

Ping Xiang · Eric J. Haas · Michael G. Zeece
John Markwell · Gautam Sarath

C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen

Received: 22 January 2004 / Accepted: 7 May 2004 / Published online: 14 July 2004
© Springer-Verlag 2004

Abstract Gly m Bd 28 K is a major soybean (*Glycine max* Merr.) glycoprotein allergen. It was originally identified as a 28 kDa polypeptide in soybean seed flour. However, the full-length protein is encoded by an open reading frame (ORF) of 473 amino acids, and contains a 23 kDa C-terminal polypeptide of as yet unknown allergenic and structural characteristics. IgE-binding (allergenic potential) of the Gly m Bd 28 K protein including the 23 kDa C-terminal portion as well as shorter fragments derived from the full-length ORF were evaluated using sera from soy-sensitive adults. All of these sera contained IgE that efficiently recognized the C-terminal region. Epitope mapping demonstrated that a dominant linear C-terminal IgE binding epitope resides between residues S256 and A270. Alanine scanning of this dominant epitope indicated that five amino acids, Y260, D261, D262, K264 and D266, contribute most towards IgE-binding. A model based on the structure of the β subunit of soybean β -conglycinin revealed that Gly m Bd 28 K contains two cupin domains. The dominant epitope is on the edge of the first β -sheet of the C-terminal cupin domain and is present on a potentially solvent-accessible loop connecting the two cupin

domains. Thus, the C-terminal 23 kDa polypeptide of Gly m Bd 28 K present in soy products is allergenic and apparently contains at least one immunodominant epitope near the edge of a cupin domain. This knowledge could be helpful in the future breeding of hypoallergenic soybeans.

Keywords Allergen · Cupin superfamily · Epitope mapping · *Glycine max* · IgE-binding · Recombinant fusion polypeptide

Abbreviations Ara h 1: *Arachis hypogaea* allergen 1 · Ara h 3: *Arachis hypogaea* allergen 3 · BCA: Bicinchoninic acid · Gly m Bd 28 K: *Glycine max*: band 28 kDa allergen · Gly m Bd 30 K: *Glycine max* band 30 kDa allergen · Gly m Bd 68 K: *Glycine max* band 68 kDa allergen · IgE: Immunoglobulin E

P. Xiang · J. Markwell · G. Sarath (✉)
Department of Biochemistry, University of Nebraska,
332 Keim Hall, East Campus, Lincoln,
NE 68583-0939, USA
E-mail: gsarath1@unl.edu
Tel.: +1-402-4724204
Fax: +1-402-4724020

E. J. Haas
Genetic Sequence Analysis Facility,
University of Nebraska Medical Center,
Omaha, Nebraska, USA

M. G. Zeece
Department of Food Science and Technology,
University of Nebraska, Lincoln,
NE 685803-0919, USA

G. Sarath
USDA-ARS Wheat Sorghum and Forage Research Unit,
University of Nebraska, Lincoln, NE 68583-0939, USA

Introduction

Soybean ranks among the eight most significant food allergens (Pedersen 1988; Zeece et al. 1999) and is the largest source of protein meal in the world (USDA-ERS 2003). Its wide usage in the food industry makes it difficult to eliminate from the diet of soy-sensitive individuals. Allergies to soybean have been increasingly studied over the last decade (Herian et al. 1990; Beardslee et al. 2000; Helm et al. 2000a, 2000b; Ogawa et al. 2000; Gu et al. 2001). There has also been strong interest in removing major allergens from soybeans, through food processing (Ogawa et al. 2000), mutations (Samoto et al. 1997; Ogawa et al. 2000) or through plant transformation (Herman 2003; Herman et al. 2003). In an elegant study, Herman et al. (2003) utilized plant transformation techniques to develop soybean lines essentially devoid of a major allergen: Gly m Bd 30 K. The transformed lines were essentially identical to non-transformed plants (Herman et al. 2003), raising the

possibility of successfully engineering hypoallergenic elite lines in the near future. However, as pointed out succinctly by Herman (2003), there are many other hurdles that have yet to be overcome before genetically modified hypoallergenic food crops are grown commercially.

Soybean seeds contain many proteins of varying allergenicity. These include the G1 and G2 glycinin proteins, Gly m Bd 30 K (a thiol protease-like 34 kDa protein), Gly m Bd 68 K (the 70 kDa α -subunit of β -conglycinin) and Gly m Bd 28 K (Beardslee et al. 2000; Helm et al. 2000a, 2000b; Ogawa et al. 2000; Xiang et al. 2002). While all of these allergens have been studied to different extents, several important allergenic features of Gly m Bd 28 K have not yet been evaluated.

Gly m Bd 28 K was first isolated from soybean meal as a 28 kDa glycosylated protein (Tsuji et al. 1997). It was later demonstrated that it had a single glycosylation site, with an N-linked glycan moiety attached to Asn—the 20th amino acid in the purified protein (Hiemori et al. 2000). Interestingly, the cloning of the cDNA for Gly m Bd 28 K has shown that the protein is about twice as large as anticipated from its initial discovery as a 28 kDa polypeptide on SDS gels (Tsuji et al. 2001). The full open reading frame (ORF) coding for Gly m Bd 28 K consists of a preproprotein of 473 amino acids, with a predicted 21 amino acid signal peptide. Overall, Gly m Bd 28 K shares significant homology with the pumpkin MP27/MP32 protein (50.4% identity) and a carrot globulin-like protein, Gea8 (45.9% identity). The Gly m Bd 28 K preprotein is suggested to be further processed in the seed into two mature proteins, an N-terminal polypeptide of about 240 amino acids and a C-terminal polypeptide of about 212 amino acids, in a manner similar to the processing of the pumpkin MP27/MP32 preprotein (Tsuji et al. 2001).

