMNPQHTIPTLND---NGFCLWES
GSTδ|Q9GNE9                 PCRAVQMTAAAVGVELNLKLTNLM--AGEHMKPEFLKINPQHCIPTLVD---NGFALWES
GSTδ|P20432                 PCRSVIMTAKAVGVELNKKLLNLQ--AGEHLKPEFLKINPQHTIPTLVD---NGFALWES
GSTδ|P42860                 PCRSVLMTAKALGIELNKKLLNLQ--AGEHLKPEFLKINPQHTIPTLVD---GDFALWES
GSTδ-like|Q2YFE3            PCRTVSTVAKILNVDMEMKKLDLL--TKEHLNPEFLKVNPFHKVPTFVD--SDGFVVDES
GSTε|Q9GPL9                 PCRAVELTAKALGLELERKLVNLL--AGENLTPEFLKLNPKHTIPVLDD---NGTIISES
GSTε|Q9GPL8                 PCRAVELTAKALGLELEQKTINLL--TGDHLKPEFVKLNPQHTIPVLDD---NGTIITES
GSTσ|P46429                 WARPT-MLLAYGGQEFEDHRVEYE------QWPEFKPNTPFGQMPVLEI---DGKKYAQS
GSTσ|V9Q3X5                 IAEPIRFILAYVGADYEDVRIEFS------DWPPLKPDTPYGTMPVLEV---DGKKLGQS
GSTσ|P41043                 LAEPLRYLFAYGNQEYEDVRVTRD------EWPALKPTMPMGQMPVLEV---DGKRVHQS
GSTσ|Q8MUR9                 LGEPLRFLLSYGNLPFDDVRITRE------EWPALKPTMPMGQMPVLEV---DGKKVHQS
GSTσ|O18598                 LGEPIRFLLSYGEKDFEDYRFQEG------DWPNLKPSMPFGKTPVLEI---DGKQTHQS
GSTσ|V9Q487                 LGEPIRFLLSYGKIEFEDERCERE------KWPSVKESMPFGQMPVLEI---DGKKTWQS
GSTσ|V9Q315                 LGEPIRFLLSYGNMEFEDDRFERE------KWPSIKPSMPFGQVPVLEF---DGKKTHQS
GSTω|Q86D85                 YAQRVHLMLDAKKIPYHAIYINLS-----EKPEWYLEKNPLGKVPALEIPGKEGVTLYES
GSTθ|Q8MUQ1                 PSRALWIFLEKTKLPYEKCLINLG--KGEHLTEEFKAINRFQKVPCITD---SQIKLAES
GSTζ|Q8MUQ5                 CSWRVRIALNLKEIPYDIKPISLIKSGGEQHCNEYREVNPMEQVPALQI---DGHTLIES
                                                                  * :           :*

GST|Accession no.MG255130   RAILSYLADQYGKD---DSLYPKD-AKKRALVDQRLYFDIRTLYHRFGEY----YYPIYF
GSTδ|V9Q2Q3                 HAIVAYLVDRFGKN---DSLYPKD-IQKRSKVNERLCFDISL-FTKVLKF----VVGPLL
GSTδ|V9Q3Y3                 RAISMYLISKYAKD---DSLYPKD-VNKRVLVDQRLFYDQD-LYNRILSV----FLP-KF
GSTδ|V9Q4H6                 RAIAMYLVSRYAKD---DSLYPKE-LQRRVLVDQRLFFDQD-LYNRIMAV----CNP-LF
GSTδ|V9Q328                 VAILQYLADKYGPG---HSLYPRD-PQQRALVNHRLAFNISTYYARIAEY----AVAPIF
GSTδ|V9Q4A2                 RAIIAYLVDNYAKD---DSLYPRD-PKKRATVNRMLFFDAGTLWMNYHIY----QMP-IF
GSTδ|O76483                 RVILAYLVSAYGKD---ENLYPKD-FRSRAIVDQRLHFDLGTLYQRVVDY----YFPTIQ
GSTδ|Q9XZ67                 RAILAYLVNAYGKD---DSLYPKN-PRQRAIVDQRLNFDIGTLFPRYSNL----YFPMLF
GSTδ|Q60GK5                 RAIITYLVNKYAKG---SSLYPED-PKARALVDQRLYFDIGTLYQRFSDY----FYPQVF
GSTδ|Q8N0B3                 RAIMTYLADQYGKD---DSLYPKD-PKKRAKVNQRLYFDMGTLYQSFGDA----YYPHMF
```

