Structure and function of seed storage proteins in faba bean (*Vicia faba* L.)

Yujiao Liu¹ · Xuexia Wu¹ · Wanwei Hou¹ · Ping Li¹ · Weichao Sha¹ · Yingying Tian¹

Received: 21 December 2016 / Accepted: 10 March 2017 / Published online: 27 April 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract The protein subunit is the most important basic unit of protein, and its study can unravel the structure and function of seed storage proteins in faba bean. In this study, we identified six specific protein subunits in Faba bean (cv. Qinghai 13) combining liquid chromatography (LC), liquid chromatography-electronic spray ionization mass (LC–ESI–MS/MS) and bio-information technology. The results suggested a diversity of seed storage proteins in faba bean, and a total of 16 proteins (four GroEL molecular chaperones and 12 plant-specific proteins) were identified from 97-, 96-, 64-, 47-, 42-, and 38-kD-specific protein subunits in faba bean based on the peptide sequence. We also analyzed the composition and abundance of the amino acids, the physicochemical characteristics, secondary structure, three-dimensional structure, transmembrane domain, and possible subcellular localization of these identified proteins in faba bean seed, and finally predicted function and structure. The three-dimensional structures were generated based on homologous modeling, and the protein function was analyzed based on the annotation from the non-redundant protein database (NR database, NCBI) and function analysis of optimal modeling. The objective of this study was to identify the seed storage proteins in faba bean and confirm the structure and function

of these proteins. Our results can be useful for the study of protein nutrition and achieve breeding goals for optimal protein quality in faba bean.

Keywords Faba bean · Protein subunit · LC–ESI–MS/MS · Bioinformatics · Structure · Function

Introduction

The seeds of flowering plants usually accumulate large quantities of seed protein including storage protein and lectins. The amount of protein found in seed varies from 10% (in cereals) to 40% (in certain legumes and oilseeds) of the seed dry weight, which forms a major source of dietary protein (Shimada et al. 2003; Shewry et al. 1995; Shewry and Halford 2002). Faba bean is consumed worldwide as a rich source of plant protein for humans and animals. In addition to its high protein content, faba bean is also a globally important nitrogen-fixing legume (Sosulski and McCurdy 1987; Cazzato and Tufarelli 2012; Webb et al. 2016).

Seed storage proteins account for almost 80% of the total seed protein in weight (Shewry and Halford 2002). In faba bean, seed storage proteins mainly include globulin and albumin (Liu et al. 2012b; Shi et al. 2012, 2013). Globulins are widely distribution and can be divided into two groups based on their sedimentation coefficient ($S_{20,w}$): the 7S vicilin-type globulins and the 11S legumin-type globulins. The 7-11S globulins tend to be deficient in the sulfur-containing amino acids cysteine and methionine and also tryptophan. Albumin, a functional protein composed of trypsin inhibitor and phytolectin, contains more sulfur-containing amino acids than globulin (Utsumi et al. 1980; Mandal and Mandal 2000).

Electronic supplementary material The online version of this article (doi:10.1007/s13205-017-0691-z) contains supplementary material, which is available to authorized users.

✉ Yujiao Liu
13997058356@163.com

¹ State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Qinghai Academy of Agricultural and Forestry Science, Xining, Qinghai 810016, People's Republic of China

The amino acid composition of seed storage proteins is important for protein quality, and thus impacts the various end uses (Shewry et al. 1995). However, seed proteins in general are deficient in some essential amino acids which compromises nutritional quality. For example, legumes are deficient in sulfur-containing amino acids (cysteine and methionine), and cereal proteins are usually deficient in lysine and tryptophan which may limit the nutrition quality of seed (Shewry et al. 1995; Mandal and Mandal 2000). Therefore, exploring the amino acid composition is also important in evaluating the nutritional quality of seed storage proteins in faba bean.

Seed storage proteins have been widely studied in soybean, rice, *Arabidopsis thaliana*, wheat, and pea (Scott et al. 1992; Meinke et al. 1981; Choi et al. 2000; Shimada et al. 2003; Wu et al. 2016; Gatehouse et al. 1983). Scientists also evaluated the protein nutrition and agricultural importance of seed storage proteins in faba bean (Xia et al. 2016; Chandra-Hioe et al. 2016; Makri et al. 2005; Carbonaro et al. 2000). Bailey and Boulter (1970) analyzed the structure of legumin in faba bean seed storage proteins using peptide mapping techniques after a tryptic digest. Valizadeh (2001) analyzed the seed storage protein profile of grain legumes grown in Iran using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). However, the seed storage protein profile of faba bean has not been explored comprehensively, and the structure of the proteins is also unknown. Our previous work only analyzed the protein composition and protein subunits for different genotypes of faba bean (Liu et al. 2012a, b; Hou et al. 2015). Given the importance of the nutritional evaluation and protein consumption of seed storage proteins in faba bean, exploring the variation of the proteins in faba bean would help breeders in their selection efforts to develop new varieties. Our study may supply a reliable data for breeding high-quality faba bean varieties. Meanwhile, the identification of proteins would provide a window into complex cellular regulatory networks (Domon and Aebersold 2006).

Biological mass spectrometry (MS) possesses an advantage when exploring protein structure and function because of its ability to precisely predict the relative molecular weight, amino acid sequence, and Post-transcriptional modification in a polypeptide or protein (Smith et al. 1997). In addition, a complex protein mixture can be distinguished by combining liquid chromatography (LC) and tandem mass spectrometry (MS–MS) (Chen 2009; Link et al. 1999). Therefore, we used LC, LC–ESI–MS/MS, and bio-information technology to analyze the seed storage proteins of faba bean in this study. We specifically evaluated as follows: (1) the seed storage protein composition and their abundance in faba bean; (2) their physicochemical characteristics and amino acid composition; (3)

and the structure and function prediction of these seed storage proteins.

Materials and methods

Materials

The Faba bean (*Vicia faba* L.), cv. Qinghai 13 cultivated by Qinghai Academy of Agricultural and Forestry Science was used in this study. Endoproteinase trypsin was obtained from Promega (Madison, WI). All other chemicals used in proteolytic digestion and high-performance liquid chromatography (HPLC) were obtained from Sigma. The following instruments were used in this study: high-speed centrifuge (Thermo, Palo Alto, CA), high-performance liquid chromatograph (Agilent), and Quadrupole ion trap mass spectrometer (Thermo). The non-redundant protein database was from NCBI's GenBank.

Preparation and separation of protein

Seed storage proteins of faba bean were extracted according to Kumamaru et al. (1988) with some modification. Mature dry seeds were ground to a fine powder after stripping coat and screened through a mesh size of 80. To isolate the seed storage proteins, 50 mg of finely ground seed powder was transferred to a 1.5 mL microcentrifuge tube containing 35 μ L β -mercaptoethanol and infiltrated for 20 min. Next, 675 μ L protein extract buffer (8 M urea, 4% SDS, 20% glycerinum, 50 mM Tris-HCl, pH = 6.8) was added which was preheated at 40 °C. The tube was vigorously vortexed for 30 s until the contents were transparent. After setting for 48 h at room temperature, the contents were centrifuged at 8000 $\times$ g for 5 min at 4 °C. The clear supernatant was resolved by SDS-PAGE.