The allergenic potential of Gly m Bd 28 K has been demonstrated by elevated serum IgE binding ability in patients with soy sensitivities. To date, most of the IgE-binding studies to Gly m Bd 28 K have been limited to the N-terminal, approximately 240 residue, fragment initially purified from soybean seeds (Tsuji et al. 1997, 2001; Heimori et al. 2000). In one study, the N-linked carbohydrates on this polypeptide apparently exhibited greater IgE-binding within the sera tested than the deglycosylated amino acid backbone (Heimori et al. 2000). However, IgE binding to this protein could not be fully abolished by the inhibition of the IgE binding to its glycan moiety (Heimori et al. 2000). Furthermore, a recombinant polypeptide, corresponding to Phe22–Lys259 of the full length Gly m Bd 28 K ORF still retained significant IgE binding ability (Tsuji et al. 2001). These data indicated that the deglycosylated N-terminal polypeptide of Gly m Bd 28 K contained IgE-binding epitopes and potentially contributed to the allergic responses documented in vivo. Since the discovery of the C-

terminally cleaved region of Gly m Bd 28 K, however, no further studies have investigated the allergenicity, IgE-binding or predicted tertiary structure of this approximately 212 amino acid novel polypeptide.

In this paper, we have dissected the potential linear allergenic domains of Gly m Bd 28 K beginning with the recombinant, full-length, 473 residue preprotein. Overlapping fragments spanning N-terminal and C-terminal portions of the Gly m Bd 28 K protein were constructed, and their IgE binding abilities were evaluated using sera from soybean-sensitive individuals. We have demonstrated that both the N- and C-terminal fragments of Gly m Bd 28 K can bind IgE, and have characterized a novel dominant sequential epitope on the C-terminal fragment of the protein. The Gly m Bd 28 K allergen is probably processed into two smaller polypeptides of 240 and 212 amino acids in the soybean seed, and both portions are expected to be present in soybean-derived foods. Our data suggest that the C-terminal polypeptide has been an overlooked contributor to soybean sensitivities. This hypothesis is supported by sequence alignments of Gly m Bd 28 K with members of the cupin family, which contain several known plant allergens (Mills et al. 2002). We have also generated a homology model of Gly m Bd 28 K and found that the novel, dominant, IgE-binding epitope identified in this work may reside on the edge of a loop region that connects the two cupin domains. This position may contribute to its allergenic potential through enhanced surface exposure.

Materials and methods

Construction of recombinant fusion proteins of Gly m Bd 28 K

The cDNA clone of Gly m Bd 28 K was obtained from Pioneer Hybrid (generous gift of Rudolf Jung, Pioneer Hybrid, Des Moines, Iowa). This clone starts at a Met, which corresponds to the 20th amino acid in the sequence reported by Tsuji et al. (2001; DDBJ accession no. AB046874). In addition, our clone has two sequence differences, namely, R4 and K333 based upon the starting residue of our clone, which correspond to H23 and R352, respectively, in the sequence reported by Tsuji et al. (2001).

Eight thioredoxin-fusion proteins representing Gly m Bd 28 K and overlapping fragments of Gly m Bd 28 K were constructed. PCR primers (P28FNDEI, 5'-catatggccttccgtgatg; P28RECORI, 5'-gaattcaaaaacatc-cataacca; P28F1R231, 5'-gaattccctccatgacctact; P28F-2M215, 5'-catatgcaagaccaagaggagg; P28F11L120, 5'-gaattcaagtctctgaccttctct; P28F3L100, 5'-catatgttgta-catgattccatctggt; P28F3F345, 5'-gaattcaaacagtcacctt; P28F13F45, 5'-gaattcaaatatcctaccacca; P28F14M40, 5'-catatgcatgggtgtaggatattt; P28F14L100, 5'-gaattccaagtc-cctgtctt) were designed for read-through, and all primers were synthesized by Sigma-Genosys (The

Woodlands, Tex.). PCR reactions were performed for 35 cycles using *Taq* DNA polymerase (Gibco-BRL). PCR products were purified from 1% agarose gels and ligated into the pBAD/TOPO vector using the TA system (Invitrogen). The ligated DNA was transformed into *Escherichia coli* Top10 cells following the manufacturer's recommended protocols (Invitrogen). Constructs containing inserts with the predicted sizes and correct orientation were verified through endonuclease mapping and sequenced in the Genomics Core Research Facility of the University of Nebraska-Lincoln.

