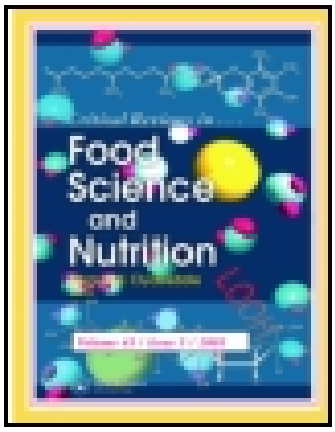

Critical Reviews in Food Science and Nutrition
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/bfsn20

Structural and Functional Analysis of Various Globulin Proteins from Soy Seed

Amandeep Singh [a] , Megha Meena [a] , Dhiraj Kumar [a] , Ashok K. Dubey [a] & Imtaiyaz Hassan [b]
[a] Division of Biotechnology, Netaji Subhas Institute of Technology , Azad Hind Fauz Marg, Sector-3, Dwarka, New Delhi , 110078 , India
[b] Centre for Interdisciplinary Research in Basic Sciences , Jamia Millia Islamia, Jamia Nagar, New Delhi , 110025 , India
Accepted author version posted online: 23 Oct 2013.

Running title: A review on globulins of Glycine max

Structural and Functional Analysis of Various Globulin Proteins from Soy Seed

AMANDEEP SINGH[1], MEGHA MEENA[1], DHIRAJ KUMAR[1], ASHOK K. DUBEY[1] AND MD. IMTAIYAZ HASSAN[2,*]

[1]*Division of Biotechnology, Netaji Subhas Institute of Technology, Azad Hind Fauz Marg, Sector-3, Dwarka, New Delhi – 110078, India.*

[2]*Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi 110025, India.*

All Correspondence should be addressed to,

Md. Imtaiyaz Hassan, Ph.D.

Centre for Interdisciplinary Research in Basic Sciences,

Jamia Millia Islamia, Jamia Nagar, New Delhi 110025, India

Telefax: +91-11-2698-3409

E-mail: mihassan@jmi.ac.in

Abstract

Storage proteins of soybean mostly consist of globulins, which are classified according to their sedimentation coefficient. Among 4 major types: 2S, 7S, 11S and 15S of globulins, 7S and 11S constitute major fraction. The 11S fraction consists only of glycinin and 7S fraction majorly consists of β-conglycinin, small amounts of γ-conglycinin and basic 7S globulin (Bg7S). Glycinin exist as a hexamer while β-conglycinin as a trimer and Bg7S as a tetramer. Glycinin subunits are coded by 5 genes of a family, whereas about 15 genes are present for β-conglycinin subunits. Bg7S gene is present in 4 copies in soybean genome. Synthesis of all proteins takes place as a single polypeptide chain, which is cleaved after folding to yield different chains or subunits. Glycinin and β-Conglycinin are made for storage purpose however; Bg7S has potential xylanase inhibition activity and protein kinase activity. Primary structure of Bg7S reveals 12 conserved cysteine residues involved in forming 6 disulfide bonds, which provides appreciable stability to protein. Secondary structure is predominately rich in β-sheets with few alpha helices. Bg7S shares structural similarity with various aspartic-proteases. In this review our aim is to discuss sequence, structure and function of various globulins present in Glycine max.

Keywords: Soybean globulins, 11S and 7S globulin, β-Conglycinin, Basic 7S globulin, and Hypo-cholesterolemic effect

INTRODUCTION

Soybean *[Glycine max]* is one of the richest source of protein, which contain approximately 40% of dry weight (Krishnan and Nelson, 2011). Tombs et al. (Tombs, 1967) described that most of the proteins in soybean are present as storage proteins in the sub-cellular particles of cotyledons called 'protein bodies'. During the developmental stages of seed maturation in soybean, synthesis of large amount of storage proteins occurs (Asakura et al., 2012a). These proteins are used later during germination, where they are digested and amino acids are used for growth and development (Kim et al., 2011b). Proteins in soybean majorly consists of 4 fractions of globulins, which are classified as per their sedimentation properties: 2S, 7S, 11S and 15S fractions at 0.5M ionic strength, which account for 8%, 35%, 52% and 5% respectively (Kinsella, 1979). At 0.1M ionic strength, one type of 7S fraction aggregates to form a differently sedimenting complex having a sedimentation coefficient of 9S (KOSHIYAMA, 1968b).

About 70-80% of storage protein is constituted by 2 protein fractions: Glycinin and β-conglycinin (Catsimpoolas., 1969). The amount of these two proteins are inversely proportional to each other (Hill and Breidenbach, 1974). Glycinin has a sedimentation coefficient of 11S (Catsimpoolas, 1968; Tombs, 1967) and β-conglycinin has a sedimentation coefficient of 7S (Catsimpoolas and Ekenstam, 1969; Koshiyama, 1965; Koshiyama and Fukushima, 1976a). β-conglycinin is relatively low in sulphur containing amino acids than glycinin and thus is nutritionally low in score (Kaviani B, 2008). Glycinin and β-conglycinin correspond to the vicilin-like and legumin-like proteins respectively found in the legume family of plants

(Derbyshire et al., 1976). Both proteins have β-barrel domains and belong to the cupin superfamily, each having two cupin domains (Dunwell et al., 2000). One basic difference is that 7S fractions are glycoproteins containing mannose and glucosamine, with a ratio of 3:1 for mannose to N-acetyl glucosamine (Vu Huu and Kazuo, 1977).

Although β-conglycinin is the major protein in 7S globulin fractions of soybeans, two more fractions: basic 7S Globulin (Bg7S) and γ-conglycinin have been identified, which together account for only a small percentage of 7S globulins (Kinsella, 1979). γ-conglycinin has not been characterized very well, so is not discussed here in detail. The Bg7S is a glycoprotein with high content of cysteine residue. Very similar proteins have been found to be distributed widely amongst the legumes (Kagawa et al., 1987). Bg7S is released in good amount from the soy seeds by soaking in water at 50-60 °C (Hirano H, 1992). Classification of globulin proteins is showed in **Figure 1.**

In this paper we shall discuss all the characteristics associated with major storage proteins of soybean: 11S and 7S globulins, starting from the basic structure of globulins to specific structural features, from gene structure to regulation, the synthesis of proteins during seed development and its mobilization during germination. We also discuss here the physiological effects of dietary soy protein. Glycinin and β -conglycinin have been given a lot of emphasis earlier in reviews by other scientists. However, we emphasize on Bg7S here, in regards with the structural evaluation and comparison. We take into consideration here the primary, secondary and tertiary structure of basic 7S globulin and compare it with other related legume proteins.

STRUCTURAL FEATURES OF STORAGE PROTEINS

11S GLOBULIN

Glycinin has a molecular weight of around 320 kDa (Catsimpoolas and Ekenstam, 1969). The quaternary structure of glycinin consists of six monomers, where each subunit is composed of an acidic chain(A) and a basic chain(B) joined by disulfide linkage (Nicholas, 1969; Staswick et al., 1981). Catsimpoolas et al. (Catsimpoolas, 1968) also showed similar results, where purified 11S was separated into more than 12 bands on polyacrylamide disc gel electrophoresis using Guanidine-HCl as denaturant. Badley et al. (BADLEY, 1975) found 12 free N-terminals, indicating the presence of 12 chains. 6 different acidic polypeptides (A1a, A1b, A2, A3, A4, and A5, A6) and five basic (B1a,B1b, B2,B3, and B4) polypeptides were found having homology in their sequences (Staswick et al., 1981).

The 5 identified subunits of glycinin are divided into two groups on the basis of homology. Group 1: A1aB1b (53.6 kDa), A2B1a (52.4 kDa), A1bB2 (52.2 kDa) and Group2: A5A4B3 (61.2 kDa) and A3B4 (55.4 kDa) or also called glycinin G1, G2, G3 , G4 and G5 subunits respectively (Nielsen, 1985). Group 1 members are relatively rich in methionine and have 2 cysteine and 3 cysteine residues in basic and acidic residue's respectively, whereas group 2 members are relatively poor in methionine and have 2 cysteine and 2 cysteine residues basic and acidic residues respectively (Nielsen, 1985; Utsumi et al., 1997).

The cysteine residues which link the acidic and basic subunits are generally found close to N-terminal of the chains. Though the cysteine residues are present very close to each other on acidic chain, the residues of cysteine forming disulfide bond are highly specific (Staswick et al., 1984). Cys107 and Cys317, numbering from the start of precursor polypeptide synthesized, link

the acidic chain of G1 subunit (Adachi et al., 2001). Cys104 and Cys307 in G2 subunit (Staswick et al., 1984), Cys109 and Cys351 in G5 subunit (Adachi et al., 2003) are experimentally confirmed to form the inter-chain disulfide bonds. By sequence similarity, Cys107 and Cys303 in G3 subunit and Cys108 of A5 chain and Cys384 of B3 chain in G4 subunit are thought of forming the inter-chain disulfide bonds.

