Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Phytochemistry 63 (2003) 1–5
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Amino acid sequence of the 26 kDa subunit of legumin-type seed storage protein of common buckwheat (*Fagopyrum esculentum* Moench): molecular characterization and phylogenetic analysis

Sangeeta Bharali, Nikhil K. Chrungoo*

Molecular Plant Physiology Laboratory, Department of Botany, North Eastern Hill University Shillong 793 022, India

Received 20 September 2002; received in revised form 17 December 2002

Abstract

The paper describes the amino acid sequence of a 26 kDa basic subunit of 13S globulin of common buckwheat (*Fagopyrum esculentum* Moench). The protein has 93 and 75% sequence homology with 11S globulin of *Coffea arabica* and β subunit of 11S globulin of *Cucurbita pepo* respectively. The subunit has the ''globally conserved'' N-terminal sequence consisting of Gly-Ile-Asp-Glu and the cysteine at P7′ from the proteolytic processing site. A conserved 7 residue domain of Pro-His-Trp-Asn-Ile-Asn-Ala, characteristic of basic subunits of legumins from non-leguminous angiosperms, is also present in this protein. A distinguishing features of this subunit is the relatively high level of lysine and methionine.

Keywords: *Fagopyrum esculentum*; Buckwheat; Lysine; Methionine; Legumin; Basic subunit; Phylogenetic analysis

1. Introduction

Although cereal grains and legume seeds are the main source of dietary proteins for human nutrition, the storage proteins in both are deficient to varying degrees in several of the essential amino acids. The major amino acid deficiency in legume seed proteins is their low content of sulphur amino acids, especially methionine. On the other hand cereal proteins are deficient in lysine (Shotwell And Larkins, 1989). Over the years many attempts have been made to improve the level of essential amino acids in seed storage proteins of crop plants through conventional breeding programmes (Larkins, 1983). However, in most cases the attempts have either led to a severe depletion in the storage protein levels or abnormalities in seed development. A variety of barley *viz*. Riso1508, developed in this way, had very high lysine content but a severe depletion of hordein level and grain yield (Herman and Larkins, 1991). The negative correlation between the seed protein content and the level of essential amino acids per unit protein has come as a major handicap in improving the amino acid composition of seed proteins in crops.

Because of inherent limitations in interspecific hybridizations, molecular approaches have provided alternative strategies to conventional breeding programmes. Such approaches have focussed either around the manipulation of the primary sequence of the proteins (Lago et al., 1990) or expression of genes for heterologous proteins rich in essential amino acids (Guerche et al., 1990; Saalbach et al., 1995). The generality of the approach by which the foreign proteins, rich in desired amino acids, may be introduced is constrained by scanty information about suitable heterologous proteins and their genes. The identification and characterization of the target proteins and cloning of full length genes coding for such proteins would be an essential prerequisite for production of transgenic plants with improved and balanced amino acid composition.

Amongst the existing plant resources, minor crops such as common buckwheat (*Fagopyrum esculentum* Moench) assume importance because of short growth span and the high nutritive value of its grains (Pomeranz and Robbins, 1992; Rout and Chrungoo, 1996). Rout et al. (1997) have already reported about the purification and N-terminal amino acid sequence of the

* Corresponding author. Fax: +91-364-550-108.
E-mail address: nchrungoo@nehu.ac.in (N.K. Chrungoo).

0031-9422/03/$ - see front matter
doi:10.1016/S0031-9422(02)00755-0

lysine rich basic subunit of the 13S globulin from grains of common buckwheat. The present paper describes the amino acid sequence of the protein and its phylogenetic relationship with similar proteins of other plants.

2. Materials and methods

2.1. Plant material

Grains of common buckwheat (*Fagopyrum esculentum* Moench var BDS-1354)) were procured from the Vivekananda Laboratory for Hill Agriculture, Almora (India).