Expression and purification of recombinant Gly m Bd 28 K proteins

E. coli containing the appropriate constructs were grown in 5 ml LB medium containing ampicillin (50 µg/ml) overnight at 37°C, and used to inoculate 250 ml LB medium containing ampicillin (50 µg/ml). Cultures were grown at 37°C with shaking at 200 rpm until the OD₆₀₀ reached approximately 0.5. At that time, cells were induced with 0.02% L-arabinose. Cells were grown for an additional 3–4 h at 37°C, and harvested by centrifugation at 5,000 g for 5 min at 4°C. Purification of the recombinant proteins was as described earlier (Beardslee et al. 2000; Xiang et al. 2002). Fusion protein purity was evaluated by SDS-PAGE as described below. Protein concentrations were determined by the BCA assay (Pierce). Purified proteins were stored in aliquots at –20°C until used.

Electrophoretic separation

Aliquots containing 4 µg of each recombinant protein were analyzed by SDS-PAGE on 12.5% SDS gels (Laemmli 1970). Proteins were stained with 0.1% Coomassie Brilliant Blue R-250.

Human sera

Sera from adult soybean-allergic patients with high soybean specific IgE levels were purchased from Plasma-Labs (Everett, Wash.). All of these adult individuals had given consent (to Plasma Labs) for collection of their plasma for research studies, and were found to be allergic to soy and peanut extracts by a skin-prick test. Sera from people with high IgE level but no reported soybean allergic reactions or those lacking soybean specific IgE were used as negative controls (Xiang et al. 2002).

Immunoblotting

Proteins separated by SDS-PAGE were transferred to polyvinylidene fluoride membranes (Millipore, Immobilon-P) in a tank apparatus overnight at 20 V in transfer buffer (10 mM Tris-HCl, 100 mM glycine, 10% methanol) at 4°C. Membranes were probed with sera

and antigen-antibody complexes were detected essentially as described previously (Xiang et al. 2002).

Peptide synthesis on SPOT membrane

Individual peptides were synthesized on a modified SPOT cellulose membrane (Sigma-Genosys) containing free hydroxyl groups using Fmoc amino acids as described previously (Xiang et al. 2002). Peptide-containing membranes were stored at –20°C until needed or used directly for the IgE-binding assay as described below.

IgE-binding assay of SPOT membrane

The SPOT membrane containing synthetic peptides was washed with Tris-buffered saline (TBS) and blocked overnight at 4°C with blocking buffer [TBST (TBS-Tween 20) containing 5% sucrose and 2× membrane blocking buffer from Sigma-Genosys]. The membrane was incubated overnight at 4°C with the appropriate human serum diluted 1:3 in blocking buffer, washed in TBST and incubated with goat anti-human IgE-horse-radish peroxidase (Bethyl-Labs, Montgomery, Tex.) diluted to 0.15 µg/ml in blocking buffer and developed using chemiluminescence.

Sequence analysis and model construction of Gly m Bd 28 K

Pfam searches of Gly m Bd 28 K sequence were conducted using Pfam 8.0 (<http://pfam.wustl.edu/hmm-search.shtml>; Bateman et al. 2002). Similarity searches of Gly m Bd 28 K sequence were performed using BLAST 2 from NCBI (<http://www.ncbi.nlm.nih.gov/>). A tentative model of the Gly m Bd 28 K structure was generated using the 3D-PSSM server (Kelley et al. 2000). 3D-PSSM was the only automated modeling program that preserved the alignment of Gly m Bd 28 K with known cupin domains. Of the models generated by 3D-PSSM, we chose the one based upon the β-conglycinin structure (PDB identifier 1ipj; Berman et al. 2000; Maruyama et al. 2001). We used percent identity and similarity to infer homology, and utilized the power of the 3D-PSSM server to model distant homologs of proteins. Although speculative, our models based upon these distant homologs provides insights into the structure and function of these new, relatively uncharacterized proteins.

Results

Construction, expression and purification of recombinant Gly m Bd 28 K

Recombinant Gly m Bd 28 K and its fragment proteins were constructed as thioredoxin fusions, and contained a

6× histidine tag at their C-termini. The respective fragment lengths and position on the intact protein are shown in Fig. 1. Based on a predicted processing site between N242 and E243 (Tsuji et al. 2001), and the initial discovery of this allergen as an approximately 28 kDa protein, we first constructed two fusion fragments: f1 (the N-terminal 231 residues) and f2 (the C-terminal 240 residues). Initial experiments revealed that both fragments bound to IgE from sera of soy-allergic individuals. To learn more about the IgE binding of these two fragments, several smaller and overlapping fusions were also generated. Purity of all recombinant proteins was verified by SDS-PAGE, and generally exceeded 90% (Fig. 2a).

Identification of IgE binding regions by immunoblotting

Sera from eight soy-sensitive patients were tested; six of these contained IgE antibodies recognizing recombinant Gly m Bd 28 K. One of the sera also bound strongly to the thioredoxin fusion partner and was not studied further. The other five sera were tested for their recognition of the recombinant Gly m Bd 28 K and its fragments. A representative immunoblot obtained using one of the five sera from soy-sensitive patients is shown in Fig. 2b. Individual sera were tested for the presence of IgE directed towards different portions of the Gly m Bd 28 K protein. The developed X-ray films were scored independently by the authors, who compared and scored the relative density obtained for each fragment from the six immunoblots on a scale of 4 (high) to 0 (negligible) and the pooled data were averaged. The average qualitative scores are shown in Table 1. Generally, there was negligible or very weak response of these soybean-sen-