```
GSTδ|A9QUN5                 RAILSYLADQYGKD---DSLYPKD-PKKRALVDQRLYFDLGTLYQRFGDY----YYPIMF
GSTδ|A7E1B7                 RAILSYLADQYGKD---DSLYPKD-PKKRALVDQRLYFDLGTLYQRFGDY----YYPIMF
GSTδ|G8XWV6                 RAILSYLADQYGKD---DSLYPKD-PKKRALVDQRLYFDLGTLYQRFGDY----YYPIMF
GSTδ|Q1M0Y4                 RAILSYLADQYGKD---DSLYRRD-AKKRALVDQRLYFDIGTLYHRFGEY----YYPIYF
GSTδ|G8XWU4                 RAILSYLADQYGKD---DSLYPKD-AKKRALVDQRLYFDIGTLYHRFGEY----YYPIYF
GSTδ|Q9GNE9                 RAICTYLAEKYGKD---DKLYPKD-PQKRAVVNQRLYFDMGTLYQRFADY----YYPQIF
GSTδ|P20432                 RAIQVYLVEKYGKT---DSLYPKC-PKKRAVINQRLYFDMGTLYQSFANY----YYPQVF
GSTδ|P42860                 RAIMVYLVEKYGKN---DSLFPKC-PKKRAVINQRLYFDMGTLYKSFADY----YYPQIF
GSTδ-like|Q2YFE3            RVIAMYLVESRKPD---SFLYPKNDLKKRIQIDRWLHYDINLS-TTISAP----MF-CVF
GSTε|Q9GPL9                 HAIMIYLVRKYGQGEGKDALYPTD-IVEQARVNEALHFESGVLFARLRFI----TELVFF
GSTε|Q9GPL8                 HAIMIYLVTKYGKD---DSLYPKD-PVKQARVNSALHFESGVLFARMRFN----FERILF
GSTσ|P46429                 LAISRYLGRKYGLA-------GND-IEEDFEIDQIVDFVND-IRASAASV----EYEQDA
GSTσ|V9Q3X5                 IPICRYLAKQYGLL-------AEN-DWDNAQIDAAVDAIND-LRMSIKNF----YFEENG
GSTσ|P41043                 ISMARFLAKTVGLC-------GAT-PWEDLQIDIVVDTIND-FRLKIAVV----SYEPED
GSTσ|Q8MUR9                 VAMSRYLANQVGLA-------GAD-DWENLMIDTVVDTVND-FRLKIAVV----SYEPDD
GSTσ|O18598                 VAISRYLGKQFGLS-------GKD-DWENLEIDMIVDTISD-FRAAIANY----HYDADE
GSTσ|V9Q487                 LAICRYLGKQMGLA-------GAN-DWEDLQIDMAIETVTD-LRLKIANY----WYETDE
GSTσ|V9Q315                 IAICRYLGKQLKLA-------GDN-DWEALQIDMAADCLTD-LRMKLGSF----FYESDE
GSTω|Q86D85                 LVLSDYIEEAYSAQ--QRKLYPAD-PFSKAQDRILIERFAGS----VIGP----YYRILF
GSTθ|Q8MUQ1                 VAIFRYLCREYQVP---DHWYPAD-SRRQALVDEYLEWQHHNTRATCAIYFQYVWLRPRM
GSTζ|Q8MUQ5                 VSIMYYLEETRPQ----RPLMPQD-VLKRAKVREICEVIASGV-QPLQNL----IV-LIH
                              :  ::