2.2. Protein isolation

Total soluble proteins were extracted from the grains following extraction with cold 50 mM Tris-Cl buffer (pH 8.0) containing 0.1 M NaCl, 5% polyvinyl pyrrolidone and 2 mM phenyl methyl sulphonyl fluoride (PMSF). The globulin fraction representing the main storage protein was isolated as per the protocol described earlier (Rout and Chrungoo, 1996). Briefly the homogenate was centrifuged at 10,000 *g* for 15 min at 4 °C. The supernatant was brought to 80% saturation level of $(NH_4)_2SO_4$; the precipitate was dissolved in 50mM Tris-Cl buffer (pH 8.0) containing 100 mM EDTA and dialysed against the same buffer for 96 h at 4 °C. The 13S globulin was isolated from the dialysate by chromatography on Sepharose 6B. The isolated globulin fraction was reduced and alkylated with 0.1 M 2-ME in 6 M Urea according to Hager et al. (1992). The alkylated fraction was fractionated on Sephadex G-120 to isolate the 26 kD basic subunit.

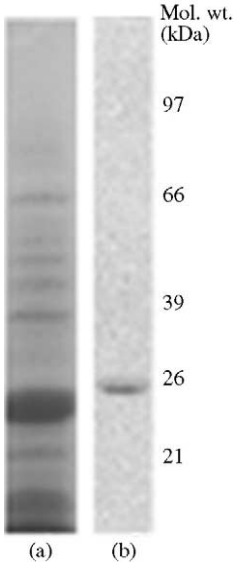

Fig. 1. SDS–PAGE profile of the crude extract (a) and the purified 26 kDa basic subunit of legumin type protein (b) from grains of common buckwheat.

2.3. Amino acid sequencing

The intact *S*-alkylated basic subunit was sequenced by Edman gas phase degradation using PSQ-1 Shimadzu gas phase sequenator coupled to PTH-amino acid analyser. Samples of the *S*-alkylated protein were digested separately with trypsin and CnBr as per Kamp (1986). The peptide mixtures were fractionated on BioGel-P6 column in 0.1 M ammonium bicarbonate (pH 8.0). Peak fractions were pooled and subjected to RP-HPLC (Vydac C-18 column) with gradient of acetonitrile in 0.1% aqueous TFA. The derived peptides were sequenced manually by the DABITC/PITC double coupling method (Chang et al., 1978).

2.4. Sequence analysis

The CLUSTAL package (Thompson et al., 1994) was used to align the amino acid sequence with amino acid sequences of basic subunits of other known legumins. The statistical significance of amino acid similarity was calculated for 200 randomisations according to Dayhoff (1978).

3. Results and discussion

The legumin-like polypeptides represent an abundant protein fraction as detected by SDS–PAGE (Fig. 1a). The polypeptides could be grouped into three size classes viz. α-β showing a molecular mass of 54 kDa, α with

Table 1
Amino acid composition of the 26 kDa basic subunit of buckwheat grain legumin

Amino acid	Residues (mol%)	Residues (mol^{-1})
Ala	9.27	18
Arg	6.18	12
Asx	10.30	20
Glx	16.49	32
Cys	0.51	1
Gly	6.70	13
His	1.03	2
Ile	5.67	11
Leu	7.21	14
Lys	6.70	13
Met	1.54	3
Phe	4.63	9
Pro	3.60	7
Ser	6.70	13
Thr	2.57	5
Trp	0.51	1
Tyr	3.60	7
Val	6.70	13

molecular mass ranging from 32-37 kDa and β with molecular weights ranging from 20 to 29 kDa. The purified 26 kD subunit resolved as a single band under denaturing SDS–PAGE (Fig. 1b).

The amino acid composition of the sununit reveals relatively high levels of lysine, threonine, leucine and methionine (Table 1). This feature indicates the nutritionally rich amino acid composition of buckwheat legumin. The protein also has high content of arginine and glutamine/glutamic acid. Presence of high levels of amides (glutamic acid, aspartic acid-asparagine and arginine), as observed in buckwheat legumin, is consistent with the role of crystalloid seed storage proteins as a nitrogen source during seed germination. Fischer et al. (1996) have reported the presence of methionine rich (MetR) legumins in lower angiosperms. Amongst the seed crops, methionine rich legumin type protein has been reported from *Perilla frutescens* (Jin et al., 2000). The identification of a legumin-type protein with a relatively higher level of methionine in common buckwheat indicates the presence of such proteins in seed crops also.