Glycinin subunits are arranged alternately into two identical hexagonal rings, stacked on one another, to form a hollow cylinder (BADLEY, 1975). At low ionic strength, particularly below or at 0.1M, a small fraction of glycinin associates solely or along with β-conglycinin reversibly, into heavier and faster sedimenting forms having a sedimentation coefficient of 15S (Catsimpoolas., 1969; Wolf and Nelsen, 1996). 11S dissociates variably depending on the denaturing conditions and different reagents used. At pH 2.2, 5S, 4S and 2S fractions are obtained at 0.2M, 0.1M and 0.01M ionic strength respectively (Wolf et al., 1958). 11S dissociates on heating to 100 °C, in 5mins small soluble aggregates and 3-4S fractions are the only left over and the whole 11s component disappears. On further heating, soluble aggregate precipitates out and only 3-4S fractions are left in solution (WOLF, 1969).

7S GLOBULIN

The β –conglycinin is major fraction of 7S globulin, exists as a hetero-trimer with molecular mass is about 180-210 kDa (Koshiyama, 1968a). The quaternary structure is comprised of 3 major subunits: α (68 kDa), α′ (72 kDa), and β (52 kDa), all of which have different physiochemical properties. α and α′ subunits are shown to have considerable similarity in the amino acid sequence. Also, β subunit was found to have relatively higher concentration of

hydrophobic amino acids (Vu Huu and Kazuo, 1977). The experimental isoelectric point for each subunit was reported as: α, 4.90; β, 5.66–6.00; α′, 5.18 (Vu Huu and Kazuo, 1977). Cysteine and metheonine were found to be absent in β subunit but one residue of cysteine is present in each α and α′ subunits (Nielsen, 1985). Molecular heterogeneity has been observed in β-conglycinin. Six different isomeric species are known: $\alpha'\beta_2$, $\alpha\beta_2$, $\alpha\alpha'\beta$, $\alpha_2\beta$, $\alpha_2\alpha'$ and α_3 (Fumio et al., 1981; Vu Huu and Shibasaki, 1978).

Each α and α′ subunits have 2 amino acids glycosylated with sugar moieties, while β subunit have only 1 glycosylated residue (Utsumi et al., 1997). These N-linked glycans help in folding the protein to their correct conformation (Chrispeels, 1991). All three subunits share a similar core region and α/α′ subunits contain extensions of 125 and 141 residues respectively, in addition to the core region (Maruyama et al., 1998). α and α′ subunits were found to be disulfide linked when isolated in oxidizing conditions (Wadahama et al., 2012).

The 7S fraction of globulins also comprises of two more proteins other than β–conglycinin: γ–conglycinin and basic 7S Globulin. γ–conglycinin was purified and was found in small amounts with preparations of β–conglycinin (Catsimpoolas and Ekenstam, 1969). Koshiyama & Fukushima (Koshiyama and Fukushima, 1976b) characterized the protein and found that its approximate molecular weight is 104 kDa and it had a sedimentation coefficient of 6.55S. Also the protein had an isoelectric pH of 5.80 and didn't dissociate and aggregate at low ionic concentrations. It is also a glycoprotein with 5.49% carbohydrate content.

Bg7S is a small cysteine rich protein that constitutes about 3% of total seed protein (Fumio et al., 1981). Hu and Esen (Hu and Esen, 1981, 1982) fractioned the soybean globulins, and reported

two small fractions of 27kDa and 16kDa. Yamauchi et al. (Yamauchi, 1984) reported that the molecular weight of Bg7S is 168kDa and its quaternary structure consists of 4 repeated units of a 43kDa protomer. The arrangement of protomers changes if pH is deviated from neutral. At weakly basic (pH 8.0) and weakly acidic (pH 6.0) conditions, it was found to occur as dimer, while it occurred as a monomer at highly acidic conditions (pH4.0). At neutral pH, the 4 protomers are arranged in pseudo-222 symmetry (Yoshizawa et al., 2011). This protomer is composed of a small subunit (16kDa) joined to a large subunit (27kDa) by means of a disulfide bond. It is so named, because of its relatively high isoelectric point (pH 9.05-9.26) compared to other globulins known (Sato et al., 1987; Yamauchi, 1984). Bg7S homologues are present widely in legume species (Kagawa et al., 1987).

GENE STRUCTURE AND REGULATION

11S GLOBULIN

Glycinin subunits are encoded by 5 genes, which are denoted Gy1 to Gy5 (Davies, 1987). Three non-allelic, highly homologous genes Gy1, Gy2, Gy3 were identified by Fisher and Goldberg (Fischer and Goldberg, 1982). Scallon et al. [43] identified the other two genes Gy4 and Gy5 two years later. Major glycinin subunits are coded by these genes which are divided into two groups on the basis of subunits they encode (Nielsen, 1984). Group I: Gy1, Gy2 and Gy3 genes code for A1aB2, A2B1a, A1b B1b or G1, G2, G3 subunits of glycinin and Group II: Gy4 and Gy5 genes code for A5A4B3 and A3B4 or G4 and G5 subunits (Beilinson et al., 2002;

Nielsen et al., 1989; Yagasaki K, 1996) respectively. There is about 45% similarity in the above two groups and about 80% similarity in the members of the group (Nielsen et al., 1989). In addition, two more genes are known to exist in the glycinin family: Gy6, a pseudo gene and a weakly expressing Gy7 gene, found to be expressed during mid-maturation stages of seed development (Beilinson et al., 2002; Nielsen et al., 1989). Recently, a pseudo gene, Gy8 has also been identified. These three genes are not similar to both the groups above and thus are identified separately as Group III. They have been derived by tandem duplication, followed by two rounds of gene replication on large scale of the ancestral glycinin genes (Li and Zhang, 2011). Linkage groups N, O, F and L are known to have Gly1-Gly7 genes. Gy1 and Gy2 genes are present 3kb apart in a single domain occurring in the linkage group N, while Gy3, Gy4 and Gy5 belongs to linkage group L,O and F respectively. Gy6 was present downstream of Gy2 on linkage group N and Gy7 was found to be present downstream of Gy3 on linkage group L (Beilinson et al., 2002; Chen and Shoemaker, 1998; Cho and Nielsen, 1989; Diers et al., 1994). Linkage group L is found to be on chromosome 19 (Garner, 2001) and linkage group N on chromosome 3 (Lee J, 2000).

The subunits are coded by 4 exons separated by 3 introns in all of the 5 major genes. Several conserved regions are known. Firstly, at 5’ end of the genes, a seed specific consensus sequence of 9bp 5'-CAACACAAT-3' (Geldberg, 1986) is present and 8bp 5'-CATGCATG-3' RY sequence (Dickinson, 1988) is found around 109bp upstream the transcription start site. Secondly, a legumin box is present around 89-90 bases upstream. Thirdly, 25 to 30bp upstream a TATA box was present (B.Goldberg, 1989; Cho and Nielsen, 1989; Vu H.Thanh, 1989). And 5'-AACAAUGGC-3' regulatory sequence is found at transcription start site (Lutcke, 1987) and

multiple repeats of 5'-AATAAA-3' at poly-adenylation sites (Fitzgerald, 1981). The legumin box of Gy6 and Gy7 has been found aberrant, thus may be the cause of pseudogenation and weak expression (Beilinson et al., 2002). Regulation of glycinin gene expression occurs mainly at mRNA level by the cis-regulatory elements described above (Goldberg, 1981a). The expression is suppressed after germination. Glycinin genes are found to be expressed progressively during seed maturation from a 2mm size bean to a fully mature sized bean (Asakura et al., 2012b).

7S GLOBULIN

β-conglycinin subunits are coded by a large family of genes with clustering of few genes (Harada JJ, 1989). Two homologous mRNA families that differ significantly in size, code for the β-conglycinin subunits (Goldberg, 1981a). Subunits α/α′ are coded by a 2.5kb mRNA, and subunit β is coded by a 1.7kb mRNA (Beachy et al., 1981; Schuler et al., 1982). A number of homologous mRNA constitutes the family of 2 differently sized mRNA's, which account for the heterogeneity in seeds along with post-translational mechanisms (Beachy, 1985; Beachy et al., 1981; Sengupta, 1981). As many as fifteen genes designated CG-1 to CG-15 have been identified. α subunit is coded by CG-2 and CG-3 (Yoshino M, 2002), α′ subunit is coded by CG-1 and β subunit is coded by CG-4. CG-8, CG-11, CG-12, CG-13, and CG-15 are found to be highly homologous to CG-4 and code for unique β subunits. CG-5, CG-6, CG-7, and CG-14 are found to be equally homologous with both mRNA families. When compared to CG-4, an extra 0.56kb protein coding region is found in the first exon of CG-1 gene which accounts for its large size (Harada JJ, 1989).