The seed storage proteins were first separated by SDS-PAGE and then stained with coomassie brilliant Blue. Then the band was excised from the gel for protein identification with MS/MS. The sample treatment followed a commonly used protocol (Shevchenko et al. 1996). In brief, the sample gel was cut from a target region of the gel and each protein gel band was destained with 50% MeOH in water containing 2% acetic acid for overnight. Reduction of disulfide linkages was carried out with dithiothreitol (DTT) followed by alkylation with iodoacetamide for 15 min at room temperature in the dark. The sample was then washed with water, and digested in-gel with sequencing grade-modified trypsin in the digestion buffer (ammonium bicarbonate 100 mM, pH 8.5). The peptides from the digestion were extracted with acetonitrile, and completely dried down in a SpeedVac device (Thermo). The dried sample was then redissolved in sample solution (2% acetonitrile, 97.5% water,

0.5% formic acid). A dissolved peptide sample was then analyzed by a NanoLC–ESI–MS/MS system.

NanoLC–ESI–MS/MS analysis

NanoLC–ESI–MS/MS analysis of digested protein samples was performed by an HPLC system with a 75 µm inner diameter and 8 cm in length reverse-phase C18 column. The particle size of the C18 was 3 µm with a pore size of 300 Å. The injection time was 20 min. The HPLC Solvent A was 97.5% water, 2% acetonitrile, and 0.5% formic acid. The HPLC Solvent B was 9.5% water, 90% acetonitrile, and 0.5% formic acid. The gradation time was 60 min from 2% Solvent B to 90% solvent B, plus 20 min for sample loading and 20 min for column washing. The column flow rate was around 800 nL per minute after splitting. The typical injection volume was 3 µL.

The HPLC system was coupled with an ion trap mass spectrometer (LCQ DECA XP PLUS, Thermo) so that a sample eluted from the HPLC column was directly ionized by an electrospray ionization (ESI) process and entered into the mass spectrometer. The ionization voltage was optimized each time and was normally in a range of 1.2–1.8 kv. The capillary temperature was set at 110 °C. The mass spectrometer was set at the data-dependent mode to acquire MS/MS data via a low-energy collision-induced dissociation (CID) process. The default collision energy was 33%, and the default charge state was 3. One full scan with one microscan with a mass range of 550 amu to 1800 amu was acquired, followed by one MS/MS scan of the most intense ion with a full mass range and three microscans. The dynamic exclusion feature was set as follows: repeat count of 1 within 0.3 min and exclusion duration of 0.4 min. The exclusion width was 4 Da.

Database search and validation

As a default, the mass spectrometric data were used to search against the most recent non-redundant protein database (NR database, NCBI) with ProtTech's ProtQuest software suite. The output from the database search was manually validated by a senior scientist before reporting. The parameters of the data acquisition and data processing were as following:

Data acquisition parameters:

Instrument: LCQ DECA XP Plus

Software: Xcalibur 2.0

Centroid mode

Minimum MS signal for Precursor-ion: 5×10^4 counts

Processing parameters:

Software: ProtQuest 2.0

Signal-to-noise: ≥ 5

De-isotoped: Yes

Filtered: Yes

Missed cleavages: ≤ 5

Modifications: Cys +57

Database: NR database from NCBI's GenBank (update on 9/9/2011)

Validation parameters:

Peptide = 1, score ≥ 13

Peptide > 1, score ≥ 15

Label-free protein quantitation

Relative protein abundance was calculated based on the method published by Griffin et al. (2010). The scoring function was based on MS abundance recorded in both MS and MS/MS datasets, spectral count (SC, number of MS/MS spectra per peptide), unique peptide number (PN), and fragment ion (MS/MS) intensities. The normalized spectral index (SI_N) was calculated for each gel slice based on the following equation. The final relative protein abundance is the percentage of each SI_N in the total SI_N in each sample.

$$SI_N = \left[\frac{pn}{\sum_{k=1}^{pn} \left(\frac{sc}{\sum_{j=1}^{sc} i_j} \right)} \right] \bigg/ \left[\frac{n}{\sum_{j=1}^n SI_j} \right] \bigg/ L,$$

where pn is the unique peptide number, sc is the spectral count, i is the fragment ion intensity of peptide k , n is the total number of proteins identified, and L is the number of amino acids in a protein.

Bioinformatics analysis

Physicochemical characteristics including the amino acid amount, amino acid composition, molecular weight, theoretical isoelectric point, instability index, aliphatic index, grand average of hydropathicity (GRAVY), extinction coefficient, and total number of negatively/positively charged residues of proteins were calculated using the ProtParam tool (<http://www.expasy.org/tools/protparam.html>). Subcellular localization was analyzed using CELLO version 2.5 (<http://cello.life.nctu.edu.tw/>) (Yu et al. 2004, 2006). The secondary structure was predicted by SPOMA (self-optimized prediction method with alignment). To determine if a protein is a membrane protein, the online servers TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) and HMMTOP (<http://cctop.enzim.ttk.mta.hu/>) were used. The predicted three-dimensional structures of identified proteins were generated using the SWISS-MODEL server (<http://swissmodel.expasy.org/>) (Biasini et al. 2014; Wang et al. 2016).

Structural evaluation and stereochemical analyses were assessed using RAMPAGE Ramachandran plot analysis (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Lovell et al. 2003).

Results

Mass spectrometry analysis of different protein subunits

The protein composition, peptide mass, and peptide sequence for subunits of seed storage proteins in faba bean were obtained using LC–ESI–MS/MS. Six protein subunits of 97, 96, 64, 47, 42, and 38 kD were identified in faba bean. In total, 16 proteins were identified in these six protein subunits, and the molecule weight ranged from 20 to 57 kD (Table S1). All peptide sequences were listed and each peptide was sequenced by LC–MS independently, so even a single peptide could lead to a highly confident protein identification. Each protein was hydrolyzed into 1–10 peptide fragments through enzymatic hydrolysis at the arginine and lysine sites, and the calculated relative abundance of peptides ranged from 5.0 to 100%. In total, 42 peptide fragments were identified in the 16 seed storage proteins of faba bean (Table S1). Each of the peptide fragments was composed of 8–23 amino acids (Table S1).

Basic information of identified seed storage proteins

For further analysis, amino acid sequence alignment was performed by searching the NR database of NCBI (<http://www.ncbi.nlm.nih.gov/protein>) based on identified peptide sequences. Table S2 lists 16 identified proteins from each sample. Since our results are based on MS/MS peptide sequencing of each individual peptide, all proteins listed here are actual proteins present in the sample instead of “potential candidates”, which is often seen in a MALDI-TOF-based peptide mapping. For any identified protein here, there should be >98% certainty if the identification is based on LC–MS/MS sequence of one peptide, and it will have >99.9% certainty if it is based on sequencing of two or more peptides.