The amino acid sequence of the protein was established after microsequencing of the relevant peptides obtained after digestions of the reduced and alkylated basic subunit with trypsin and or following chemical cleavage with CnBr (Fig. 2). The sequence has been deposited in the SWISS-PROT protein data bank with accession no. P83004. Using an alignment that permitted maximum homology, the buckwheat protein showed the highest (>90%) homology with 11S storage protein from *Coffea arabica*. The percentage homology with other legumin type proteins varied between 68% for legumin1 from common buckwheat to 76% for legumin type protein from *Magnolia salicifolia*. A statistical evaluation of the alignments revealed that the

```
                                                                        60
Bwleg (P83004)           gidenvctmklrenikspqeadfynpkagrittansqklpalrslqmsaergflysngiy
11Sglbcoffea (ACC61983) gleetlctvklseniglpqeadvfnpragrittvnsqkipilsslqlsaergflysnaif
Prunin (CAA55009)        gleetfcslrlkenignperadifspragristlnshnlpilrflrlsaergffyrngiy
Magnolia (S54206)        gleeiqcsskltyniadptradvynpqagritslnsqklpilnvlqlsaergvlyrnall
Citrin (T10089)          gfeetictmklrhnidkpshadvynpragrvttvnrfnlpilrdlqlsaekgnlypnall
Ricinus (AAF73008)       gleetfctrrlrhninkpseadiynpragrvtsvnshnlpilrylqlsiqkavlyknalm
11Sglbcucumber (P13744) gleetictlrlkqnigrsvradvfnprggristanyhtlpilrqvrlsaergvlysnamv
Fagopyrum leg1(T10696)   gleqafcnlkfkqnvnrpsradvfnpragrintvnsnnlpilefiqlsaqhvvlyknail
glycinin (BAA74953)      gveenictlklheniarpsradfynpkagristlnsltlpalrqfqlsaqyvvlykngiy

                                                                       120
Bwleg (P83004)           aphwninahsalyvtrgnakvqvvgdegnkvfddevkqgqliivpqyfavikkagnqgfe
11Sglbcoffea (ACC61983) aphwninahsalyvirgnariqvvdhkgnkvfddevkqgqliivpqyfavikkagnqgfe
Prunin (CAA55009)        sphwnvnahsvvyvirgnarvqvvnengdaildqevqqgqlfivpqnhgviqqagnqgfe
Magnolia (S54206)        apqwnvnahslvyatrgngrvqivgeqgrpvfdgelregqlvvvpqsfavvkkagnegfe
Citrin (T10089)          apqwnlnahsivyvtrgngrmqivaengenvfdgqiregqlivvpqgfavvkragnrgle
Ricinus (AAF73008)       tphwninahsiryitrgsgrvqivnengdsvfdgqvqrgqmftvpqnfvvitkasnegle
11Sglbcucumber (P13744) aphytvnshsvmyatrgnarvqvvdnfgqsvfdgevregqvlmipqnfvvikrasdrgfe
Fagopyrum leg1(T10696)   gprwnlnahsalyvtrgegrvqvvgdegrsvfddnvqrgqilvvpqgfavvlkagregle
glycinin (BAA74953)      sphwnlnansviyvtrgqgkvrvvncqgnavfdgelrrgqllvvpqnfvvaeqageqgfe
                          *******                                       *

                                                                       180
Bwleg (P83004)           yvafktndnaminplvgrlsafraipeevlrssfqisseeaeelkygrqealllseqsqq
11Sglbcoffea (ACC61983) yvafktndnaminplvgrlsafraipeevlrssfqisseeaeelkygrqealllseqsqq
Prunin (CAA55009)        yfafkteenafintlagrtsflralpdevlanayqisreqarqlkynrqetialsssqqr
Magnolia (S54206)        yvafktndnamnsplvgktsvirampedvlinsyrisreearrlkynreeiavfaprfss
Citrin (T10089)          wisfktndvamtsqlagrasvlrglpldviqnsfqvsrdeaqrlkynrqeltvftpgprs
Ricinus (AAF73008)       wvsfktndnakinqlagrvsairsmpeevvanafqvsvedarrlkdnrqevtllspgsrs
11Sglbcucumber (P13744) wiafktndnaitnllagrvsqmrmlplgvlsnmyrisreeaqrlkygqqemrvlspgrsq
Fagopyrum leg1(T10696)   wvelknddnaitspiagktsvlraipvevlansydistkeafrlkngrqevevflpfqsr
glycinin (BAA74953)      yivfkthhnavtsyl---kdvfraipsevlahsynlrqsqvselkyegnwgplvnpesqq

Bwleg (P83004)           gkrevadekererf
11Sglbcoffea (ACC61983) gkreva--------
Prunin (CAA55009)        ravv----------
Magnolia (S54206)        qqraaa--------
Citrin (T10089)          qwgl------tva-
Ricinus (AAF73008)       t-------------
11Sglbcucumber (P13744) grre----------
Fagopyrum leg1(T10696)   deke------rerf
glycinin (BAA74953)      gsprvkva------
```