GST|Accession no.MG255130   AKQAADPEKM----KKLEEAF----EFLNKFLE---SQEFVAGNKLTIADLAIVSSVSTA
GSTδ|V9Q2Q3                 RTHEPTEELR----NDCIDGL----ETVERFLS---ASKFIAGDDLTVADYCYYCTITFV
GSTδ|V9Q3Y3                 FGKQTDPSSI----EKVNEGL----ETLNRMLD---GKQWLAGDNITLADYAIANTLSAL
GSTδ|V9Q4H6                 YGKTVAESDA----DRMKDGL----DTLSRMLD---GKQWLAGDHVTLADYAVAVSLATL
GSTδ|V9Q328                 FDYKRTPEGL----NKLKIAL----NVLNTILERQ-GTKFAAGEHMTLADLSLVAATMCL
GSTδ|V9Q4A2                 YGQPGDPEKQ----KKMLEAF----ALFDKLLE---GKEWAAGSHVTIADYALAVTATSA
GSTδ|O76483                 LGAHLDQTKK----AKLAEAL----GWFEAMLK---QYQWSAANHFTIADIALCVTVSQI
GSTδ|Q9XZ67                 RGDEYNQENA----DKLNEAL----GWLNIFLE---KSAFVAGDNLTIADISIIVTITNL
GSTδ|Q60GK5                 AGAPADKAKN----EKVQEAL----QLLDKFLE---GQKYVAGPNLTVADLSLIASVSSL
GSTδ|Q8N0B3                 GGAPLDEDKK----KKLGDAL----VFLDGFLE---KSAFVAGEDLTLADLAIVASISTI
GSTδ|A9QUN5                 AKASPDAEKM----KKLEEAY----QFLDKFLE---GQKFVAGNSLTIADIATIASVSTA
GSTδ|A7E1B7                 AKASPDAEKM----KKLEEAY----QFLDKFLE---GQKFVAGNSLTIADIATIASVSTA
GSTδ|G8XWV6                 AKASPDAEKM----KKLEEAY----QFLDKFLE---GQKFVAGNSLTIADIATIASVSTA
GSTδ|Q1M0Y4                 AKQAADPEKM----KKLEEAF----EFLNKFLE---SQEFVAGNKLTIADLAIVSSVSTA
GSTδ|G8XWU4                 AKQAADPEKM----KKLEEAF----EFLNKFLE---SQEFVAGNKLTIADLAIVSSVSTA
GSTδ|Q9GNE9                 AKQPANAENE----KKMKDAV----DFLNTFLD---GHKYVAGDSLTIADLTVLATVSTY
GSTδ|P20432                 AKAPADPEAF----KKIEAAF----EFLNTFLE---GQDYAAGDSLTVADIALVATVSTF
GSTδ|P42860                 AKAPADPELY----KKMEAAF----DFLNTFLE---GHQYVAGDSLTVADLALLASVSTF
GSTδ-like|Q2YFE3            RGHQVQDYQV----EQGKETL----KTLDGVMQSFEGKFLTGADQFTLADIAMYFSLNTM
GSTε|Q9GPL9                 ARKPEIPEDR---IEYVRTAY----RLLEDSL----QSDYVAGSRMTIADLSCISSVASM
GSTε|Q9GPL8                 FGKSDIPEDR---VEYVQKSY----ELLEDTL----VDDFVAGPTMTIADFSCISTISSI
GSTσ|P46429                 ANKEVK------HEENMKNKYPFQLNKLSEIIT--KNNGFLALGRLTWADFVFVGMFDYL
GSTσ|V9Q3X5                 STKNNM------KEKLMKEAVPFYYKKLEEMVK--NNGGYLAGGKLTWGDLYLTTMLELF
GSTσ|P41043                 EIKEKK------LVTLNAEVIPFYLEKLEQTVK--DNDGHLALGKLTWADVYFAGITDYM
GSTσ|Q8MUR9                 EIKEKK------LVTLNNEVIPFYLEKLDDIAR--DNNGYLANSKLSWADIYFTAILDYL
GSTσ|O18598                 NSKQKK------WDPLKKETIPYYTKKFDEVVK--ANGGYLAAGKLTWADFYFVAILDYL
GSTσ|V9Q487                 AQKEKK------KEPLFNETLPFLLPRLDNLVK--ENGGYLANGKLSWADLYFIAVLDYL
GSTσ|V9Q315                 AVKEKK------KEPIIKEFLPFFLPRLDNLVK--ENGGYLANGKLSWADFYFAGILDYV
GSTω|Q86D85                 AADGIPPGAI----TEFGAGL----DIFEKELKA-RGTPYFGGDKPGMIDYMIWPWCERV
GSTθ|Q8MUQ1                 FGTKVDPKQAEKYRGQMEGTL----DFIEREYLG-SGARFIAGDEITVADLLAACEIEQP
GSTζ|Q8MUQ5                 VGEEKKKEWA---QHWITRGF----RAIEKLLST-SAGKFCVGDEITLADCCLVPQVFNA
                                                          ..                    *