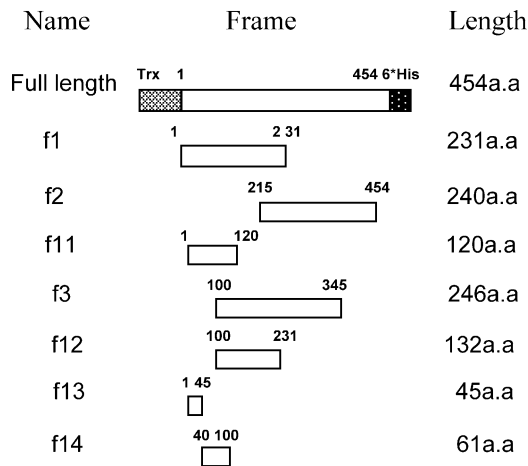


Fig. 1 Construction of recombinant proteins based on the sequence of Gly m Bd 28 K. Each construct contains a 13 kDa thioredoxin fusion partner on the N-terminus and a 3 kDa 6× His tag on the C-terminus. For simplicity, the fusion partners are indicated for the full-length Gly m Bd 28 K construct and omitted on other constructs in this figure

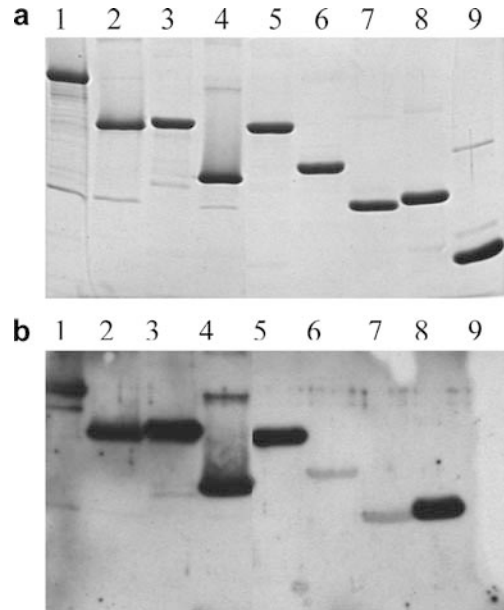


Fig. 2 a SDS-PAGE of purified recombinant Gly m Bd 28 K and its fragments. b Representative immunoblot of recombinant Gly m Bd 28 K and its fragments with a serum sample from a soybean-sensitive individual (#1); 4 µg of each polypeptide was loaded and separated on a 12.5% polyacrylamide gel. Immunoblotting was performed as described in Materials and methods. Lanes: 1 Full length protein; 2–8 fragments f1, f2, f11, f3, f12, f13, f14, respectively; 9 thioredoxin

sitive sera to the thioredoxin fusion partner. The non-allergic sera used as a control did not detect any of the target Gly m Bd 28 K fragments tested, confirming that reactivity of the sera to Gly m Bd 28 K correlates to soy sensitivities in human patients.

The profile of IgE-binding to the Gly m Bd 28 K fragments showed that both the 231 residue N-terminal region, f1, and the 240 residues C-terminal fragment, f2, contributed to observed IgE binding of the full-length protein, indicating that epitopes existed in both halves of the allergen. Based upon the binding frequency and relative densities scored from immunoblots, the C-terminal f2 region appeared to display stronger IgE binding than the N-terminal f1 region (Table 1). This finding was novel, since this was the first observation of IgE binding to the C-terminal fragment. We next narrowed the region of the polypeptide that contained IgE-binding domains. Fragment f11 (encompassing the first 120 residues from the N-terminus) bound to IgE in sera from soy-sensitive individuals, as did fragment f3 (residues L100–F345); however, fragment f12 (residues L100–R231) did not bind to IgE with significant avidity. These results indicate that IgE binding epitope(s) are located within the first 120 amino acids from the N-terminus, and for the first time indicated IgE-binding sites present between residues 231 and 345 of the full-length Gly m Bd 28 K protein.

We also performed less detailed studies of the N-terminal epitope binding region. These analyses revealed that IgE-binding to residues M1–M40 (fragment

Table 1 IgE binding ability of recombinant full-length Gly m Bd 28K and fusion fragments. Relative density rating: 0 no significant IgE binding, 1 weak IgE binding, 2 moderate IgE binding, 3 strong IgE binding, 4 very strong IgE binding

Fragment name Fragment region	Full-length 1-454	Relative density of immunoblot							
		f1 1-231	f2 215-454	f11 1-120	f3 100-345	f12 100-231	f13 1-45	f14 40-100	
Serum #1	3	3	4	3	3	1	1	3	
Serum #2	3	3.5	4	4	3	2	2	3	
Serum #3	1.5	3	3	2	2.5	1.5	1.5	2.5	
Serum #4	1	0.5	2	0.5	0.5	0	0	0.5	
Serum #5	2.5	1	3	0.5	3.5	0	0	1	
Nonallergic sera	0	0	0	0	0	0	0	0	

f13) was relatively weak, whereas much greater binding (on average) of IgE to fragment f14 (M40-L100) was documented for the five soy-sensitive sera tested (Table 1). These results indicate that strong IgE binding epitope(s) exist between residues M40 and L100 of Gly m Bd 28 K.

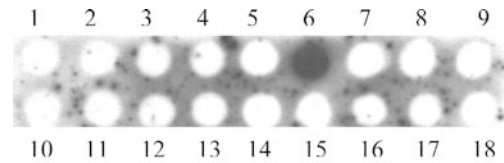
The novel, potentially strong, allergenic region of the C-terminal portion of the full-length Gly m Bd 28 K molecule was mapped in greater detail and is described in the following sections.