In total, 16 proteins, four GroEL molecular chaperones, and 12 plant-specific proteins (phosphate ABC transporter periplasmic substrate-binding protein, Tu elongation factor Tu, electron transfer flavoprotein subunit alpha, alkyl hydroperoxide reductase C22 subunit, citrate synthase, motA/TolQ/ExbB proton channel family protein, convicilin, legumin A precursor, putative sucrose-binding protein, LEGB7_VICFA, and Htlv-1 Gp21 Ectodomain MALTOSE-Binding Protein Chimera) were identified (Table S2). The 97-kD subunit was recognized as GroEL molecular chaperone, and its relative abundance reached up to 100%, which was much higher than in the 64-, 42-, and 38-kD subunits (Table S2). Among those 12 identified plant-specific proteins, convicilin in the 64-kD subunit, legumin (a precursor in the 47- and 42-kD subunits), putative sucrose-binding protein in the 47-kD subunit, and

LEGB7_VICFC in the 38-kD subunit were recognized as specific proteins which had been identified in leguminous plants like *Vicia faba*, *Vicia*, and others (Table S2). Six proteins from the 96-kD subunit were all identified as specific proteins, and the sequence coverage ranged from 4.01 to 15.84% (Table S2).

The results from CELLO revealed various subcellular location patterns of those identified proteins. The results exhibited GroEL chaperonin, elongation factor Tu, electron transfer flavoprotein subunit alpha, alkyl hydroperoxide reductase C22 subunit, citrate synthase, convicilin, and putative sucrose-binding protein all located in the cytoplasm; phosphate ABC transporter periplasmic substrate-binding protein and Htlv-1 Gp21 ectodomain MALTOSE-binding protein chimera all located periplasmic; motA/TolQ/ExbB proton channel family protein located at the inner membrane; legumin A precursor located at the outer membrane; and LEGB7_VICFA which showed a wide distribution to the outer membrane, extracellular region, and cytoplasm (Table S2). The diverse subcellular localization of those seed proteins in faba bean imply a variety of functions.

Analysis of protein primary structure

The sequence of amino acids is the foundation of the native state of a protein, so we first analyzed the protein primary structure using Expert Protein Analysis (ExPASy, <http://expasy.org/tools>) in order to precisely predict the potential function and basic structure of seed storage proteins in faba bean. As shown in Table 1, the number of amino acids in the 16 identified proteins ranged from 187 to 497 (Table 1). The maximal number and minimal number of amino acids was 497 in convicilin and 187 in alkyl hydroperoxide reductase C22 subunit, respectively, which resulted in a different molecular weight of 57501.1 and 20525.2 D, respectively. The theoretical isoelectric point of identified proteins ranged from 4.90 to 9.55, which is very important to protein purification by the isoelectric focusing method. The instability index ranged from 18.59 in electron transfer flavoprotein subunit alpha to 61.92 in the legumin A precursor. The aliphatic index ranged from 75.49 to 112.6. The grand average of hydropathicity (GRAVY) was −0.957 to 0.374. Convicilin and the alkyl hydroperoxide reductase C22 subunit showed the maximum and minimum value, respectively, in the total number of negatively/positively charged residues (Table 1).

Amino acid composition assay of different proteins

There was a vast difference among different proteins both in amino acid amount and amino acid composition. In legumes, the low content of cysteine and methionine in

Table 1 The primary structure of seed storage proteins in faba bean

	GroEL chaperonin	Phosphate ABC transporter periplasmic substrate-binding protein	Elongation factor Tu	Electron transfer flavoprotein subunit alpha	Alkyl hydroperoxide reductase C22 subunit	Citrate synthase
Number of amino acids	494	341	392	311	187	424
Molecular weight	52,458.9	36,477.8	42,553.4	31,437.8	20,525.2	47,239.2
Isoelectric point	5.15	9.55	5.15	4.90	4.96	5.92
Instability index	23.43	32.49	28.86	18.59	21.69	25.64
Aliphatic index	100.85	84.75	89.72	112.06	89.20	89.76
Grand average of hydropathicity	-0.111	-0.167	-0.173	0.374	-0.079	-0.091
Extinction coefficient	8940	23,380	17,880	12,950	30,946	48,820
Total number of negatively charged residues (Asp + Glu)	76	30	59	34	26	52
Total number of positively charged residues (Arg + Lys)	64	46	42	24	16	44
(Cys + Met)	11	9	15	4	3	19

	MotA/TolQ/ExbB proton channel family protein	Convictin	Legumin A precursor	Putative sucrose-binding protein	LEGB7_VICFA	Htlv-1 Gp21 ectodomain MALTOSE-binding protein chimera
Number of amino acids	266	497	482	482	335	450
Molecular weight	27,416.4	57,501.1	54,667.9	54,614.4	37,827.8	49,332.2
Isoelectric point	9.38	6.16	7	6.86	5.46	5.58
Instability index	34.85	58.87	61.92	54.39	58.83	23.04
Aliphatic index	101.77	75.49	77.1	81.18	79.13	87.93
Grand average of hydropathicity	0.288	-0.957	-0.777	-0.508	-0.789	-0.265
Extinction coefficient	22,460	38,850	35,870	20,400	25,900	73,340
Total number of negatively charged residues (Asp + Glu)	17	81	66	71	47	55
Total number of positively charged residues (Arg + Lys)	22	77	66	70	38	48
(Cys + Met)	4	3	8	15	1	10

Table 2 The amino acid information of identified proteins from faba bean seeds

Amino acid (number/%)	Protein					
	GroEL chaperonin	Phosphate ABC transporter periplasmic substrate-binding protein	Elongation factor Tu	Electron transfer flavoprotein subunit alpha	Alkyl hydroperoxide reductase C22 subunit	Citrate synthase
Ala (A)	53/10.7	44/12.9	30/7.7	57/18.3	19/10.2	44/10.4
Arg (R)	26/5.3	13/3.8	21/5.4	7/2.3	3/1.6	20/4.7
Asn (N)	15/3	18/5.3	8/2	14/4.5	5/2.7	18/4.2
Asp (D)	36/7.3	14/4.1	26/6.6	19/6.1	13/7	27/6.4
Cys (C)	0/0	4/1.2	5/1.3	0/0	2/1.1	4/0.9
Gln (Q)	14/2.8	10/2.9	9/2.3	8/2.6	6/3.2	9/2.1
Glu (E)	40/8.1	16/4.7	33/8.4	15/4.8	13/7	25/5.9
Gly (G)	50/10.1	28/8.2	40/10.2	30/9.6	16/8.6	31/7.3
His (H)	3/0.6	4/1.2	11/2.8	4/1.3	6/3.2	13/3.1
Ile (I)	34/6.9	21/6.2	25/6.4	24/7.7	13/7	27/6.4
Leu (L)	40/8.1	21/6.2	27/6.9	21/6.8	13/7	37/8.7
Lys (K)	38/7.7	33/9.7	21/5.4	17/5.5	13/7	24/5.7
Met (M)	11/2.2	5/1.5	10/2.6	4/1.3	1/0.5	15/3.5
Phe (F)	9/1.8	13/3.8	11/2.8	5/1.6	10/5.3	20/4.7
Pro (P)	11/2.2	13/2.8	19/4.8	9/2.9	9/4.8	19/4.5
Ser (S)	24/4.9	25/7.3	16/4.1	17/5.5	6/3.2	15/3.5
Thr (T)	30/6.1	18/5.3	27/6.9	14/4.5	13/7	24/5.7
Trp (W)	0/0	1/0.3	0/0	1/0.3	4/2.1	4/0.9
Tyr (Y)	6/1.2	12/3.5	12/3.1	5/1.6	6/3.2	18/4.2
Val (V)	54/10.9	28/8.2	41/10.5	40/12.9	16/8.6	30/7.1
Pyl (O)	0/0	0/0	0/0	0/0	0/0	0/0
Sec (U)	0/0	0/0	0/0	0/0	0/0	0/0
Total	494	341	392	311	187	424