GST|Accession no.MG255130   --------DIMGF-DV-SKYSNVAKWFEKCKKIVPGYE--ELNHSGCLKFKEMCDNLAKK
GSTδ|V9Q2Q3                 -----DIPQKGVI-DL-KKYKNIQRWMDLIHKTCPLFT--KYDQIAEKAFQIYLEDGPY-
GSTδ|V9Q3Y3                 -----EFAPEFGI-DP-TKQPNIKQWLPRVESSSPKYG--ESL----KEFREALKKMAQK
GSTδ|V9Q4H6                 -----EYNPEFAI-DA-AKHPNITQWISRLEESSNTYR--EHI----ELFFETVRKATEE
GSTδ|V9Q328                 --------EAVEF-DL-TPWARVQRWYADFKQAAPRLW--AVAEPAMLELRAFSNSPPDL
GSTδ|V9Q4A2                 --------QWSGV-DL-KVYDNVTRWLERTKEGITCFD--EINNSSFAALKELLQKLKEK
GSTδ|O76483                 --------EAFQF-DL-HPYPRVRAWLQKCKDELQGHGYKEINETGAETLAGLFRSKLKQ
GSTδ|Q9XZ67                 --------DAFKF-DF-SEHENLTKWFERTKKALEPYDWEDIDETGAQMLADFLKREH--
GSTδ|Q60GK5                 --------EASDI-DF-KKYANVKRWYETVKSTAPGYQ--EANEKGLEAFKGLVNSMLKK
GSTδ|Q8N0B3                 --------EAVEY-DL-SPYKNINSWYSKVKAAAPGYK--EANEEGAKGFGQMFKAMTGK
GSTδ|A9QUN5                 --------AILGF-DI-TRYPNVNKWFENAKKVIPGYD--ELNHSGCLEFRKMWDNLTQK
GSTδ|A7E1B7                 --------AILGF-DI-TRYPNVNKWFENAKKVIPGYD--ELNHSGCLEFKKMWDNLTQK
GSTδ|G8XWV6                 --------AILGF-DI-TRYPNVNKWFENAKKVIPGYD--ELNHSGCLEFKKMWDKLTQK
GSTδ|Q1M0Y4                 --------DIMGF-DV-SKYSNVAKWFEKCKKIVPGYE--ELNHSGCLKFKEMCDNLAKK
GSTδ|G8XWU4                 --------DIMGF-DV-SKYSNVAKWFEKCKKIVPGYE--ELNHSGCLKFKEMCDNLAKK
GSTδ|Q9GNE9                 --------DVAGF-EL-AKYPHVAAWYERTRKEAPGAA--I-NEAGIEEFRKYFEK----
GSTδ|P20432                 --------EVAKF-EI-SKYANVNRWYENAKKVTPGWE--E-NWAGCLEFKKYFE-----
GSTδ|P42860                 --------EVAGF-DF-SKYANVAKWYANAKTVAPGFD--E-NWEGCLEFKKFFN-----
GSTδ-like|Q2YFE3            -----EVYP-KYF-KF-DDYPNLKSWYHRVAEALKQYDTEGTIPKAIETMKQFIQQRAAE
GSTε|Q9GPL9                 -----V--GFIPM-ER-SEFPRVHGWIERMKQ-LPYYE--EINGAGATELAEFIVDMLAK
GSTε|Q9GPL8                 -----M--GVVPL-EQ-SKHPRIYAWIDRLKQ-LPYYE--EANGGGGTDLGKFVLAKKEE
```