IgE binding epitopes in the C-terminal fragment

The documentation of strong IgE-binding in an as yet undocumented portion of the allergen permitted us to identify the IgE binding epitope(s) within the C-terminal portion of Gly m Bd 28 K. We tested IgE binding of the region encompassing residues E221 to C353, which contains the relatively strong IgE-binding domains documented for fragment f2 and f3 (Table 1). Eighteen 15-mer peptides designed with 8 amino acid overlap were synthesized on modified cellulose paper and probed with an aliquot of soy-sensitive sera obtained by pooling equal amounts of sera from the five soy-sensitive individuals (see Table 1). Only one peptide, peptide 6, corresponding to the region S256-A270 of Gly m Bd 28 K, exhibited strong IgE binding (Fig. 3).

Alanine scanning of the C-terminal epitope

To identify the amino acids critical for IgE binding in peptide 6, a series of related peptides were synthesized on a SPOT membrane with alanine substitutions at each position of the 15-mer epitope, and their IgE binding abilities evaluated (Fig. 4). Of the 15 amino acids on this IgE epitope, alanine substitutions at Y260, D261, D262 and K264 essentially abolished binding to the IgE present in the pooled soy-sensitive sera, whereas substitution of D266 with alanine greatly decreased IgE binding. Thus, the central IgE-binding region of this peptide is localized between Y260 and D266. The amino acids most critical for IgE binding were identified as Y260, D261, D262 and K264. However, residue D266 also contributed towards IgE binding.



Spot	Position	Sequence
1	221-235	EEEEQTSRSWRKLLLE
2	228-242	RSWRKLLLETVEGKVN
3	235-249	ETVFGKVNEDIENKD
4	242-256	NEKIENKDTAGSPAS
5	249-263	DTAGSPASNYLYDDK
6	256-270	SYNLYDDKKADFKNA
7	263-277	KKADFKNAYGWSKAL
8	270-284	AYGWSKALHGGEYPP
9	277-291	LHGGEYPPLEPDIG
10	284-298	PLSEPDIGVLLVKLS
11	291-305	GVLLVKLSAGSMLAP
12	298-312	SAGSMLAPHVNPISD
13	305-319	PHVNPISDEYTIIVLS
14	312-326	DEYTIIVLSGYGELHI
15	319-333	SGYGELHIGYPNGSK
16	326-340	IGYPNGSKAMKTKIK
17	333-347	KAMKTKIKQGDVFFVV
18	340-353	KOGDVFVVPYRFFFC

Fig. 3 Epitope mapping of Gly m Bd 28 K C-terminal IgE-binding region E221-C353. Eighteen 15-mer peptides representing the region E221-C353 were manually synthesized on a SPOT membrane as described in Materials and methods. Adjacent peptides had an eight amino acid overlap. The membrane was probed with a pool of soybean-sensitive sera and bound IgE was detected by chemiluminescence using a goat anti-human IgE-horseradish peroxidase conjugate. Peptide 6 (*bold, underlined*), corresponding to residues S256-A270 of Gly m Bd 28 K protein, exhibited strong IgE binding

Sequence analysis and model construction of Gly m Bd 28 K

Sequence analysis indicated that Gly m Bd 28 K belongs to the cupin superfamily and contains two cupin domains, as has been documented for several legume protein allergens (Mills et al. 2002). The cupin domains span residues F18-S148 near the N-terminal end and Y257-R407 near the C-terminal portion of the full-length protein. Therefore, processing by cleavage at residue 242, as predicted to occur in soybeans (Tsuji et al. 2001), would yield two proteins, each containing one cupin domain.

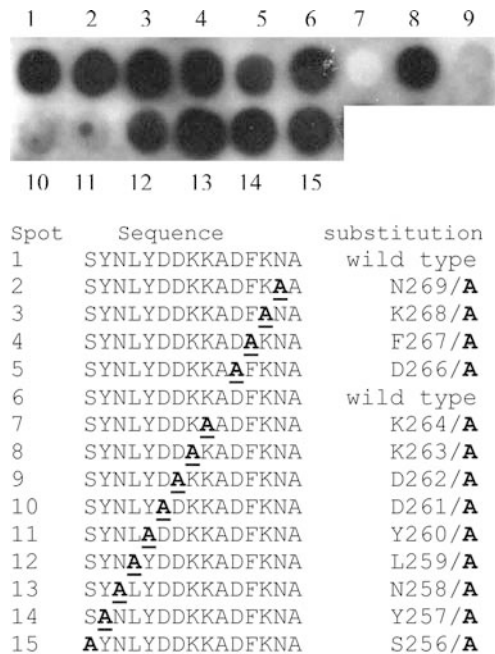
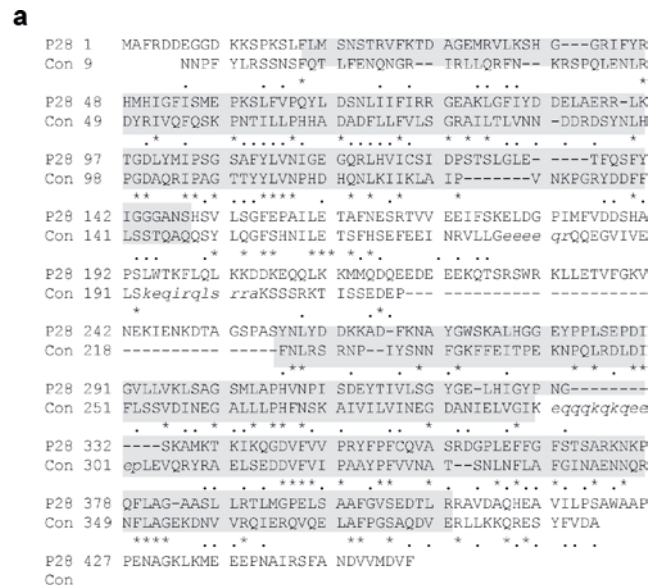