Amino acid (number/%)	Protein					
	MotA/TolQ/ExbB proton channel family protein	Convicilin	Legumin A precursor	Putative sucrose-binding protein	LEGB7_VICFA	Htlv-1 Gp21 ectodomain MALTOSE-binding protein chimera
Ala (A)	43/16.2	22/4.4	33/6.8	34/7.1	21/6.3	56/12.4
Arg (R)	9/3.4	44/8.9	48/10	19/3.9	28/8.4	8/1.8
Asn (N)	9/3.4	34/6.8	39/8.1	33/6.8	25/7.5	28/6.2
Asp (D)	8/3	19/3.8	24/5	20/4.1	10/3	26/5.8
Cys (C)	1/0.4	2/0.4	6/1.2	7/1.5	1/0.3	3/0.7
Gln (Q)	13/4.9	32/6.4	31/1.26.4	12/2.5	28/8.4	17/3.8
Glu (E)	9/3.4	62/12.5	42/8.7	51/10.6	37/11	29/6.4
Gly (G)	28/10.5	28/5.6	32/6.6	26/5.4	25/7.5	33/7.3
His (H)	3/1.1	7/1.4	5/1	11/2.3	7/2.1	6/1.3
Ile (I)	17/6.4	23/4.6	21/4.4	27/5.6	15/4.5	26/5.8
Leu (L)	28/10.5	46/9.3	45/9.3	49/10.2	29/8.7	44/9.8
Lys (K)	13/4.9	33/6.6	18/3.7	51/10.6	10/3	40/8.9
Met (M)	3/1.1	1/0.2	2/0.4	8/1.7	0/0	7/1.6
Phe (F)	11/4.1	19/3.8	18/3.7	31/6.4	10/3	17/3.8
Pro (P)	8/3	22/4.4	23/4.8	21/4.4	16/4.8	23/5.1
Ser (S)	24/9	38/7.6	33/6.8	31/6.4	23/6.9	18/4
Thr (T)	14/5.3	18/3.6	18/3.7	19/3.9	13/3.9	21/4.7

Table 2 continued

Amino acid (number/%)	Protein					
	MotA/TolQ/ExbB proton channel family protein	Convicilin	Legumin A precursor	Putative sucrose-binding protein	LEGB7_VICFA	Htlv-1 Gp21 ectodomain MALTOSSE-binding protein chimera
Trp (W)	3/1.1	3/0.6	3/0.6	1/0.2	2/0.6	9/2
Tyr (Y)	4/1.5	15/3	13/2.7	10/2.1	10/3	16/3.6
Val (V)	18/6.8	29/5.8	28/5.8	21/4.4	25/7.5	23/5.1
Pyl (O)	0/0	0/0	0/0	0/0	0/0	0/0
Sec (U)	0/0	0/0	0/0	0/0	0/0	0/0
Total	266	497	482	482	335	450

Table 3 The secondary structure of identified seed storage proteins in faba bean

Protein	Alpha helix		Extended strand		Beta turn		Random coil	
	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)
GroEL chaperonin	232	46.96	91	18.42	56	11.34	115	23.28
Phosphate ABC transporter periplasmic substrate-binding protein	141	41.35	71	28.82	37	10.85	92	26.98
Elongation factor Tu	99	25.26	112	28.57	49	12.50	132	33.67
Electron transfer flavoprotein subunit alpha	136	43.73	60	19.29	42	13.50	73	23.47
Alkyl hydroperoxide reductase C22 subunit	56	29.95	48	25.67	20	10.70	63	33.69
Citrate synthase	178	41.98	74	17.45	43	10.14	129	30.42
MotA/TolQ/ExbB proton channel family protein	156	58.65	30	11.28	16	6.02	64	24.06
Convicilin	136	27.36	127	25.55	43	8.65	191	38.43
Legumin A precursor	148	30.71	111	23.03	58	12.03	165	34.23
Putative sucrose-binding protein	181	37.55	114	23.65	47	9.75	140	29.05
LEGB7_VICFA	117	34.93	67	20.00	41	12.24	110	32.84
Htlv-1 Gp21 ectodomain MALTOSSE-binding protein chimera	193	42.89	86	19.11	52	11.56	119	26.44

major storage proteins may limit the nutritional quality of the seed for monogastric animals (Shewry et al. 1995), so the content of sulfur-containing amino acids (cysteine and methionine) is an important criteria when evaluating protein quality.

The amino acid composition and percentage in specific proteins are shown in Table 2. As shown, the convicilin contained 62 glutamic and 46 leucine, which made up 12.5 and 9.3% of the total amino acid amount, respectively. In the legumin A precursor, the content of arginine was the highest and next was leucine, which accounted for 10 and 9.3% of the total amino acid amount, respectively. There were 51 lysine and 51 glutamic in putative sucrose-binding protein, accounting for the highest percentage at 10.6% each. Leucine was the second-most abundant amino acid and made up 10.2% of the total amino acid amount in

putative sucrose-binding protein. In LEGB7_VICFA, there were 37 glutamic and 29 leucine, accounting for 11.0 and 8.7%, respectively. The sulfur-containing amino acid number in the four proteins above was 3, 8, 15, and 1, which corresponds with 0.6, 1.6, 3.2, and 0.3%, respectively. In addition, elongation factor Tu and citrate synthase showed a high sulfur-containing amino acid content with percentages of 3.9 and 4.4%, respectively. These results suggest that the proteins in faba bean are diverse in amino acid amount and composition.