Fig. 2. Amino acid sequence of the basic subunit of legumin-type protein (Bwleg, Swissprot accession No. P83004) from grains of common buckwheat and its alignment with amino acid sequences of basic subunits of other legumin-type proteins. The invariant residues are shown in bold. The "*" mark represents a conserved domain of invariant/ similar residues. [Text in parentheses indicates the accession No. of the sequence; arrow at the start of the sequence marks the N-terminal glycine.]

sequence homologies were highly significant. Although Fujina et al. (2001) have reported the presence of allergenic proteins in buckwheat grains, the buckwheat legumin isolated by us did not show any significant homology with amino acid sequences of known allergens.

The sequence reveals all the characteristic features of legumin subfamily. These include the N-terminal domain composed of Gly-Ile-Asp-Glu and the cysteine at P7′ from the proteolytic processing site. The ''globally conserved'' position of P7′ Cysteine indicates its significance in the secondary structure formation. Hager et al. (1992) have indicated the importance of this residue in formation of a disulphide bridge linking the acidic and basic chains. The basic subunits of legumins from angiosperms show conserved domain consisting of Pro-His-Trp-Asn-Ile/Leu/Val-Asn-Ala between residue 64 and 70 from the N-terminus. This feature distinguishes these subunits of angiosperms from those of gymnosperms. The buckwheat legumin also has a similar conserved domain of Pro-His-Trp-Asn-Ile-Asn-Ala. A low complexity region was detected in the sequence between residues 169–180. This allows a prediction that any mutation or insertion in this region might not have a detrimental effect on the structure of the protein.

Phylogenetic analysis of amino acid sequences of legumin type proteins including the one from common buckwheat reported here reveals a clear diversification into angispermous and gymnospermous groups (Fig. 3).

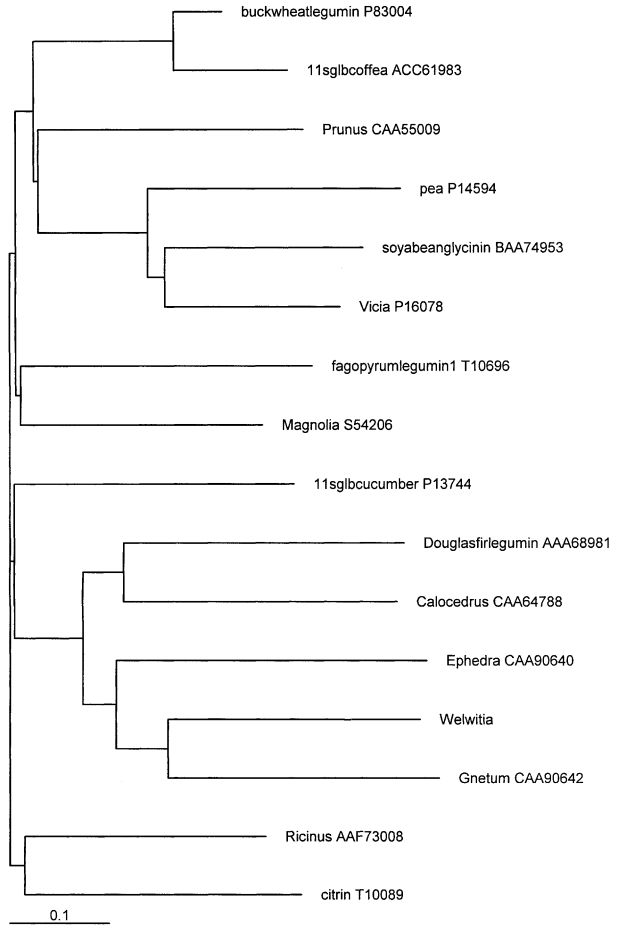

Fig. 3. Phylogenetic tree based on the amino acid sequences of the basic subunit of legumin type proteins of various angiosperms and gymnosperms. The text in the figure refers to accession number of the sequence.