Fig. 4 Alanine scanning of residues important for IgE-binding in peptide 6 shown in Fig. 3. Fifteen peptides designated as spots 1–15 were manually synthesized on a SPOT membrane. Peptides 1 and 6 represented the wild type sequence. Alanine substitutions are *underlined* and in *bold*. The membrane was probed with the soybean-sensitive sera (1:3 dilution) and bound IgE was detected as described in Fig. 3

A tentative model of the Gly m Bd 28 K structure was generated by 3D-PSSM (Kelley et al. 2000), using soybean β -conglycinin as a template for the new structure (Fig. 5a). This model was chosen because the β -conglycinin sequence gave the best alignment with Gly m Bd 28K in the cupin domains. In this model, the dominant IgE-binding epitope (S256–A270) present on the C-terminal half of the protein is expected to be near the beginning of the C-terminal cupin domain, and is colored green in Fig. 5b. Interestingly, this IgE-binding epitope in the C-terminal domain of Gly m Bd 28 K occurs in a similar position to two epitopes of a peanut cupin allergen, Ara h 1 (epitope 14, D393–E402 and epitope 15, N409–K418) identified by Shin et al. (1998).

Discussion

In this study, we have identified two major IgE-binding sites in Gly m Bd 28 K using recombinant fusion fragments designed from the entire ORF and sera from soybean-sensitive patients. Based on the intensity and binding frequency of IgE in sera obtained from soybean-sensitive individuals, the C-terminal half of Gly m Bd 28 K showed slightly stronger binding as compared to the N-terminal half of this protein. Furthermore, using synthetic peptides, we located an important IgE binding region in the C-terminal 23 kDa polypeptide



b

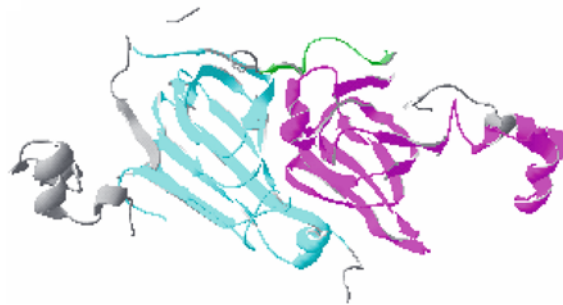


Fig. 5a,b A putative model of Gly m Bd 28 K based on the solved structure of the β -subunit of soybean β -conglycinin. **a** Sequence alignment of Gly m Bd 28 K (P28) to the β -subunit of soybean β -conglycinin (Con). Asterisks Residues identical in all aligned sequences, dots conserved residues. The aligned cupin domains are shaded. Residues that were absent in the PDB 1IPK structure and modeled using alanine placeholders are shown in lowercase italics. **b** Ribbon diagram of a putative structure of Gly m Bd 28 K. The two cupin domains are represented in blue and purple, respectively, and the immunodominant epitope is highlighted in green

of Gly m Bd 28 K, spanning residues S256–A270 (SYNLYDDKKADFKNA). The residues important for IgE-binding on this epitope were identified through alanine scanning and shown to be Y260, D261, D262, K264 and D266. Since all of these residues are in the middle of the epitope, it probably explains why the peptides corresponding to D249–K263 and K263–L277, which are adjacent in the linear sequence of the protein to this epitope, did not show IgE binding (Fig. 3). Finding a dominant allergenic epitope in proteins is not unusual (Xiang et al. 2002); however, it is known that Gly m Bd 28 K also contains strong IgE-binding regions in the N-terminal half of the protein, and other epitopes with less affinity for IgE binding (Ogawa et al. 2000; see also Table 1, Fig. 2, this work). The expectation is that each allergic individual will have a unique signature of

epitope recognition, but that dominant epitopes will be present for most allergenic proteins. Also, our mapping methods were directed at linear epitopes and as such would not have detected potential conformational epitopes. Future research using other methods could be used to highlight both the linear and conformational epitopes present on the entire Gly m Bd 28 K ORF.

Our results indicate that the sera used in this study cross-reacted to the Gly m Bd 28 K C-terminal portion of the protein, and efficiently recognized a dominant linear epitope in this, as yet uncharacterized, region of this known soy allergen. It is also possible that the IgE would cross-react to other related soybean and peanut proteins. As an example, we have shown in a previous study that both soybean and peanut glycinins contain a conserved IgE-binding epitope (Xiang et al. 2002).