```
GSTσ|P46429                 -------KKMLRMPDLEEQYPIFKKPIETVLSNPKLKAYLDSAPKKEF------------
GSTσ|V9Q3X5                 -------DHDLEC-EITKDYPHLATLRNKVVNIPSIKTWIEKRPKSER------------
GSTσ|P41043                 -------NYMVKR-DLLEPYPALRGVVDAVNALEPIKAWIEKRPVTEV------------
GSTσ|Q8MUR9                 -------NYMTKS-DLVANHPNLQRVVDNVTSIESIRSWIDKRPKTEI------------
GSTσ|O18598                 -------NHMAKE-DLVANQPNLKALREKVLGLPAIKAWVAKRPPTDL------------
GSTσ|V9Q487                 -------NTMVGF-DLTKDYPNLAELKNKVLEIPAIKEWVDKRPKTDL------------
GSTσ|V9Q315                 -------NHMAGF-DITKDYANLAALKNKVLELPAIKEWIAKRPVSEV------------
GSTω|Q86D85                 DLLKFALGDKYEL-DK-ERFGKLLQWRELMEKDDAVKQ----SFISTEDHTKFLQSRKNG
GSTθ|Q8MUQ1                 --------RMAGY-DPCEGRPNLTQWMARVRESTNPYY--DQAHK---LVNKFAQDTASK
GSTζ|Q8MUQ5                 --------RRFHV-DL-RPYPIILRID-RELEGHPAFR--AAHPS---------------
                                              .       .

                                            % identity (compared to MG255130)
GST|Accession no.MG255130   --------------   100.0
GSTδ|V9Q2Q3                 --------------    32.9
GSTδ|V9Q3Y3                 --------------    43.8
GSTδ|V9Q4H6                 EKKNAQQK------    33.6
GSTδ|V9Q328                 SALRHPYHPVRASK    34.6
GSTδ|V9Q4A2                 LAEMSVD------K    42.7
GSTδ|O76483                 --------------    46.0
GSTδ|Q9XZ67                 --------------    48.4
GSTδ|Q60GK5                 --------------    55.6
GSTδ|Q8N0B3                 --------------    60.6
GSTδ|A9QUN5                 --------------    81.5
GSTδ|A7E1B7                 --------------    82.4
GSTδ|G8XWV6                 --------------    81.9
GSTδ|Q1M0Y4                 --------------    98.1
GSTδ|G8XWU4                 --------------    99.5
GSTδ|Q9GNE9                 --------------    58.3
GSTδ|P20432                 --------------    61.6
GSTδ|P42860                 --------------    61.1
GSTδ-like|Q2YFE3            AEKH----------    31.2
GSTε|Q9GPL9                 NAKL----------    39.5
GSTε|Q9GPL8                 NAKA----------    39.6
GSTσ|P46429                 --------------    22.7
GSTσ|V9Q3X5                 --------------    21.2
GSTσ|P41043                 --------------    11.0
GSTσ|Q8MUR9                 --------------    17.5
GSTσ|O18598                 --------------    16.8
GSTσ|V9Q487                 --------------    23.8
GSTσ|V9Q315                 --------------    21.7
GSTω|Q86D85                 ENNYDILA------    18.8
GSTθ|Q8MUQ1                 AKL-----------    26.5
GSTζ|Q8MUQ5                 ---NQPDCPPEAAK    23.3
```

Supplementary Figure 2. Alignment (Clustal Omega Multiple Sequence Alignment) and percent identity of amino acid sequence of *P. americana* rGST of this study (accession number MG255130) with GSTs of other insects in Figure 1.