Secondary structure prediction of different proteins

Protein secondary structure is the basic spatial structure of a protein. As shown in Table 3, alpha helix, extended strand, beta turn, and random coil are the most common

secondary structures in identified proteins. Abundance of alpha helices in the identified proteins ranged from 25.26% in elongation factor Tu to 58.65% in the motA/TolQ/ExbB proton channel family. Electron transfer flavoprotein subunit alpha contained the most beta turns with an abundance of 13.50%, and the content of beta turns in the motA/TolQ/ExbB proton channel family was the lowest. Elongation factor Tu possessed the most extended strands with an abundance of 28.57%, whereas the percentage of extended strands in the motA/TolQ/ExbB proton channel family was only 11.28%. The amount of random coils ranged from 23.28% in GroEL chaperonin to 38.43% in convicilin.

Transmembrane domain prediction

To characterize a protein to be a membrane protein, we further detected the transmembrane domain of those identified proteins depending on the online servers TMHMM and HMMTOP (data from HMMTOP not shown), and acquired consistent results. The results suggest that the motA/TolQ/ExbB proton channel family protein contained three transmembrane spanning domains, consistent with the roles that the motA/TolQ/ExbB proton channel family protein undertakes in intracellular trafficking, secretion, and vesicular transport. The results also suggest that convicilin and putative sucrose-binding protein possess one transmembrane domain each (Fig. 1). There were no transmembrane domains found in GroEL chaperonin, phosphate ABC transporter periplasmic substrate-binding protein, elongation factor Tu, electron transfer flavoprotein subunit alpha, alkyl hydroperoxide reductase C22 subunit, citrate synthase, legumin A precursor, legumin A precursor, LEGB7_VICFA, or Htlv-1 Gp21 ectodomain MALT-TOSE-binding protein chimera (Fig. 1). The various transmembrane domains of proteins in faba bean seeds are closely related to the diverse functions.

Prediction of three-dimensional structure for identified proteins

Model quality estimation is an essential component of protein structure prediction, as the accuracy of a model determines its usefulness for practical applications (Biasini et al. 2014). A reliable three-dimensional structure based on homologous modeling was constructed using SWISS-MODEL and is shown in Fig. 2. The optimum templates information is listed in Table 4. We used homology modeling for structure prediction of GroEL chaperonin, elongation factor Tu, alkyl hydroperoxide reductase C22 subunit, citrate synthase, and legumin A precursor, where all the sequence identities were >60% between the target and template sequences (Table 4). Htlv-1 Gp21 ectodomain MALT-TOSE-binding protein chimera even shared a

high sequence identity of 100% with the template (PDB ID 1mg1), which is consistent with the three-dimensional structure characteristic predicted by Kobe et al. (1999). LEGB7_VICFA shared a sequence identity of 45.90% with the template (Table 4). As shown in Table 4, we also used homology modeling for structure prediction of phosphate ABC transporter periplasmic substrate-binding protein, motA/TolQ/ExbB proton channel family protein, and putative sucrose-binding protein, where the sequence identity was 23.26, 27.44, and 34.61%, respectively.

The GMEQ (global quality estimation score), in which the most likely structural similarity is the value at which the joint distribution is maximized, is an important indicator expressed as a number between zero and one (Biasini et al. 2014). GroEL chaperonin, elongation factor Tu, electron transfer flavoprotein subunit alpha, alkyl hydroperoxide reductase C22 subunit, citrate synthase, and Htlv-1 Gp21 ectodomain MALT-TOSE-binding protein chimera showed high GMEQ values of 0.8, 0.79, 0.8, 0.9, 0.82, and 0.99, respectively, meaning a high reliability of the hypothetical three-dimensional structure. Phosphate ABC transporter periplasmic substrate-binding protein, motA/TolQ/ExbB proton channel family protein, convicilin, legumin A precursor, putative sucrose-binding protein, and LEGB7_VICFA showed relatively low values of 0.51, 0.58, 0.53, 0.73, 0.56, and 0.65, respectively. Coverage, another property of model quality, ranged from 0.71 to 1.

In addition, structural evaluation and stereochemical analyses were assessed using RAMPAGE Ramachandran plot analysis

(<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>)

(Lovell et al. 2003; Wang et al. 2016). The refined models showed good proportions of residues in favored, allowed, and outlier regions except LEGB7_VICFA (Online Resource 1 and Table 4). The models of GroEL chaperonin, elongation factor Tu, electron transfer flavoprotein subunit alpha, alkyl hydroperoxide reductase C22 subunit, citrate synthase, motA/TolQ/ExbB proton channel family protein, and Htlv-1 Gp21 ectodomain MALT-TOSE-binding protein chimera showed >95% of the residues in the allowed region of the Ramachandran plot (Online Resource 1 and Table 4).

Functional annotation and prediction of identified proteins

Structure similarity is more consistent than sequence similarity (Taylor and Orengo 1989), and the structures of homologous proteins are more conserved than their sequences (Chothia and Lesk 1986). The biological function cannot be predicted by comparison of sequence similarity alone (Illergard et al. 2009). In this study, protein

Fig. 1 The transmembrane domain prediction of the seed storage proteins in faba bean **a** GroEL chaperonin. **b** Phosphate ABC transporter periplasmic substrate-binding protein. **c** Elongation factor Tu. **d** Electron transfer flavoprotein subunit alpha. **e** Alkyl

hydroperoxide reductase C22 subunit. **f** Citrate synthase. **g** MotA/TolQ/ExbB proton channel family protein. **h** Convicilin. **i** Legumin A precursor. **j** Putative sucrose-binding protein. **k** LEGB7_VICFA. **l** Htlv-1 Gp21 ectodomain MALT0SE-binding protein chimera

function were explicated combined the protein structures and function annotation from the NR database. The objective was to assign a precise function to the proteins in faba bean. The major function of these proteins is listed in Online Resource 2 (Table S3).

The GroEL chaperonin was identified in the 97-, 64-, 42-, and 38-kD subunits with a relative abundance of 100, 8.4, 13.2, and 5.9%, respectively. Sequence similarity searches showed that GroEL chaperonin belongs to the HSP60 family. It was reported that GroEL chaperonin can

Fig. 2 Three-dimensional structure predictions of seed storage proteins in faba bean **a** GroEL chaperonin. **b** Phosphate ABC transporter periplasmic substrate-binding protein. **c** Elongation factor Tu. **d** Electron transfer flavoprotein subunit alpha. **e** Alkyl hydroperoxide reductase C22 subunit. **f** Citrate synthase. **g** MotA/TolQ/ExbB proton channel family protein. **h** Convicilin. **i** Legumin A precursor. **j** Putative sucrose-binding protein. **k** LEGB7_VICFA. **l** Htlv-1 Gp21 ectodomain MALTOSE-binding protein chimera

capture the polypeptide substrate and bind co-chaperone GroES and ATP to promote productive folding (Bartasaghi et al. 2012; Walter 2002). Rusanganwa and Gupta (1993) revealed a possible role of the GroEL or HSP60 chaperonin in the nodulation (symbiosis) and nitrogen fixation processes (Rusanganwa and Gupta 1993).