CG-2 and CG-3 have shown to be very tightly linked, which together are found to be tightly linked to CG-4 (Davies CS, 1985). CG-1 is found to segregate and inherit independently of CG-2 (Tsukada Y, 1986).CG-1 has been mapped along with Gy5 on linkage group F (Chen and Shoemaker, 1998). CG-1 expression is obtained even if contiguous 159bp are present to the transcription start site (Chen, 1988, 1986). However, for high expression, -159 to -257 upstream regions are required. The seed specific consensus elements 5'-CATGCAT-3' and 5'-CAACACA-3' are present. TATA box is present around 31 bases upstream the transcription start site (Harada JJ, 1989). The most characterized genes CG-1 to CG-4 are found to have five introns positioned in between six exons having high homology in the sequences (Doyle, 1986; Harada JJ, 1989). Regulation of β-conglycinin genes occurs mainly at transcriptional level by the cis-regulatory elements described above. α and α′ subunits are expressed in the early stage in seed development. On the other hand, the mRNA for β-subunit appears at relatively later stage along with glycinin genes during seed maturation (Asakura et al., 2012b; Sengupta, 1981). The mRNA's are not detected in the germinating seedling cotyledons (Walling, 1986).

Bg7S gene is 1281bp long and doesn't contain introns. Transcription initiates about 61 bases upstream the start codon. TATA box is present 25bases upstream and CAAT box is found 116 bases upstream. AATAAA poly-adenylation signals are present at two places i.e. 74 and 94bp downstream the stop site TGA. Also, CTTCTGGAAATTTCA 153 bp upstream, CTACTTGAAACTTAT 706 bp upstream, and CTTCTAAAATCTTCC 932 bp upstream are found, which are very similar to Heat-shock element sequences. Bg gene is present at least in 4 copies in the total genome (Watanabe and Hirano, 1994). The regulation of expression of gene

occurs at transcriptional level and a gradual increase in the amount of expression of gene is found during seed development (Asakura et al., 2012b).

SITE OF EXPRESSION: SYNTHESIS AND MOBILIZATION

Synthesis of prepro-glycinin takes place on the rough endoplasmic reticulum (ER) of parenchyma cells in the cotyledons (Chrispeels et al., 1982). Each protomer is synthesised as a single polypeptide chain consisting of a signal peptide, an acidic subunit and a basic subunit joined with each other requiring two proteolytic cleavages for maturation (Sengupta et al., 1981; Tumer et al., 1981). Signal peptide which targets the protomer into the lumen of ER is removed co-translationally to form pro-glycinin (Tumer et al., 1981). Non-covalent interactions between the pro-glycinin subunits lead to formation of trimer, following the formation of intra-molecular disulfide bonds between the acidic and the basic subunits, and are transported in Golgi (Ereken-Tumer N, 1982; Nam Y-W, 1997). Pro-glycinin subunits are transferred to dense vesicles and a sorting mechanism then transports them to a large central vacuole (Dorel et al., 1989). Second cleavage takes place in the central vacuole between an evolutionary conserved aspargine-glycine bond to form separate acidic and basic chains in each subunits (Hara-Nishimura I, 1987). Acidic chains move away from the association site due to a change in their hypervariable region causing association to form hexamer (Adachi et al., 2003; Dickinson CD, 1989).

β-conglycinin's quaternary structure is a trimer of 3 different subunits. Synthesis of each α, α' and β subunit takes place on the rough ER in the cotyledons, as a polypeptide consisting of a pro-peptide joined with a signal peptide (Sengupta et al., 1981). The removal of signal peptide occurs co-translationally. Inside the ER lumen, glycans or carbohydrate moieties are attached on the

nitrogen of aspargine, one on β-subunit and two each on α and α' subunits (Sengupta et al., 1981). Non-covalent interactions between the pro-conglycinin subunits lead to formation of a trimer (Chrispeels et al., 1982), which is transported to Golgi and then to dense transport vesicles for their deposition in protein storage bodies or protein bodies. After their assembly into trimers and their transport, the pro-peptides are cleaved to yield a mature form of β-conglycinin (Lelievre J-M, 1992).

Bg7S is synthesized on the rough ER as a single polypeptide consisting of 24 amino acid long signal peptide, a large subunit chain and a small subunit chain. The signal poly-peptide is co- and post-translationally cleaved and the polypeptide is transferred into the ER, where disulfide bonds are formed and it is folded in proper conformation (Watanabe and Hirano, 1994). The protein is transferred via Golgi vesicles to middle lamella in the cell wall of cells in the cotyledons. It is known to be facing the inter-cellular space while residing in the middle lamella (Nishizawa et al., 1994).

MOBILIZATION

Mobilization or breakdown of seed storage proteins during germination initiates with limited proteolysis followed by extensive degradation. α and α' subunits of β-conglycinin and glycinin acidic subunits are the first to be mobilized early during seed germination (Wilson et al., 1986). The first step in degradation of α and α' subunits of β-conglycinin is catalyzed by Protease C1, so named. This endo-protease cleaves at regions rich in glutamic acid (Qi X, 1994; Qi X, 1992), and at serine residues. It belongs to subtilisin family (Liu X, 2001), and is localized in plant vacuole. Its N-terminal sequence is responsible for targeting to endo-membrane system

(Fanglian He, 2007). N-terminal regions of α and α' subunits, non-homologous to β subunit are targeted by this enzyme and consequently, 48 kDa and 50 kDa proteins corresponding to core regions of α and α' subunits appear (Qi X, 1992). Cupin domains in core regions of all three subunits are separated later during germination by another protease. β subunit, 48kDa and 50kDa intermediates are further degraded into smaller units of approximately half their size (Wilson et al., 1986). Lys-Thr and Thr-Ile bonds in the linker region of cupin domains represent easily accessible sites for initiating cleavage. A papain-like cysteine protease, protease C2, catalyzes such cleavage. Seo et al. (Seo et al., 2001) stated that protease C2 is actually starting protease that initiates the process of mobilization and doesn't require prior proteolysis by Protease C1 and also is able to act on the accessible N-terminal regions of α and α' subunits.

Glycinin proteolysis also follows a similar pattern. Protease C1 acts on C-terminal of acidic subunits of glycinin at a region rich in acidic amino acids and reduces their size by about 1-2 kDa (Qi X, 1994). The acidic chains are then cleaved reducing their size from 40 kDa to 21 kDa proteolytic intermediates, still disulfide bonded with the basic subunits (Wilson et al., 1986). Basic chains contain more hydrophobic regions and are more compact, thus less prone to proteolysis due to the structural advantage, thus are processed relatively later during germination (Kim et al., 2011a).

FUNCTION

11S globulin fraction is found purely for storage purpose, and has no physiological activity. Major fraction of Bg7S and β-conglycinin is also found for storage purpose. Both of these proteins are synthesized in large amounts during seed development and are utilized during

germination where the proteolysis products are used for growth (Meinke et al., 1981; Tombs, 1967). However, Bg7S has been found to be associated with biological activity. Computational analysis shows an aspartic-type endo-peptidase activity domain in addition to nutrient reservoir activity. It is highly similar in structure to inhibitors of endo- β-glucanase present in other legume species, like xyloglucan-specific endo-β-1, 4-glucanase inhibitor protein (XEGIP) in tomato and *Tritinum aestivum* xylanase inhibitor (TAXI) in wheat, which inhibit enzymes of glycoside hydroxylase (GH) family . But in contrast, Soybean Bg7S is unable to inhibit GH12 or GH11 enzymes but may possibly have similar role in alteration of activity of other related enzymes (Yoshizawa et al., 2011).

Bg7S is capable of binding to hormone-like peptide leginsulin and insulin-like growth factors IGF-I and IGF-II (Hirano H, 1992; Komatsu and Hirano, 1991). Comparison to human insulin receptor shows a good structural similarity between the two, even though no significant sequence homology is observed. This, along with the fact that Bg7S is expressed in the middle lamella, proposes a possible role of Bg7S as insulin receptor-like function in signal transduction and physiological growth control (Komatsu and Hirano, 1991; Watanabe and Hirano, 1994). In addition to this, Komatsu et al. (Komatsu et al., 1994) showed that the low molecular mass 16 kDa subunit has protein kinase associated with it. They also showed that insulin binds to both, sugar moiety of 27 kDa subunit and 16 kDa subunit separately.