Acknowledgements

Part of this work was carried out with financial assistance from DST, New Delhi under project No. SP/SO/A-41.

References

Chang, J.Y., Brauer, D., Wittman-Liebold, B., 1978. Microsequence analysis of peptides and proteins using 4-N dimethylaminobenzene 4′-isothiocyanate/phenylisothiocyanate double coupling method. FEBS Lett. 93, 205–214.

Dayhoff, M.O., 1978. A model of evolutionary change in protein matrices for detecting distance relationships. In: Dayhoff, M.O. (Ed.), Atlas of Protein Sequences and Structures, Vol. 5, suppl. 3. National Biochemical Research Foundation, Washington DC, pp. 345–358.

Fisher, H., Chen, L., Wallisch, S., 1996. The evolution of angiosperm seed protein: A methionine-rich legumin subfamily present in lower angiosperm clades. J. Mol. Evol 43, 399–404.

Fujina, K., Funatsuki, H., Inada, M., Shimono, Y., Kikuta, Y., 2001. Expression, cloning and immunological analysis of buckwheat (*Fagopyrum esculentum* Moench) seed storage proteins. J. Agric. Food Chem. 49, 1825–1829.

Guerche, P., DeAlmeida, E., Schwarztein, M.A., Gander, E., Krebbers, E., Pelletier, G., 1990. Expression of the 2S albumin from *Bertholletia excelsa* in *Brassica napus*. Mol. Gen. Genet. 221, 306–314.

Hager, K.P., Jensen, U.W.E., Gilroy, J., Richardson, M., 1992. The N-terminal amino acid sequence of the β-subunit of the legumin like protein from seeds of *Ginkgo biloba*. Phytochemistry 31, 523–525.

Herman, R.G., Larkins, B., 1991. Plant Molecular Biology. Plenum, New York.

Jin, U.H., Jin, B.R., Lee, J.W., Cho, Y.S., Kwon, O.C., Kim, Y.K., Chung, C.W., 2000. Characterization of methionine rich storage protein cDNA from Perilla (*Perilla frutescens*). Aust. J. Pl. Physiol. 27, 701–707.

Kamp, R.M., 1986. Separation of peptides. In: Wittman-Liebold, B., Salnikov, J., Erdman, V.A. (Eds.), Advanced Methods in Protein Microsequence Analysis. Springer-Verlag, Berlin, pp. 8–20.

Lago, W.J.P., Scott, M.P., Nielsen, N.C., 1990. Assembly properties of modified subunits in the glycinin subunit family. In: Herman, R.G., Larkins, B.A. (Eds.), Plant Molecular Biology II. Plenum, New York, pp. 635–640.

Larkins, B.A., 1983. Genetic Engineering of Plants: An Agricultural Perspective. Plenum, New York. pp. 92-118.

Pomeranz, Y., Robbins, G.S., 1992. Amino acid composition of buckwheat. J. Agric. Food Chem. 20, 270.

Rout, M.K., Chrungoo, N.K., 1996. Partial characterization of Lysine rich 13S globulin from buckwheat (*Fagopyrum esculentum* Moench): Its antigenic homology with seed proteins of some other crops. Biochem. Mol. Biol. Intl 40, 587–595.

Rout, M.K., Chrungoo, N.K., Rao, K.S., 1997. N-terminal amino acid sequence of the lysine rich subunit of 13S globulin from grains of common buckwheat (*Fagopyrum esculentum* Moench). Phytochemistry 45, 865–867.

Saalbach, I., Waddell, D., Pickardt, T., Schieder, O., Muntz, K., 1995. Stable expression of the sulphur rich 2S albumin gene in transgenic *Vicia narbonensis* increases the Methionine content of seeds. J. Plant. Physiol 145, 674–681.

Shotwell, M.A., Larkins, B.A. 1989. Molecular biology and biochemistry of seed storage proteins: In The Biochemistry of Plants. A comprehensive Treatise, Vol XV. Academic Press, New York, pp. 297-345.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting position specific gap penalties and weight matrix choice. Nucleic Acid Res. 22, 4673–4680.