Phosphate ABC transporter periplasmic substrate-binding protein belongs to the periplasmic-binding protein type 2 family, which plays a role in the uptake of a variety of soluble substrates such as phosphate, sulfate, polysaccharides, lysine/arginine/ornithine, and histidine. Besides transport proteins, the family includes ionotropic glutamate receptors and unorthodox sensor proteins involved in signal transduction. The above function was confirmed based on the homology modeling of 4omb, which was classified as a transport protein and was a component of the conserved bacterial phosphate-specific transport system (Neznansky et al. 2014).

Elongation factor Tu is a multifunctional protein and has been extensively studied in several species. This protein is a GTP-binding protein and plays an important role in protein biosynthesis as it promotes the binding of aminoacyl-tRNAs to ribosomes (Dale et al. 2004; Kim et al. 2000; Daviter et al. 2003; Carlin et al. 1992). The structure of *Escherichia coli* ribosome elongation factor Tu complex was available at <3 Å resolution by C_s-corrected cryo-EM (Fischer et al. 2015).

The electron transfer flavoprotein serves as a specific electron acceptor for various mitochondrial dehydrogenases. Electron transfer flavoprotein transfers electrons to the main respiratory chain via electron transfer flavoprotein-ubiquinone oxidoreductase. Electron transfer

Table 4 The optimal three-dimensional structure information of the proteins in faba bean

Protein name	Template	Seq identity (%)	Seq similarity	Coverage	GMEQ	Ramachandran plot		
						Favored region (%)	Allowed region (%)	Outlier region (%)
GroEL chaperonin	2ynj	67.48	0.49	1	0.8	95.5	3.7	0.9
Phosphate ABC transporter periplasmic substrate-binding protein	4omb	23.36	0.32	0.8	0.51	90.9	6.8	2.3
Elongation factor Tu	5afi	84.87	0.56	0.99	0.79	97.3	2.7	0
Electron transfer flavoprotein subunit alpha	2a1u	59.02	0.46	0.98	0.8	97.4	2.6	0
Alkyl hydroperoxide reductase C22 subunit	1n8j	71.51	0.53	0.99	0.9	98.6	1.4	0
Citrate synthase	2h12	68.74	0.51	0.99	0.82	97.4	2.1	0.5
MotA/TolQ/ExbB proton channel family protein	5sv0	27.44	0.33	0.81	0.58	96.6	2.2	1.2
Convicilin	3s7e	51.40	0.45	0.72	0.53	93.3	5.6	1.1
Legumin A precursor	3ksc	89.57	0.57	0.95	0.73	92.1	6.5	1.4
Putative sucrose-binding protein	5e1r	34.61	0.39	0.82	0.56	90.1	7.0	2.9
LEGB7_VICFA	2d5 h	45.90	0.43	0.98	0.65	88.7	9.1	2.2
Htlv-1 Gp21 ectodomain MALTOSE-binding protein chimera	1mg1	100	0.61	1.00	0.99	95.1	4.5	0.4

flavoprotein is a heterodimer that consists of an alpha and a beta subunit which binds one molecule of FAD per dimer. The homologous pair of proteins (FixA/FixB) are essential for nitrogen fixation (Tsai and Saier 1995; Roberts et al. 1996). Based on the model of 2a1u, the crystal structure of this protein has been constructed (Toogood et al. 2005).

Alkyl hydroperoxide reductase C22 subunit, a kind of oxidoreductase which belongs to the family of antioxidant proteins, directly reduces organic hydroperoxides in its reduced dithiol form. It is reported that alkyl hydroperoxide reductase acts as a specific alkyl hydroperoxide-scavenging enzyme for protection against radical oxygen damage (Jacobson et al. 1989), though elimination of reactive nitrogen intermediates has also been demonstrated to occur (Chen et al. 1998). These data argue for a significant role of alkyl hydroperoxide reductase in the resistance to peroxide damage.

Citrate synthase is a hexameric enzyme that catalyzes the entry of carbon into the citric acid cycle (Gerike et al. 1998; Textor et al. 1997). It is often regarded as the first enzyme in the citric acid cycle (Patton et al. 1993). It is also reported that the crystal structure of the closed form of citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus* is 1.9 Å (Russell et al. 1997).

The MotA/TolQ/ExbB proton channel family protein is mainly localized in the membrane and functions in intracellular trafficking, secretion, and vesicular transport using the motive force at the inner membrane (Jarosik and Hansen 1995; Sun and Webster 1987). Studies have also shown that the Ton subcomplex (including motA/TolQ/ExbB) forms pH-sensitive cation-selective channels, providing insight into the mechanism by which it may harness the proton motive force to produce energy (Celia et al. 2016).

Convicilins, a set of seed storage proteins, differ from vicilins, a related group of seed storage proteins, mainly because of the presence of the N-terminal extension, an additional sequence of amino acids in the sequence that corresponds to the first exon. Convicilins have been described only in species of the legume tribe Viciaeae. Legumin A precursor is a seed storage globulin found in *Vicia narbonensis* and provides the major nitrogen source for the developing plant (Tandang-Silvas et al. 2010).

Putative sucrose-binding protein has been described as a sucrose-binding and sucrose-transporting protein as well as an outgroup member of the large vicilin storage protein family (Hei et al. 2001).

The LEGB7_VICFA protein, found in the seeds of many leguminous and non-leguminous plants, is an important source of sulfur-containing amino acids in seed meals. The

LEGB7_VICFA protein belongs to the 11S seed storage protein (globulin) family (Heim et al. 1989).

It is hypothesized that the Htlv-1 Gp21 ectodomain MALTOSE-binding protein chimera possesses dual roles of conserved fusion function and an ability to anchor diverse surface-exposed subunit structures to the virion envelope and infected cell surface (Kobe et al. 1999; Center et al. 1998).

Discussion and conclusion

Seed storage proteins are vital for plant growth and development as they (1) provide nutrients for seed respiratory metabolism, seed germination, and seedling establishment; (2) regulate physiological and biochemical reactions and metabolic processes. Studies show that the seed storage proteins of faba bean mainly consist of albumin and globulin (Utsumi et al. 1980; Bewley 1997; Li et al. 2003; Liu et al. 2012a). In this study, 16 proteins (Four GroEL molecular chaperones and 12 plant specific proteins) were identified. The convicilin, legumin A precursor, putative sucrose-binding protein, and LEGB7_VICFC are all globulins and specifically found within leguminous plants. Meanwhile, the convicilin in the 64-kD subunit, legumin A precursor in the 42- and 47-kD subunits, putative sucrose-binding protein in the 47-kD subunit, and LEGB7_VICFA in the 38-kD subunit were found at a relatively high abundance of 91.6, 86.7, 69.2, 30.8, and 89.1% respectively, whereas the abundance of other proteins ranged from 4.6 to 12.3% except GroEL chaperonin and phosphate ABC transporter periplasmic substrate-binding protein. This suggests that in faba bean seeds, globulin occupies a dominate part of seed storage proteins and provides the main nutrition and amino acid supply.