BASIC 7S GLOBULIN

PRIMARY STRUCTURE

Bg7S is synthesized as single polypeptide chain of 427 amino acids from a large open reading frame of 1281bp. It is composed of a signal peptide, a low molecular weight subunit chain and a high molecular weight subunit chain. Signal peptide, 24 amino acids long, is cleaved while its transport into the lumen of ER. Amino acid residues ranging from 25 to 275 correspond to 27 kDa or high molecular weight subunit (α) and amino acids from 276 to 427 correspond to 16 kDa or low molecular weight subunit (β) (Hirano H., 1990; Kagawa and Hirano, 1989). Cleavage occurs between two serine residues 275 and 276 to generate two subunits, numbered starting from the first residue of precursor polypeptide chain. The disulfide bond formed between Cys209–Cys418 links the subunits and is critical for stabilization (Yoshizawa et al., 2011). Initially partial amino acid sequencing was performed to sequence N-terminal residues once and internal residues of Bg7S (Hirano et al., 1986; Kagawa et al., 1987). Later, full sequence was reported by cloning and identification of cDNA encoding Bg7S (Kagawa and Hirano, 1989). Primary structure analysis of Bg7S reveals aspartic type endo-peptidase homology (Source: InterPro), however, recently Bg7S was found to lack protease activity, as a catalytic aspartate residue in the active site of Bg7S was replaced by Ser289 (Yoshizawa et al., 2011).

Potential evolutionary relatives of Bg7S were obtained by BLAST (Stephen F. Altschul, 1997) and 5 sequences with good resemblance were taken into account for analysis. Sequence identity of Bg7S with close relatives is: (63%) Bg7S *[Medicago truncatula]*, (62%) conglutin gamma *[Lupinus albus]*, (40%) xyloglucan-specific endoglucanase inhibitor protein *[Capsicum annuum]*, (39%) xyloglucan-specific fungal endoglucanase inhibitor protein *[Solanum lycopersicum]* and (39%) aspartyl protease-like protein [Arabidopsis thaliana]. *Triticum astevium* xylanase inhibitor (TAXI-IA) is found to have homologous structure to Bg7S but is not taken

into account for this alignment due to very low sequence similarity. Multiple sequence alignment of these sequences as shown in **Figure 2** reveals 12 conserved cysteine residues. Interestingly, 6 disulfide bonds are found in the crystal structure of Bg7S, formed by these 12 conserved residues, thus it can be inferred that these disulfide bond positions are relatively conserved. Along with this, the fact that high degree of sequence homology is there, similar folding patterns might be adopted by all referred proteins. The disulfide bonds are formed between Cys51–Cys141, Cys65–Cys78, Cys70–Cys94, Cys81–Cys89, Cys185–Cys394 (Inter-subunit) and Cys304–Cys345, numbered starting from the mature polypeptide (Yoshizawa et al., 2011). Also aspartate residue of catalytic site is conserved at one position and is replaced by serine at another in 4 out of 6 sequences taken into account, which explains the loss of peptidase activity.

The analysis using Prosite (Hofmann et al., 1999) showed 4 potential protein kinase C phophorylation sites, 1 caesin kinase II phosphorylation site and 8 N-myristoylation sites distributed well on primary structure of Bg7S. No potential N-linked glycosylation site was observed in the sequence. However, Bg7S is a glycoprotein, suggesting the existence of O-linked glycosylation. The analysis of the sequence for search of motifs reveals hits in 8 different families dispersed throughout the sequence of Bg7S. These have been illustrated in the **Table 1**. Two signatures of each of the following families were observed: PRSTNOIDEP1R having an ancestary from 7TM, PEPSIN, NGELATINASE, RIBOSOMALS2, OTCASE, TGF beta 2, HOLDHDRGNASE and FLGPRINGFLGI.

Phylogenetic analysis of Bg7S was performed with related proteins found by BLAST search. Sequences from different plant species like *Solanum, Capsicum, Daucus* and many others have been taken to prepare a phylogenetic tree, shown in **Figure 3**. The analysis showed the division of proteins in two major clusters. Cluster I consists of Basic 7S globulin from *Glycine max* or soybean along with Bg7S from *Medicago truncatula* and Conglutin gamma from 2 Lupinus species. Pepsin A *[Ricinus communis]* though belongs cluster I but has relatively low homology to other members. Presence of pepsin A in cluster I points to a potential aspartyl- protease activity domain in other members. However, during lineation, these proteins lost the protease activity. Cluster II consists of XEGIP's and EDGP from various plant species. Interestingly, ATP binding protein *[Ricinus communis]* belongs to this cluster pointing to possible ATP binding activity of these proteins.

SECONDARY AND TERTIARY STRUCTURE

We analyzed the structure of Bg7S, which consists of 19 β-strands and 9 α-helices, which suggests that Bg7S is a β-rich protein. Structural classification of proteins or SCOP (Lo Conte, 2000) analysis of Bg7S sequence suggests that it belongs to class of all-β proteins, with a fold of acid proteases and family pepsin-like acid protease. The structural topology of monomer of Bg7S is illustrated in **Figure 4** (Yoshizawa et al., 2011). The cleft between α domain and β domain serves as the pseudo-active site of Bg7S. The cleft is homologous to cleft in aspartic protease, and should contain two aspartic residues; however Asp265 in mature polypeptide is replaced by

serine, while Asp374 of mature polypeptide remain conserved. This explains the loss of protease activity in Bg7S.

Recently, crystal structure of Bg7S from soybean has been reported (Yoshizawa et al., 2011). We tried to find out homologous structures in protein data bank by DALI server (P, 2010), the results of which are showed in the **Table 2**. The structure of xylanase inhibitor (PDB id-1t6g) was found to be very similar in structure to Bg7S (Z score- 39.7 and RMSD- 2.3). The secondary structure elements are well conserved between these two proteins, despite of some deletions and insertions in loop.

HYPO-CHOLESTEROLEMIC EFFECTS OF SOYBEAN PROTEIN

The control or reduction of lipidemic levels of high cholesterolemic subjects by dietary soybean proteins is widely accepted (C.R. Sirtori, 2001). Major fraction of 7S globulin, β-conglycinin have been in highlight in recent times due to its ability to up-regulate high affinity low density lipoprotein (LDL)-receptors in liver (M.R. Lovati, 1992). Also its effect in reduction of triglycerides in human plasma is of importance (Ferreira et al., 2012; H. Kambara, 2002). Also, the same has been seen in hyper-lipidemic rats (T. Aoyama, 2001). This up-regulation was induced mainly by α' subunit and relatively less by α subunit, while β subunit is ineffective. It is likely that 7S globulin in dissociated into constituent subunits or is degraded in cytosol of Hep G2 cells, taken as model system (Manzoni et al., 2003). Recently, a 142 amino acid long recombinant polypeptide chain of N-terminus of α' subunit was synthesised, consisting of only the extension region. This peptide was found to modulate the uptake(+87%) and degradation(+100%) of LDL, at a concentration of 2μM(Consonni et al., 2011).

A small peptide FVVNATSN, prepared by hydrolysis of β chain of β-conglycinin by a protease obtained from *Bacillus amyloliquefaciens*, was found to up-regulate transcription of LDL receptors. About +148% expression was obtained by exposure of this peptide at 100μM conc. using Hep G2 cells (S.J. Cho, 2008). Similar peptides derived from 11S globulin are also known to control the lipidemic levels(M.R. Lovati, 1987). Recently Fassini et al. (Priscila G Fassini, 2011) confirmed that 11S globulin reduce artherogenic factors like total cholesterol and triglycerides in animals exposed to a hyper-cholesterolemic diet.

β-Conglycinin controls levels of cholesterol by down-regulation of two proteins: SREBP1 and adinopectin. Sterol regulatory element-binding protein 1 (SREBP1), the major protein controlling fatty acid synthesis, is down-regulated by feeding β-Conglycinin. Adinopectin is a protein involved in glucose metabolism (Tachibana et al., 2010). In addition to this, down-regulation of mRNA's of ChREBP (carbohydrate response element-binding protein) in mouse fed with sucrose and of peroxisome proliferator-activated receptor (PPAR) γ2 in mouse fed with High fat diet, both of which are involved in fatty liver disease in mouse (Yamazaki et al., 2012).

PERSPECTIVES

Here we discussed most of the recent issues related to major globulin proteins protein soybean. The gene sequence and its regulation are described in detail for both 11S globulin and Bg7S. The Bg7S is a cysteine-rich glycoprotein present as a storage protein in soybean seeds. Mature Bg7S is composed of subunits of high and low molecular mass that are linked by

disulfide bonds. The crystal structure of Bg7S has reported recently which has provided significant clues for understanding the physiological role of Bg7S. The completion of the soybean genome sequence allowed us to identify conserved motifs and facilitating the identification of compartment-specific cis-regulatory sequences that connect seed globulin genes into regulatory networks. Although, major function of globulin proteins are storage, however, recent findings suggests some important role and therefore, these proteins may have many significant future application. We hope this concise review provided all essential information about major globulin proteins present on soybean with special reference to Bg7S.