```
         1         11        21        31        41        51       60
         |         |         |         |         |         |        |
         MTIDFYYLPGSAPCRSVLLAAKAIGVDLNLKVTNLMAGEHLTPEFLKMNPQHTIPTLNDN
           <-β1>    <----α1----->   <-β2>          <-α2->       <β3>
BCPred   ...............................................MNPQHTIPTLNDN
AAP      ..............................................NPQHTIPTLNDN
FBCPred  MTIDFYYLPGSAPCRSVLLA............................PQHTIPTLNDN
Pep1     MTIDFYYLPGSAPCRSVLLA----------------------------------------
Pep2     ----------------------------------------------MNPQHTIPTLNDN
Pep3     ------------------------------------------------------------
Pep4     ------------------------------------------------------------
Pep5     ------------------------------------------------------------
Pep6     ------------------------------------------------------------

         61        71        81        91        101       111      120
         |         |         |         |         |         |        |
         GFCLWESRAILSYLADQYGKDDSLYPKDAKKRALVDQRLYFDIRTLYHRFGEYYYPIYFA
          <β4> <----α3---->          <-----α4------>   <----α5----->
BCPred   GFCLWES...................................................PIYFA
AAP      GFCLWESR........QYGKDDSLYPKDAKKRALVD........................
FBCPred  GFCLWESRA..SYLADQYGKDDSLYPKDAKK...........................YFA
Pep1     ------------------------------------------------------------
Pep2     GFCLWESRA----------------------------------------------------
Pep3     GFCLWESRAILSYLADQYGK------------------------------------------
Pep4     -----------SYLADQYGKDDSLYPKDAKKRALVD------------------------
Pep5     ------------------------------------------------------------PIYFA
Pep6     ------------------------------------------------------------

         121       131       141       151       161       171      180
         |         |         |         |         |         |        |
         KQAADPEKMKKLEEAFEFLNKFLESQEFVAGNKLTIADLAIVSSVSTADIMGFDVSKYSN
              <-------α6------->            <------α7------>       <-
BCPred   KQAADPEKMKKLEEA...................................................
AAP      ..........................................................
fbcpred  KQAADPEKMKKLEEAFE.................................................
Pep1     ------------------------------------------------------------
Pep2     ------------------------------------------------------------
Pep3     ------------------------------------------------------------
Pep4     ------------------------------------------------------------
Pep5     KQAADPEKMKKLEEAFE--------------------------------------------
Pep6     ------------------------------------------------------------

         181       191       201       211
         |         |         |         |
         VAKWFEKCKKIVPGYEELNHSGCLKFKEMCDNLAKK 216
         ---α8----->       <-------α9------>
BCPred   ....................................
AAP      ......KCKKIVPGYEELNHSGCLKF..........
FBCPred  ....................................
Pep1     ------------------------------------
Pep2     ------------------------------------
Pep3     ------------------------------------
Pep4     ------------------------------------
Pep5     ------------------------------------
Pep6     ------KCKKIVPGYEELNHSGCLKF----------
```

GST-dv2 B-cell epitope prediction

BCPREDS Server 1.0

Submitted sequence: 216 amino acids
Epitope length: 20 amino acids
Classifier Specificty: 85%
Prediction method: bcpred
Use overlap filter: yes

BCPred Predictions

Position	Epitope	Score
48	MNPQHTIPTLNDNGFCLWES	0.945
116	PIYFAKQAADPEKMKKLEEA	0.932