It is interesting that the relative abundance of GroEL chaperones in the 97-kD subunit reached up to 100%. This suggests that the GroEL chaperones existed as a much high amount in faba bean. It is reported that HSP60 has a 15–30% contribution in cellular protein folding and maintenance and also plays a vital role in mitochondrial protein transportation (Trivedi et al. 2016). In addition, the elevated expression level of HSP/molecular chaperones assist in the assembly of the highly abundant secretory storage proteins in the endoplasmic reticulum lumen during seed development (Trivedi et al. 2016). Taken together, that may be why there is a high abundance of GroEL chaperones in seed storage proteins.

The seed storage proteins in faba bean exhibited various functions depending on their diverse three-dimensional structures. In addition to those mentioned above, the proteins with a relatively low abundance possessed different

functions such as protein transport, signal transduction, protein biosynthesis, electron transfer, peroxide resistance, catalyzing, and intracellular trafficking in order to maintain the physiological and biochemical activities in faba bean seeds. Together, protein composition and abundance is closely associated with protein function in faba bean seeds.

Amino acid composition, as the assessment standard for protein quality, often varies between different seed storage proteins. The 16 proteins identified in faba bean were mainly composed of glutamic, arginine, lysine, glycine, leucine, cysteine, and methionine, which are essential to human life. In this study, the number of sulfur-containing amino acids (cysteine and methionine), which may limit the nutritional quality of the legume seed, reached up to 15, 19, and 15 in elongation factor Tu, citrate synthase, and putative sucrose-binding protein, respectively. The percentages of sulfur-containing amino acid were 3.9, 4.4, and 3.2% out of the total number of amino acids in elongation factor Tu, citrate synthase, and putative sucrose-binding protein, respectively, which was higher than the percentages in the other proteins. This result suggests that generation of the ideal faba bean using traditional breeding methods or genetic engineering may need to increase the cysteine and methionine levels in seed storage proteins. In the future, we can purposefully promote the amount of the elongation factor Tu, citrate synthase, and putative sucrose-binding proteins which contained a relative high amount of sulfur-containing amino acids to optimize the protein quality of faba bean seeds.

Acknowledgements This work was supported by China Agriculture Research System (CARS-09), National Natural Science Foundation of China (31460377) and the National Basic Research Program of China (973 Program, 2010CB134408).

Compliance with ethical standards

Conflict of interest Authors declare there is no conflict of interest.

References

- Bailey CJ, Boulter D (1970) The structure of legumin, a storage protein of broad bean (*Vicia faba*) seed. *Eur J Biochem* 17(3):460–466
- Bartasaghi A, Lecumberry F, Sapiro G, Subramaniam S (2012) Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure* 20(12):2003–2013
- Bewley JD (1997) Seed germination and dormancy. *Plant Cell* 9(7):1055–1066
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42 (Web Server issue):29
- Carbonaro M, Grant G, Cappelloni M, Pusztai A (2000) Perspectives into factors limiting in vivo digestion of legume proteins: antinutritional compounds or storage proteins? *J Agric Food Chem* 48(3):742–749
- Carlin NI, Lofdahl S, Magnusson M (1992) Monoclonal antibodies specific for elongation factor Tu and complete nucleotide sequence of the *tuf* gene in *Mycobacterium tuberculosis*. *Infect Immun* 60(8):3136–3142
- Cazzato E, Tufarelli V (2012) Quality, yield and nitrogen fixation of faba bean seeds as affected by sulphur fertilization. *Acta Agric Scand Sect B* 62(8):732–738
- Celia H, Noinaj N, Zakharov SD, Bordignon E, Botos I, Santamaria M, Barnard TJ, Cramer WA, Lloubes R, Buchanan SK (2016) Structural insight into the role of the Ton complex in energy transduction. *Nature* 538(7623):60–65
- Center RJ, Kobe B, Wilson KA, Teh T, Howlett GJ, Kemp BE, Pombourios P (1998) Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. *Protein Sci* 7(7):1612–1619
- Chandra-Hioe MV, Wong CH, Arcot J (2016) The potential use of fermented chickpea and faba bean flour as food ingredients. *Plant Foods Hum Nutr* 71(1):90–95
- Chen J (2009) Proteomics in agriculture biology. China science publishing & media Lid (CSPM)
- Chen L, Xie QW, Nathan C (1998) Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol Cell* 1(6):795–805
- Choi SB, Wang C, Muench DG, Ozawa K, Franceschi VR, Wu Y, Okita TW (2000) Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature* 407(6805):765–767
- Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5(4):823–826
- Dale T, Sanderson LE, Uhlenbeck OC (2004) The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry* 43(20):6159–6166
- Daviter T, Wieden HJ, Rodnina MV (2003) Essential role of histidine 84 in elongation factor Tu for the chemical step of GTP hydrolysis on the ribosome. *J Mol Biol* 332(3):689–699
- Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312(5771):212–217
- Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H (2015) Structure of the *E. coli* ribosome-EF-Tu complex at $<3 \text{ \AA}$ resolution by Cs-corrected cryo-EM. *Nature* 520(7548):567–570
- Gatehouse JA, Lycett GW, Delauney AJ, Croy RR, Boulter D (1983) Sequence specificity of the post-translational proteolytic cleavage of vicilin, a seed storage protein of pea (*Pisum sativum* L.). *Biochem J* 212(2):427–432
- Gerike U, Hough DW, Russell NJ, Dyal-Smith ML, Danson MJ (1998) Citrate synthase and 2-methylcitrate synthase: structural, functional and evolutionary relationships. *Microbiology* 144(Pt 4):929–935
- Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, Koziol JA, Schnitzer JE (2010) Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* 28(1):83–89
- Hei U, Wang Q, Kurz T, Borisjuk L, Golombek S, Neubohn B, Adler K, Gahrz M, Sauer N, Weber H, Wob U (2001) Expression patterns and subcellular localization of a 52 kDa sucrose-binding protein homologue of *Vicia faba* (VfSBPL) suggest different functions during development. *Plant Mol Biol* 47(4):461–474
- Heim U, Schubert R, Baumlein H, Wobus U (1989) The legumin gene family: structure and evolutionary implications of *Vicia faba* B-type genes and pseudogenes. *Plant Mol Biol* 13(6):653–663
- Hou WW, Zhang XJ, Shi JB, Liu YJ (2015) Genetic diversity analysis of faba bean (*Vicia faba* L.) germplasms using sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Genet Mol Res* 14(4):13945–13953