ACKNOWLEDGMENTS

MIH is grateful to the University Grants Commission for financial assistance.

REFERENCES

Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura, K., Mikami, B., and Utsumi, S. (2003). Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. Proceedings of the National Academy of Sciences of the United States of America *100*, 7395-7400.

Adachi, M., Takenaka, Y., Gidamis, A.B., Mikami, B., and Utsumi, S. (2001). Crystal Structure of Soybean Proglycinin A1aB1b Homotrimer. Journal of molecular biology *305*, 291-305.

Asakura, T., Tamura, T., Terauchi, K., Narikawa, T., Yagasaki, K., Ishimaru, Y., and Abe, K. (2012a). Global gene expression profiles in developing soybean seeds. Plant Physiol Biochem *52*, 147-153.

Asakura, T., Tamura, T., Terauchi, K., Narikawa, T., Yagasaki, K., Ishimaru, Y., and Abe, K. (2012b). Global gene expression profiles in developing soybean seeds. Plant Physiology and Biochemistry *52*, 147-153.

B.Goldberg, T.L.S.a.R. (1989). The glycinin Gy1 gene from soybean. Nucleic acids research *17*, 1.

BADLEY, R.A., ATKINSON, D., HAUSER, H., OLDA."JI, D., GREEN, J. P., and STUBBS, J. M. .. ,214-228. (1975). The structure, physical and chemical properties of the soybean protein glycinin. Biochim Biophys Acta *412*, 17.

Beachy, R.N., Chen, 2. L., Horsch, R. B., Rogers, S. G., Hoffman, N. J., and Fraley, R. T. (1985). Accumulation and assembly of soybean β-conglycinin in seeds of transformed petunia plants. EMBO J *4*, 7.

Beachy, R.N., Jarvis, N.P., and Barton, K.A. (1981). Biosynthesis of subunits of the soybean 7S storage protein. J Mol Appl Genet *1*, 19-27.

Beilinson, V., Chen, Z., Shoemaker, C., Fischer, L., Goldberg, B., and Nielsen, C. (2002). Genomic organization of glycinin genes in soybean. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *104*, 1132-1140.

C.R. Sirtori, M.R.L. (2001). Soy protein and cardiovascular disease. Curr Atheroscler Rep *3*, 7.

Catsimpoolas, N. (1968). Micro isoelectric focusing in polyacrylamide gel columns. Anal Biochem *26*, 480-482.

Catsimpoolas, N., and Ekenstam, C. (1969). Isolation of alpha, beta, and gamma conglycinins. Archives of biochemistry and biophysics *129*, 490-497.

Catsimpoolas., N. (1969). A Note on the Proposal of an Immunochemical System of Reference and Nomenclature for the Major Soybean Globulins. Cereal Chem *46*, 3.

Chen, A., and Shoemaker, R. (1998). Four genes affecting seed traits in soybeans map to linkage group F. Journal of Heredity *89*, 211-215.

Chen, L., Pan, N. S., and Beachy, R. N. (1988). A DNA sequence element that confers seed-specific enhancement of a constitutive promoter. EMB0 J *7*, 6.

Chen, L., Schuler, M. A., and Beachy, R. N. (1986). Functional analysis of regulatory elements in a plant embryo-specific gene. Proc Natl Acad Sci USA *83*, 5.

Cho, T.J., and Nielsen, N.C. (1989). The glycinin Gy3 gene from soybean. Nucleic acids research *17*, 4388.

Chrispeels, M., Higgins, T., and Spencer, D. (1982). Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. The Journal of Cell Biology *93*, 306-313.

Chrispeels, M.J. (1991). Sorting of Proteins in The Secretory System. Annual Review of Plant Physiology and Plant Molecular Biology *42*, 21-53.

Consonni, A., Lovati, M.R., Parolari, A., Manzoni, C., Morazzoni, P., Magni, C., and Duranti, M. (2011). Heterologous expression and purification of the soybean 7S globulin alpha' subunit extension region: in vitro evidence of its involvement in cell cholesterol homeostasis. Protein expression and purification *80*, 125-129.

Davies, C.S., Cho, T.-J., and Nielsen, N.C. (1987). Tests for linkage of seed protein genes Lxl, Lx3, and Cgyl to markers on known linkage groups. Soybean Genet Newsletter *14*, 5.

Davies CS, C.J., Nielsen NC (1985). Inheritance and biochemical analysis of four electrophoretic variants of β-conglycinin from soybean. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *71*, 8.

Derbyshire, E., Wright, D.J., and Boulter, D. (1976). Legumin and vicilin, storage proteins of legume seeds. Phytochemistry *15*, 3-24.

Dickinson, C.D., Evans, R.P., and Nielsen, N.C. (1988). RYrepeats are conserved in the 5'-flanking regions of legume seed-protein genes. Nucl Acids Res *16*.

Dickinson CD, H.E., Nielsen NC (1989). Role of posttranslational cleavage in glycinin assembly. The Plant cell *1*, 11.

Diers, B.W., Beilinson, V., Nielsen, N.C., and Shoemaker, R.C. (1994). Genetic mapping of the <i>Gy4 and <i>Gy5 glycinin genes in soybean and the analysis of a variant of <i>Gy4</i&gt. TAG Theoretical and Applied Genetics *89*, 297-304.

Dorel, C., Voelker, T.A., Herman, E.M., and Chrispeels, M.J. (1989). Transport of proteins to the plant vacuole is not by bulk flow through the secretory system, and requires positive sorting information. The Journal of Cell Biology *108*, 327-337.

Doyle, J.J., Schuler, M. A., Godette, W. D., Zenger, V., Beachy, R. N., and Slightom, J. L. (1986). The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris: Structural homologies of genes and proteins. J Biol Chem *261*, 11.

Dunwell, J.M., Khuri, S., and Gane, P.J. (2000). Microbial Relatives of the Seed Storage Proteins of Higher Plants: Conservation of Structure and Diversification of Function during Evolution of the Cupin Superfamily. Microbiology and Molecular Biology Reviews *64*, 153-179.

Ereken-Tumer N, R.J., Nielsen NC (1982). Structural characterization of the glycinin precursor. The Journal of biological chemistry *257*, 3.

Fanglian He, F.H., Karl A. Wilson and Anna Tan-Wilson. (2007). Protein storage vacuole acidification as a control of storage protein mobilization in soybeans. Journal of experimental botany *58*, 12.

Ferreira, E., Silva, M., Demonte, A., and Neves, V. (2012). beta-conglycinin combined with fenofibrate or rosuvastatin have exerted distinct hypocholesterolemic effects in rats. Lipids in health and disease *11*, 11.

Fischer, R.L., and Goldberg, R.B. (1982). Structure and flanking regions of soybean seed protein genes. Cell *29*, 651-660.

Fitzgerald, M., and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell *24*, 10.

Fumio, Y., Masaaki, S., Wataru, S., Yoshiro, K., and Kazuo, S. (1981). Isolation and Identification of a New Type of β-Conglycinin in Soybean Globulins. Agricultural and Biological Chemistry *45*, 2863-2868 %U http://ci.nii.ac.jp/naid/130000028688/en/.

Garner, M.E., Hymowitz, T. Xu, S.J. Hartman, G.L. USDA, ARS (2001). Physical map location of the Rps1-k allele in soybean. Crop science *41(5)*, 4.

Geldberg, R.B. (1986). Regulation of plant gene expression. Philos Trans R Soc Lond B Biol Sci *314*, 11.

Goldberg, R.B., Hoscheck, G., Ditta, G.S., and Breidenbach, R.W. (1981a). Developmental regulation of clone superabundant embryo mRNAs in soybean. Dev Biol *83*, 14.

H. Kambara, H.M., K. Takamatsu, M. Kito (2002). Triglyceride-lowering effect of soybean b-conglycinin in humans. Ther Res *23*, 5.

Hara-Nishimura I, N.M. (1987). Proglobulin processing enzymes in vacuoles isolated from developing pumpkin cotyledons. Plant physiology *85*, 6.

Harada JJ, B.S., Goldberg RB (1989). Soybean b-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. The Plant cell *1*, 12.

Hill, J.E., and Breidenbach, R.W. (1974). Proteins of Soybean Seeds: I. Isolation and Characterization of the Major Components. Plant physiology *53*, 742-746.

Hirano, H., Kagawa, H., Kamata, Y., and Yamauchi, F. (1986). Structural homology among the major 7s globulin subunits of soybean seed storage proteins. Phytochemistry *26*, 41-45.