The strong IgE binding to the C-terminal portion of Gly m Bd 28 K suggests that the two polypeptides arising from the predicted processing of the full length Gly m Bd 28 K (Tsuji et al. 2001) are present in soybean-derived foods and elicit IgE production in soybean-sensitive patients. This C-terminal polypeptide derived from Gly m Bd 28 K has a predicted size of 23 kDa and a predicted pI of 5.62, which will place it in a region where several abundant soybean proteins and allergens, such as the soybean Kunitz trypsin inhibitor and the basic chains of the glycinins (Burks et al. 1994; Helm et al. 2000a, 2000b), would normally migrate in 1- and 2-D gels. This may have precluded its prior identification as an IgE-binding protein in soy extracts and would make its present direct identification in soybean extracts quite challenging. The availability of IgG antibodies specifically directed towards the 23 kDa fragment should readily resolve this problem.

Interestingly, Gly m Bd 28 K belongs to the cupin protein superfamily, which contains a number of legume food allergens, such as the peanut Ara h 1, Ara h 3, soybean β -conglycinin and soybean glycinin G1 and G2 proteins (Mills et al. 2002). Most of these food allergens contain two cupin domains, as we have documented for Gly m Bd 28 K. Even though the overall sequence identity between Gly m Bd 28 K and other members of the cupin family is not high, Gly m Bd 28 K clearly shares somewhat greater sequence similarity to the 7S vicillin family, e.g., soybean conglycinin (sequence identity of 23%) than to the 11S legumin family, e.g., soybean glycinin (sequence identity of 20%). Thus, the modified structure of the β -subunit of soybean β -conglycinin protein was used as a template to model a potential structure of Gly m Bd 28 K. However, the region linking the N- and C-terminal cupin domains in Gly m Bd 28 K is longer than most of the vicillins, and considerably longer than the similar region of soybean β -conglycinin. This linker region seems to be a hypervariable site among legume seed storage proteins and contains specific sites for proteolytic cleavage in vivo. For example, the soybean glycinins are proteolytically processed in vivo in the

linker region joining the cupin domains to form the acidic and basic chains of the mature proteins (Shutov et al. 1995, 1996, 1998, 1999). Therefore, these regions of plant cupin allergens could be points of attack in the mammalian digestive tract, yielding potentially allergenic peptides.

The higher overall similarity of Gly m Bd 28 K to the 7S vicillins, except for the extended linker region between the two cupin domains, is also shared by two other cupin family proteins, namely the MP27/MP32 precursor protein from pumpkin and the globulin-like protein Gea8 from carrot (Shutov et al. 1998). Together, these three related proteins may form a separate branch in the seed storage globulin evolutionary tree built by Shutov et al. (1998). The extended linker region between the two cupin domains of these proteins may be a prerequisite for the legumin-like processing, and provide a site for proteolytic attack as occurs for the MP27/MP32 precursor protein (Shutov et al. 1998). It is still unclear if Gly m Bd 28 K is processed into two separate polypeptides, and if these two chains are held together by hydrophobic interactions as occurs for soybean glycinins (Mills et al. 2002). However, the recognition of both parts of the Gly m Bd 28 K protein by IgE from soybean-sensitive individuals suggests that they are indeed present in vivo. Furthermore, the C-terminal epitope of Gly m Bd 28 K identified in this work is at the beginning of the second cupin domain, and near the predicted cleavage site (N242; Tsuji et al. 2001) of this protein. Structurally, this linker region forms a small flat cavity between the two cupin domains and the dominant epitope is expected to be surface exposed (Fig. 5b). This type of structure may represent a potential allergenic region in several proteins. For example, two epitopes identified in the peanut allergen, Ara h 1 (Shin et al. 1998) are also located in a partially solvent exposed edge of the predicted cupin domain of the Ara h 1 protein.

The detection of a novel IgE-binding epitope on the C-terminal portion of the full-length Gly m Bd 28 K protein suggests that soybean seeds could contain as yet undocumented allergens. The presence of this and other epitopes on partially solvent-exposed surfaces on several legume cupin allergens might be indicative of regions on these proteins that are more prone to becoming accessible to the mammalian immune system following the digestion process. It is known that cupin domains are generally more resistant to proteolytic digestion (Mills et al. 2002) and proteins with strong β -sheet structure, such as the soybean Kunitz trypsin inhibitor, are quite stable to a number of physical and chemical denaturants (Roychaudhuri et al. 2003).

Acknowledgements We thank Dr. Rudolf Jung at Pioneer-Hybrid International for the gift of the soybean Gly m Bd 28 K clone. This work was supported in part by NIH Grant Number 1 P20 RR16469 from the BRIN Program of the National Center for Research Resources (G.S.). This work is published as Journal Series No. 14,087 from the Agriculture Research Division, University of Nebraska.