- Illergard K, Ardell DH, Elofsson A (2009) Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins* 77(3):499–508
- Jacobson FS, Morgan RW, Christman MF, Ames BN (1989) An alkyl hydroperoxide reductase from *Salmonella typhimurium* involved in the defense of DNA against oxidative damage. Purification and properties. *J Biol Chem* 264(3):1488–1496
- Jarosik GP, Hansen EJ (1995) Cloning and sequencing of the *Haemophilus influenzae* exbB and exbD genes. *Gene* 152(1):89–92
- Kim SI, Kim HY, Kwak JH, Kwon SH, Lee SY (2000) Cloning, overexpression and purification of *Bacillus subtilis* elongation factor Tu in *Escherichia coli*. *Mol Cells* 10(1):102–107
- Kobe B, Center RJ, Kemp BE, Pountourios P (1999) Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proc Natl Acad Sci USA* 96(8):4319–4324
- Kumamaru T, Satoh H, Iwata N, Omura T, Ogawa M, Tanaka K (1988) Mutants for rice storage proteins: 1. screening of mutants for rice storage proteins of protein bodies in the starchy endosperm. *Theor Appl Genet* 76(1):11–16
- Li X, Miao X, Qiu A (2003) Comparison of protein composition and structure in several legume species. *Cereals Oils* 6:19–20
- Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17(7):676–682
- Liu Y, Hou W, Li P, Shi J (2012a) Analysis on protein composition and subunit of albumin and globulin in faba beans of different genotype by SDS-PAGE in Qinghai Province. *J Gansu Agric Univ* 2(47):68–71
- Liu Y, Hou W, Shi J (2012b) Identification of special gene and analysis of the protein-subunit for different genotypes of Faba bean. *Acta Bot Boreali-occidentalia Sin* 32(1):0054–0059
- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50(3):437–450
- Makri E, Papalamprou E, Doxastakis G (2005) Study of functional properties of seed storage proteins from indigenous European legume crops (lupin, pea, broad bean) in admixture with polysaccharides. *Food Hydrocoll* 19(3):583–594
- Mandal S, Mandal R (2000) Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. *Curr Sci* 79(5):576–589
- Meinke DW, Chen J, Beachy RN (1981) Expression of storage-protein genes during soybean seed development. *Planta* 153(2):130–139
- Neznansky A, Blus-Kadosh I, Yerushalmi G, Banin E, Opatowsky Y (2014) The *Pseudomonas aeruginosa* phosphate transport protein PstS plays a phosphate-independent role in biofilm formation. *Faseb J* 28(12):5223–5233
- Patton AJ, Hough DW, Townner P, Danson MJ (1993) Does *Escherichia coli* possess a second citrate synthase gene? *Eur J Biochem* 214(1):75–81
- Roberts DL, Frerman FE, Kim JJ (1996) Three-dimensional structure of human electron transfer flavoprotein to 2.1-Å resolution. *Proc Natl Acad Sci USA* 93(25):14355–14360
- Rusanganwa E, Gupta RS (1993) Cloning and characterization of multiple groEL chaperonin-encoding genes in *Rhizobium meliloti*. *Gene* 126(1):67–75
- Russell RJ, Ferguson JM, Hough DW, Danson MJ, Taylor GL (1997) The crystal structure of citrate synthase from the hyperthermophilic archaeon *pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry* 36(33):9983–9994
- Scott MP, Jung R, Muntz K, Nielsen NC (1992) A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean. *Proc Natl Acad Sci USA* 89(2):658–662
- Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68(5):850–858
- Shewry PR, Halford NG (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. *J Exp Bot* 53(370):947–958
- Shewry PR, Napier JA, Tatham AS (1995) Seed storage proteins: structures and biosynthesis. *Plant Cell* 7(7):945–956
- Shi J, Hou W, Liu Y, Ma X (2012) Albumin genetic diversity analysis of Faba bean (*Vicia faba* L.) germplasms. *J Plant Genet Resour* 13(1):22–28
- Shi J, Hou W, Liu Y (2013) Comparative study on seed storage protein of Faba bean. *J Plant Genet Resour* 13(2):304–307
- Shimada T, Fuji K, Tamura K, Kondo M, Nishimura M, Hara-Nishimura I (2003) Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 100(26):16095–16100
- Smith DL, Deng Y, Zhang Z (1997) Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J Mass Spectrom* 32(2):135–146
- Sosulski F, McCurdy A (1987) Functionality of flours, protein fractions and isolates from field peas and FAA bean. *Food Sci* 52(4):1010–1014
- Sun TP, Webster RE (1987) Nucleotide sequence of a gene cluster involved in entry of E colicins and single-stranded DNA of infecting filamentous bacteriophages into *Escherichia coli*. *J Bacteriol* 169(6):2667–2674
- Tandang-Silvas MR, Fukuda T, Fukuda C, Prak K, Cabanos C, Kimura A, Itoh T, Mikami B, Utsumi S, Maruyama N (2010) Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochim Biophys Acta* 7(42):6
- Taylor WR, Orengo CA (1989) Protein structure alignment. *J Mol Biol* 208(1):1–22
- Textor S, Wendisch VF, De Graaf AA, Muller U, Linder MI, Linder D, Buckel W (1997) Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. *Arch Microbiol* 168(5):428–436
- Toogood HS, van Thiel A, Scrutton NS, Leys D (2005) Stabilization of non-productive conformations underpins rapid electron transfer to electron-transferring flavoprotein. *J Biol Chem* 280(34):30361–30366
- Trivedi D, Huda K, Gill S (2016) Molecular chaperone: structure, function, and role in plant abiotic stress tolerance. *Abiotic Stress Response Plants*:131–134
- Tsai MH, Saier MH Jr (1995) Phylogenetic characterization of the ubiquitous electron transfer flavoprotein families ETF-alpha and ETF-beta. *Res Microbiol* 146(5):397–404
- Utsumi S, Yokoyama Z, Mori T (1980) Comparative studies of subunit composition of *Vicia faba* L. seed. *Agric Biol Chem* 44(3):595–601
- Valizadeh M (2001) Seed storage protein profile of grain legumes grown in Iran, Using SDS-PAGE. *J Agric Sci Technol* 3:287–292
- Walter S (2002) Structure and function of the GroE chaperone. *Cell Mol Life Sci* 59(10):1589–1597
- Wang W, Xia M, Chen J, Deng F, Yuan R, Zhang X, Shen F (2016) Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in *Gossypium raimondii* and *G. arboreum*. *Data Brief* 9:345–348
- Webb A, Cottage A, Wood T, Khamassi K, Hobbs D, Gostkiewicz K, White M, Khazaei H, Ali M, Street D, Duc G, Stoddard FL, Maalouf F, Ogbonnaya FC, Link W, Thomas J, O'Sullivan DM

- (2016) A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnol J* 14(1):177–185
- Wu X, Ebine K, Ueda T, Qiu QS (2016) AtNHX5 and AtNHX6 are required for the subcellular localization of the SNARE complex that mediates the trafficking of seed storage proteins in arabidopsis. *PLoS ONE* 11(3):e0151658
- Xia K, Pittelli S, Church J, Colon W (2016) Kinetic stability of proteins in beans and peas: implications for protein digestibility, seed germination, and plant adaptation. *J Agric Food Chem* 64(40):7649–7657
- Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci* 13(5):1402–1406
- Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. *Proteins* 64(3):643–651