Hirano H, K.H., Okubo K (1992). Characterization of proteins released from legume seeds in hot water. Phytochemistry *31*, 5.

Hirano H., W.T. (1990). Microsequencing of proteins electrotransferred onto immobilizing matrices from polyacrylamide gel electrophoresis: application to an insoluble protein. Electrophoresis *11*, 8.

Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999). The PROSITE database, its status in 1999. Nucleic acids research *27*, 215-219.

Hu, B., and Esen, A. (1981). Heterogeneity of soybean seed proteins: one-dimensional electrophoretic profiles of six different solubility fractions. Journal of agricultural and food chemistry *29*, 497-501.

Hu, B., and Esen, A. (1982). Heterogeneity of soybean proteins: two-dimensional electrophoretic maps of three solubility fractions. Journal of agricultural and food chemistry *30*, 21-25.

Kagawa, H., and Hirano, H. (1989). Sequence of a cDNA encoding soybean basic 7S globulin. Nucleic acids research *17*, 8868.

Kagawa, H., Yamauchi, F., and Hirano, H. (1987). Soybean basic 7 S globulin represents a protein widely distributed in legume species. FEBS letters *226*, 145-149.

Kaviani B, K.A. (2008). Improvement of the nutritional value of soybean [Glycine max (L.) Merr.] seed with alteration in protein subunit of glycinin (11S globulin) and beta-conglycinin (7S globulin). Turk J Biol *32*, 7.

Kim, H.T., Choi, U.-K., Ryu, H.S., Lee, S.J., and Kwon, O.-S. (2011a). Mobilization of storage proteins in soybean seed (Glycine max L.) during germination and seedling growth. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics *1814*, 1178-1187.

Kim, H.T., Choi, U.K., Ryu, H.S., Lee, S.J., and Kwon, O.S. (2011b). Mobilization of storage proteins in soybean seed (Glycine max L.) during germination and seedling growth. Biochim Biophys Acta *1814*, 1178-1187.

Kinsella, J. (1979). Functional properties of soy proteins. Journal of the American Oil Chemists' Society *56*, 242-258.

Komatsu, S., and Hirano, H. (1991). Plant basic 7 S globulin-like proteins have insulin and insulin-like growth factor binding activity. FEBS letters *294*, 210-212.

Komatsu, S., Koshio, O., and Hirano, H. (1994). Protein Kinase Activity and Insulin-binding Activity in Plant Basic 7S Globulin. Bioscience, biotechnology, and biochemistry *58*, 1705-1706 %U http://ci.nii.ac.jp/naid/110002677179/en/.

Koshiyama, I. (1965). Purification of the 7S component of soybean proteins. . Agric Biol Chem *29* 3.

Koshiyama, I. (1968a). Chemical and physical properties of a 7S protein in soybean globulins. Cereal Chem, 10.

KOSHIYAMA, I. (1968b). Factors influencing conformation changes in a 7S protein of soybean globulins by ultracentrifugal investigations. Agric BioI Chern (Tokyo) 9.

Koshiyama, I., and Fukushima, D. (1976a). Identification of the 7S globulin with β-conglycinin in soybean seeds. Phytochemistry *15*, 157-159.

Koshiyama, I., and Fukushima, D. (1976b). Purification and some properties of γ-conglycinin in soybean seeds. Phytochemistry *15*, 161-164.

Krishnan, H.B., and Nelson, R.L. (2011). Proteomic analysis of high protein soybean (Glycine max) accessions demonstrates the contribution of novel glycinin subunits. Journal of agricultural and food chemistry *59*, 2432-2439.

Lee J, Z.J., Xu S, Kollipara KP, Cregan PB, Singh RJ, Hymowitz T (2000). Development of the universal soybean map. 8th Biennial Conf Cellular and Mol Biol of the Soybean.

Lelievre J-M, D.C., Dickinson LA, Nielsen NC (1992). Synthesis and assembly of soybean β-conglycinin in vitro. Plant Mol Biol *18*, 16.

Li, C., and Zhang, Y.M. (2011). Molecular evolution of glycinin and beta-conglycinin gene families in soybean (Glycine max L. Merr.). Heredity *106*, 633-641.

Liu X, Z.Z., Barnaby N, Wilson KA, Tan-Wilson A. (2001). Soybean subtilisin-like protease involved in initiating storage protein degradation. Seed Science Research *11*, 14.

Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G., and Chothia, C. (2000). SCOP: a structural classification of proteins database. . Nucleic Acids Res *28*, 3.

Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Scheele, G.A. (1987). Selection of AUG differs in plants and animals. EMBO J *6*, 6.

M.R. Lovati, C.M., A. Canavesi, M. Sirtori, V. Vaccarini, M. Marchi, G. Gaddi, C.R. Sirtori (1987). Soybean protein diet increases low-density lipoprotein receptor activity in mononuclear cells from hypercholesterolemic patients. J Clin Invest *80*, 5.

M.R. Lovati, C.M., A. Corsini, A. Granata, R. Frattini, R. Fumagalli, C.R. Sirtori (1992). Low-density receptor activity is modulated by soybean globulins in cell culture. J Nutr *122*, 8.

Manzoni, C., Duranti, M., Eberini, I., Scharnag, H., Marz, W., Castiglioni, S., and Lovati, M.R. (2003). Subcellular localization of soybean 7S globulin in HepG2 cells and LDL receptor up-regulation by its alpha' constituent subunit. The Journal of nutrition *133*, 2149-2155.

Maruyama, N., Katsube, T., Wada, Y., Oh, M.H., Barba De La Rosa, A.P., Okuda, E., Nakagawa, S., and Utsumi, S. (1998). The roles of the N-linked glycans and extension regions of soybean beta-conglycinin in folding, assembly and structural features. European journal of biochemistry / FEBS *258*, 854-862.

Meinke, D.W., Chen, J., and Beachy, R.N. (1981). Expression of storage-protein genes during soybean seed development. Planta *153*, 130-139.

Nam Y-W, J.R., Nielsen NC (1997). Adenosine 5'-triphosphate is required for the assembly of 11S seed proglobulins in vitro. Plant physiology *115*, 11.

Nicholas, C. (1969). Isolation of glycinin subunits by isoelectric focusing in urea-mercaptoethanol. FEBS letters *4*, 259-261.

Nielsen, N. (1985). The structure and complexity of the 11S polypeptides in soybeans. Journal of the American Oil Chemists' Society *62*, 1680-1686.

Nielsen, N.C. (1984). The Chemistry of Legume Storage Proteins. Philosophical Transactions of the Royal Society of London B, Biological Sciences *304*, 287-296.

Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallon, B.J., Fischer, R.L., Sims, T.L., Drews, G.N., and Goldberg, R.B. (1989). Characterization of the glycinin gene family in soybean. The Plant cell *1*, 313-328.

Nishizawa, N.K., Mori, S., Watanabe, Y., and Hirano, H. (1994). Ultrastructural Localization of the Basic 7S Globulin in Soybean (Glycine max) Cotyledons. Plant and Cell Physiology *35*, 1079-1085.

P, H.L.R. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res *38*, 5.

Priscila G Fassini, R.W.N., Ederlan S Ferreira, Maraiza A Silva, Valdir A Neves and Aureluce Demonte (2011). Soybean glycinin improves HDL-C and suppresses the effects of rosuvastatin on hypercholesterolemic rats. Lipids in health and disease *10*, 7.

Qi X, C.R., Wilson KA, Tan-Wilson AL. (1994). Characterization of a soybean b-conglycinin-degrading protease cleavage site. Plant physiology *104*, 9.

Qi X, W.K., Tan-Wilson AL. (1992). Characterization of the major protease involved in the soybean beta-conglycinin storage protein mobilization. Plant physiology *99*, 9.

S.J. Cho, M.A.J., C.H. Lee (2008). Identification of LDL-receptor transcription stimulating peptides from soybean hydrolysate in human hepatocytes. J Agric Food Chem *56*, 5.

Sato, K., Yamagishi, T., Kamata, Y., and Yamauchi, F. (1987). Subunit structure and immunological properties of a basic 7s globulin from soybean seeds. Phytochemistry *26*, 903-908.

Schuler, M.A., Ladin, B.F., Pollaco, J.C., Freyer, G., and Beachy, R.N. (1982). Structural sequences are conserved in the genes coding for the α, α′ and β-subunits of the soybean 7S seed storage protein. Nucleic acids research *10*, 8245-8261.

Sengupta, C., Deluca, V., Bailey, D.S., and Verma, D.P.S. (1981). Post-translational processing of 7S and 11S components of soybean storage proteins. Plant molecular biology *1*, 19-34.