```
1          11         21         31         41         51         60
|          |          |          |          |          |          |
MTIDFYYLPGSAPCRSVLLAAKAIGVDLNLKVTNLMAGEHLTPEFLKMNPQHTIPTLNDN 60
.................................................EEEEEEEEEEEEE
GFCLWESRAILSYLADQYGKDDSLYPKDAKKRALVDQRLYFDIRTLYHRFGEYYYPIYFA 120
EEEEEEE.................................................EEEEE
KQAADPEKMKKLEEAFEFLNKFLESQEFVAGNKLTIADLAIVSSVSTADIMGFDVSKYSN 180
EEEEEEEEEEEEEEE............................................
VAKWFEKCKKIVPGYEELNHSGCLKFKEMCDNLAKK 216
....................................
```

BCPREDS Server 1.0

Submitted sequence: 216 amino acids
Epitope length: 20 amino acids
Classifier Specificty: 85%
Prediction method: aap
Use overlap filter: yes

AAP Predictions

Position	Epitope	Score
187	KCKKIVPGYEELNHSGCLKF	1
49	NPQHTIPTLNDNGFCLWESR	1
77	QYGKDDSLYPKDAKKRALVD	1

```
1          11         21         31         41         51         60
|          |          |          |          |          |          |
MTIDFYYLPGSAPCRSVLLAAKAIGVDLNLKVTNLMAGEHLTPEFLKMNPQHTIPTLNDN 60
.................................................EEEEEEEEEEEEE
GFCLWESRAILSYLADQYGKDDSLYPKDAKKRALVDQRLYFDIRTLYHRFGEYYYPIYFA 120
EEEEEEEE........EEEEEEEEEEEEEEEEEEE........................
KQAADPEKMKKLEEAFEFLNKFLESQEFVAGNKLTIADLAIVSSVSTADIMGFDVSKYSN 180
............................................................
VAKWFEKCKKIVPGYEELNHSGCLKFKEMCDNLAKK 216
......EEEEEEEEEEEEEEEEEEEE..........
```

BCPREDS Server 1.0

Submitted sequence: 216 amino acids
Epitope length: 20 amino acids
Classifier Specificty: 85%
Prediction method: fbcpred
Use overlap filter: yes

FBCPred Predictions

Position	Epitope	Score
118	YFAKQAADPEKMKKLEEAFE	1
72	SYLADQYGKDDSLYPKDAKK	0.999
1	MTIDFYYLPGSAPCRSVLLA	0.98
50	PQHTIPTLNDNGFCLWESRA	0.975

```
1          11         21         31         41         51         60
|          |          |          |          |          |          |
MTIDFYYLPGSAPCRSVLLAAKAIGVDLNLKVTNLMAGEHLTPEFLKMNPQHTIPTLNDN 60
EEEEEEEEEEEEEEEEEEEE.............................EEEEEEEEEEE
GFCLWESRAILSYLADQYGKDDSLYPKDAKKRALVDQRLYFDIRTLYHRFGEYYYPIYFA 120
EEEEEEEEE..EEEEEEEEEEEEEEEEEEEE.........................EEE
KQAADPEKMKKLEEAFEFLNKFLESQEFVAGNKLTIADLAIVSSVSTADIMGFDVSKYSN 180
EEEEEEEEEEEEEEEEE...........................................
VAKWFEKCKKIVPGYEELNHSGCLKFKEMCDNLAKK 216
....................................
```

Supplementary Figure 3. Results of the *P. americana*-GST B-cell epitope prediction by using three different methods of the BepiPred 1.0 server. There were 2, 3, and 4 epitopes derived from the BCPred, AAP, and FBCPred prediction. These epitopes were aligned with the *P. americana*-GST sequence and they were matched to five different regions of the GST linear sequence. Peptides of these regions were synthesized. Because the regions 2 and the next were closed to each; therefore, a peptide that overlapped the two regions was synthesized. All six peptides were tested for binding to IgG and IgE in *P. americana* allergic patients' sera to determine their B-cell epitopes and allergenic epitopes.