References

- Bateman A, Birney E, Cerruti L, Durbin R, Ewinger L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) The Pfam protein families database. *Nucleic Acids Res* 30:276–280
- Beardslee TA, Zeece MG, Sarath G, Markwell JP (2000) Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int Arch Allergy Immunol* 123:299–307
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
- Burks AW, Cockrell G, Connaughton C, Guin J, Allen W, Helm RM (1994) Identification of peanut agglutinin and soybean trypsin inhibitor as minor legume allergens. *Int Arch Allergy Immunol* 105:143–149
- Gu X, Beardslee T, Zeece M, Sarath G, Markwell J (2001) Identification of IgE-binding proteins in soy lecithin. *Int Arch Allergy Immunol* 126:218–225
- Helm RM, Cockrell G, Connaughton C, Sampson HA, Bannon GA, Beilinson V, Livingstone D, Nielsen NC, Burks AW (2000a) A soybean G2 glycinin allergen 1. Identification and characterization. *Int Arch Allergy Immunol* 123:205–212
- Helm RM, Cockrell G, Connaughton C, Sampson HA, Bannon GA, Beilinson V, Nielsen NC, Burks AW (2000b) A soybean G2 glycinin allergen 2. Epitope mapping and three-dimensional modeling. *Int Arch Allergy Immunol* 123:213–219
- Herian AM, Taylor SL, Bush RK (1990) Identification of soybean allergens by immunoblotting with sera from soy-allergic adults. *Int Arch Allergy Appl Immunol* 92:193–198
- Herman EM (2003) Genetically modified soybeans and food allergies. *J Exp Bot* 54:1317–1319
- Herman EM, Helm RM, Jung R, Kinney AJ (2003) Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol* 132:36–43
- Hiemori M, Bando N, Ogawa T, Shimada H, Tsuji H, Yamanishi R, Terao J (2000) Occurrence of IgE antibody-recognizing N-linked glycan moiety of a soybean allergen, Gly m Bd 28 K. *Int Arch Allergy Immunol* 144:238–245
- Kelley LA, MacCallum RM, Sternberg MJE (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 299:499–520
- Laemmli UK (1970) Cleavage of structure proteins during assembly of the head of bacteriophage. *Nature* 227:680–685
- Maruyama N, Adachi M, Takahashi K, Yagasaki K, Kohno M, Takenaka Y, Okuda E, Nakagawa S, Mikami B, Utsumi S (2001) Crystal structures of recombinant and native soybean beta-conglycinin beta homotrimers. *Eur J Biochem* 268:3595–3604
- Mills EN, Jenkins J, Marigheto N, Belton PS, Gunning AP, Morris VJ (2002) Allergens of the cupin superfamily. *Biochem Soc Trans* 30:925–929
- Ogawa A, Samoto M, Takahashi K (2000) Soybean allergens and hypoallergenic soybean products. *J Nutr Sci Vitaminol* 46:271–279
- Pedersen HE (1988) Allergenicity of soybean proteins. In: Applewhite TH (ed) *Vegetable protein utilization in human foods and animal feedstuffs*. American Oil Chemists Society, Champaign, Ill., pp 204–211
- Roychoudhuri R, Sarath G, Zeece M, Markwell J (2003) Reversible denaturation of the soybean Kunitz trypsin inhibitor. *Arch Biochem Biophys* 412:20–26
- Samoto M, Fukuda Y, Takahashi K, Heimori M, Tsuji H, Ogawa T, Kawamura Y (1997) Substantially complete removal of three major allergenic soybean proteins (Gly m Bd 30 K, Gly m Bd 28 K and the alpha-subunit of conglycinin) from soy protein by using mutant soybean, Tohoku 124. *Biosci Biotechnol Biochem* 61:2148–2150
- Shin DS, Compadre CM, Maleki SJ, Kopper RA, Sampson H, Huang SK, Burks AW, Bannon GA (1998) Biochemical and structural analysis of the IgE binding sites on ara h 1, an abundant and highly allergenic peanut protein. *J Biol Chem* 273:13753–13759
- Shutov AD, Kakhovskaya IA, Braun H, Baumlein H, Muntz K (1995) Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. *J Mol Evol* 41:1057–1069
- Shutov AD, Kakhovskaya IA, Bastrygina AS, Bulmaga VP, Horstmann C, Muntz K (1996) Limited proteolysis of beta-conglycinin and glycinin, the 7S and 11S storage globulins from soybean [*Glycine max* (L.) Merr.]. Structural and evolutionary implications. *Eur J Biochem* 241:221–228
- Shutov AD, Braun H, Chesnokov YV, Baumlein H (1998) A gene encoding a vicilin-like protein is specifically expressed in fern spores. Evolutionary pathway of seed storage globulins. *Eur J Biochem* 252:79–89
- Shutov AD, Blattner FR, Baumlein H (1999) Evolution of a conserved protein module from Archaea to plants. *Trends Genet* 15:348–349
- Tsuji H, Bando N, Hiemori M, Yamanishi R, Kimoto M, Nishikawa K, Ogawa T (1997) Purification of characterization of soybean allergen Gly m Bd 28 K. *Biosci Biotechnol Biochem* 61:942–947
- Tsuji H, Hiemori M, Kimoto M, Yamashita H, Kobatake R, Adachi M, Fukuda T, Bando N, Okita M, Utsumi S (2001) Cloning of cDNA encoding a soybean allergen, Gly m Bd 28 K. *Biochim Biophys Acta* 1518:178–182
- USDA-ERS (2003) Oil crops year book (online). Available at <http://ers.usda.gov/publications>
- Xiang P, Beardslee TA, Zeece MG, Markwell J, Sarath G (2002) Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Arch Biochem Biophys* 408:51–57
- Zeece MG, Beardslee TA, Markwell JP, Sarath G (1999) Identification of an IgE-binding region in soybean acidic glycinin G1. *Food Agric Immunol* 11:83–90