Sengupta, C., Deluca, V., Bailey, D. S., and Verma, D. P. S. (1981). Posttranscriptional processing of 7s and 11 S components of soybean storage proteins. Plant MoI Biol *1*, 16.

Seo, S., Tan-Wilson, A., and Wilson, K.A. (2001). Protease C2, a cysteine endopeptidase involved in the continuing mobilization of soybean beta-conglycinin seed proteins. Biochimica et biophysica acta *1545*, 192-206.

Staswick, P.E., Hermodson, M.A., and Nielsen, N.C. (1981). Identification of the acidic and basic subunit complexes of glycinin. Journal of Biological Chemistry *256*, 8752-8755.

Staswick, P.E., Hermodson, M.A., and Nielsen, N.C. (1984). Identification of the cystines which link the acidic and basic components of the glycinin subunits. The Journal of biological chemistry *259*, 13431-13435.

Stephen F. Altschul, T.L.M., Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 14.

T. Aoyama, M.K., T. Saito, K. Fukui, K. Takamatsu, T. Yamamoto, Y. Hashimoto, M. Hirotsuka, M. Kito (2001). Reduction by phytate-reduced soybean b-conglycinin of plasma triglycerides level of young and adult rats. Biosci Biotechnol Biochem *65*, 5.

Tachibana, N., Iwaoka, Y., Hirotsuka, M., Horio, F., and Kohno, M. (2010). β-Conglycinin Lowers Very-Low-Density Lipoprotein-Triglyceride Levels by Increasing Adiponectin and Insulin Sensitivity in Rats. Bioscience, biotechnology, and biochemistry *74*, 1250-1255.

Tombs, M.P. (1967). Protein Bodies of the Soybean. Plant physiology *42*, 797-813.

Tsukada Y, K.K., Harada K, Kaizumu N (1986). Genetic analysis of subunits of two major storage protein (β-conglycinin and glycinin) in soybean seeds. Jpn J Breed *36*, 11.

Tumer, N.E., Thanh, V.H., and Nielsen, N.C. (1981). Purification and characterization of mRNA from soybean seeds. Identification of glycinin and beta-conglycinin precursors. The Journal of biological chemistry *256*, 8756-8760.

Utsumi, S., Matsumura, Y., Mori, T., Damodaran, S., and Paraf, A. (1997). Food Proteins and Their Applications.

Vu H.Thanh, N.E.T.a.N.C.N. (1989). The glycinin Gy2 gene from soybean. Nucleic acids research *17*, 1.

Vu Huu, T., and Kazuo, S. (1977). Beta-conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. Biochimica et Biophysica Acta (BBA) - Protein Structure *490*, 370-384.

Vu Huu, T., and Shibasaki, K. (1978). Major proteins of soybean seeds. Subunit structure of .beta.-conglycinin. Journal of agricultural and food chemistry *26*, 692-695.

Wadahama, H., Iwasaki, K., Matsusaki, M., Nishizawa, K., Ishimoto, M., Arisaka, F., Takagi, K., and Urade, R. (2012). Accumulation of β-Conglycinin in Soybean Cotyledon through the Formation of Disulfide Bonds between α′- and α-Subunits. Plant physiology *158*, 1395-1405.

Walling, L., Drews, G. N., and Goldberg, R. B. (1986). Transcriptional and posttranscriptional regulation of soybean seed protein mRNA levels. Proc Natl Acad Sci USA *83*, 5.

Watanabe, Y., and Hirano, H. (1994). Nucleotide Sequence of the Basic 7S Globulin Gene from Soybean. Plant physiology *105*, 1019-1020.

Wilson, K.A., Rightmire, B.R., Chen, J.C., and Tan-Wilson, A.L. (1986). Differential Proteolysis of Glycinin and beta-Conglycinin Polypeptides during Soybean Germination and Seedling Growth. Plant physiology *82*, 71-76.

Wolf, W.J., and Nelsen, T.C. (1996). Partial Purification and Characterization of the 15S Globulin of Soybeans, a Dimer of Glycinin. Journal of agricultural and food chemistry *44*, 785-791.

Wolf, W.J., Rackis, J.J., Smith, A.K., Sasame, H.A., and Babcock, G.E. (1958). Behavior of the 11S Protein of Soybeans in Acid Solutions. I. Effects of pH, Ionic Strength and Time on Ultracentrifugal and Optical Rotatory Properties2. Journal of the American Chemical Society *80*, 5730-5735.

WOLF, W.T., T. (1969). Heat denaturation of soybean 11S protein. Cereal Chem *46*, 14.

Yagasaki K, K.N., Kitamura K (1996). Inheritance of glycinin subunits and characterization of glycinin molecules lacking the subunits in soybean (Glycine max L. Merr.). Breed Sci *46*, 5.

Yamauchi, F.S., K.; Yamagishi, T. (1984). Isolation and partial characterization of a salt-extractable globulin from soybean seeds. Agric Biol Chem *48*, 6.

Yamazaki, T., Kishimoto, K., Miura, S., and Ezaki, O. (2012). Dietary β-conglycinin prevents fatty liver induced by a high-fat diet by a decrease in peroxisome proliferator-activated receptor γ2 protein. The Journal of nutritional biochemistry *23*, 123-132.

Yoshino M, K.A., Tsutsumi K, Nakamura I, Takahashi K, Shimamoto Y (2002). Structural variation around the gene encoding the a subunit of soybean b-conglycinin and correlation with the expression of the a subunit. Breed Sci *52*, 7.

Yoshizawa, T., Shimizu, T., Yamabe, M., Taichi, M., Nishiuchi, Y., Shichijo, N., Unzai, S., Hirano, H., Sato, M., and Hashimoto, H. (2011). Crystal structure of basic 7S globulin, a xyloglucan-specific endo-beta-1,4-glucanase inhibitor protein-like protein from soybean lacking inhibitory activity against endo-beta-glucanase. The FEBS journal *278*, 1944-1954.

Figure Legends:

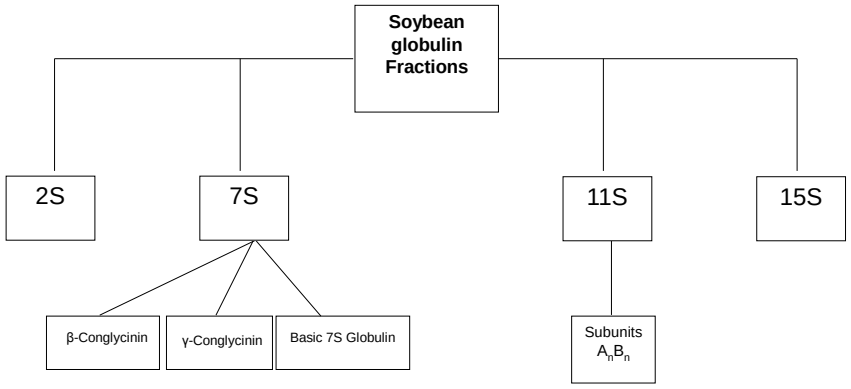

Figure 1

Figure 1: Classification of globulin fractions of soybean seed proteins.

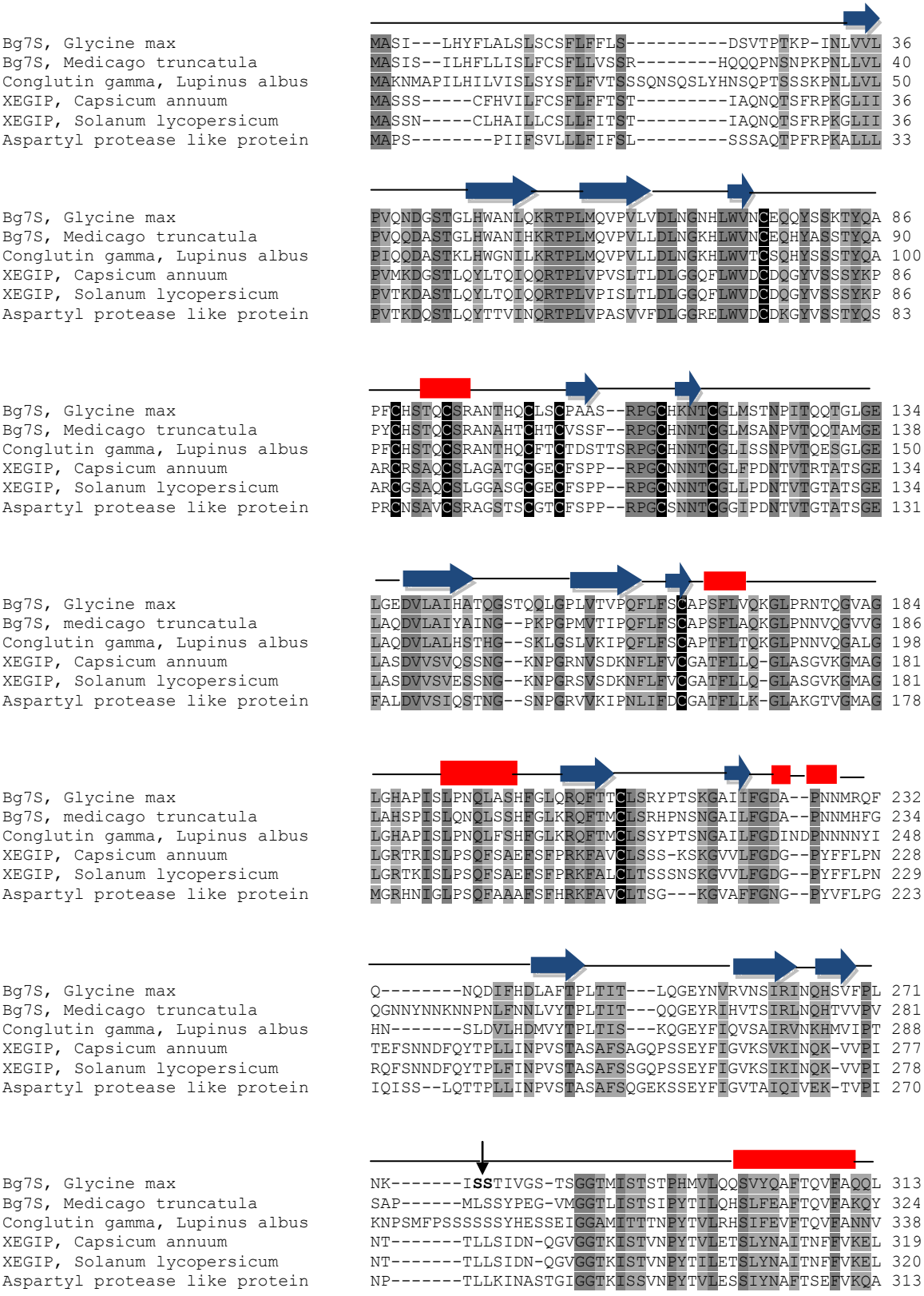

Figure 2: Multiple sequence alignment of Bg7S with similar proteins. 12 highly conserved cysteine residues are highlighted in black. The highly conserved and less conserved residues are highlighted in dark and light grey, respectively. An arrow indicates the site of cleavage during maturation to yield low and high kDa subunits. Secondary structure elements of Bg7S are shown on the top of amino-acid sequences; loop residues are shown in black line, β-strands as filled arrows (blue) and α-helices are shown in rectangles (red).

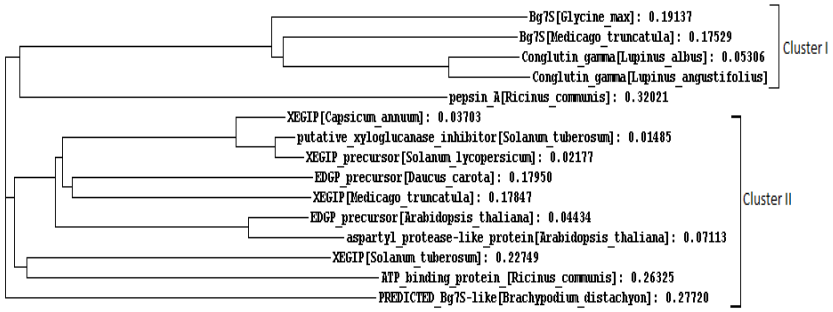

Figure 3

Figure 3: Phylogenetic tree showing relation of Bg7S with other related proteins. Tree was prepared using ClustalW2, with multiple sequence alignment covering all the 427 amino acid residues.

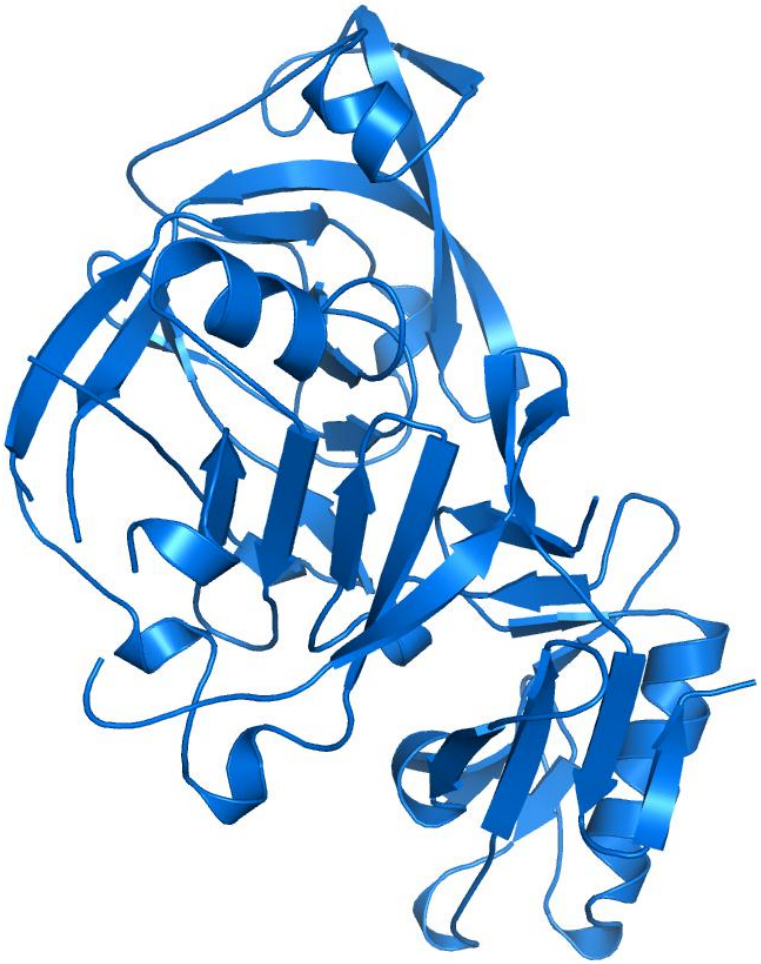

Figure 4: Structural representation of monomer of Bg7S *[Glycine max]* drawn in PyMol by using atomic coordinates of PDB id 3AUP.

Table 1: Motifs present on sequence of basic 7S globulin

Motif	Signature sequence	Corresponding residues in parent sequence
PRSTNOIDEP1R	LSCSFLFFLSDSVTPTKPINLVVLP	13 to 37
	QAFTQVFAQQLPKQAQVKSVA	303 to 323
PEPSIN	KRTPLMQVPVLVDLNGNHLWV	53 to 73
	GAIIFGDAPNNMRQ	218 to 231
NGELATINASE	QGSTQQLGPLVTV	146 to 158
	EYNVRVNSIRINQHSVFPL	253 to 271
RIBOSOMALS2	GLHWANLQKR	45 to 54
	VQKGLPRNTQGVAGL	171 to 185
OTCASE	QTGLGELGEDVLAI	129 to 142
	LVVFDLARSRVG	394 to 405
TGFBETA2	MASILHYFLALSL	1 to 22
	VLVDLNGNHLWVNCEQQYSSKTY	71 to 93

HOLDHDRGNASE	VFPLNKISSTIVGSTSGGTMISTSTPH	268 to 294
	SGEDLMVQAQPGVTCLGVMN	355 to 374
FLGPRINGFLGI	KNTCGLMSTNPITQQTGLGE	115 to 134
	SSTIVGSTSGGTM	275 to 287

Table 2: List of proteins showing close structural similarity to Basic 7S Globulin

Name of protein	PDB code	Number of residues		Z-score	RMSD	Sequence Identity (%) (%)
		Total	Super-imposed			
Basic 7S Globulin [Glycine max]	3aup-A	377	377	67.8	0.0	100
TAXI-IA in complex with ANXY	1t6g-A	368	330	39.7	2.3	30
TAXI-IIA In complex with Bacillus subtilis xylanase	3hd8-A	370	332	39.4	2.3	30
Human UROPEPSIN	1flh-A	326	288	27.5	2.5	20
Aspartic proteinase [Mucor pusillus]	1mpp-A	357	305	29.7	2.6	16
Pocrine PEPSIN	3pep-A	326	286	27.0	2.7	19
Bovine CHYMOSIN	4cms-A	320	285	26.8	2.7	18
Human β-SECRETASE 1	3k5g-A	377	296	26.8	2.9	19
Yeast PROTEINASE A	2jxr-A	330	289	26.2	2.9	17
Human CATHEPSIN-D	1lyw-B	241	214	15.8	3.1	19

ACCEPTED MANUSCRIPT