ISSN 0006-2979, Biochemistry (Moscow), 2006, Vol. 71, No. 11, pp. 1183-1191. © Pleiades Publishing, Inc., 2006.
Published in Russian in Biokhimiya, 2006, Vol. 71, No. 11, pp. 1459-1468.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM06-009, September 10, 2006.

Cloning and Structural Analysis of an Indian Little Millet (*Panicum sumatrense*) Zein-Like Storage Protein: Implications for Molecular Assembly

S. Sivakumar[1]*, O. L. Franco[2], B. Thayumanavan[1], A. M. Murad[2], A. Manickam[1], M. Mohan[1], and M. Mridula[3]

[1]*Center for Plant Molecular Biology, Tamil Nadu Agricultural University, Coimbatore-641 003, India; fax: +91-422-551-1454; E-mail: subarayansivakumar@yahoo.com*
[2]*Centro de Analises Proteomicas e Bioquimicas, Pos-Graduacao em Ciencias Genomicas e Biotecnologia, Universidade Catolica de Brasilia, Brasilia – DF, Brazil; fax: +55-61-3347-4797; E-mail: ocfranco@pos.ucb.br*
[3]*Project Trainee, Bioinformatics Center, Indian Institute of Science, Bangalore, India*

Received January 18, 2006
Revision received May 30, 2006

Abstract—Zeins are prolamin storage proteins that accumulate in kernel endosperm of several cereals. For cloning of genes coding for zein-like proteins that accumulate in enhanced quantities in the filling stages of little millet (*Panicum sumatrense* Roth.) developing grains, RT-PCR was performed using specific primers. A 750-bp cDNA was directly sequenced and *in silico* analysis showed high identity degree to α-prolamins. This family is composed of zeins from *Zea mays*, coixins from *Coix lachryma-jobi*, and α-kafirins from *Sorghum bicolor*. The putative conserved domain of zein-like proteins was identified by primary structure comparisons. Furthermore, threading analyses indicated that the millet zein-like protein forms an antiparallel α-helical hairpin with two opposite surfaces: one hydrophobic and the other hydrophilic that probably could be involved in protein storage assembly. Knowledge about zein-like α-prolamins in little millet will lead to cloning and transfer of this gene to other major food crops, such as cereals and legumes, with inferior nutritional quality for monogastric animals.

DOI: 10.1134/S0006297906110034

Key words: homology searches, molecular modeling, Indian little millet, zein, phylogenetic relationships

The current world population, approximately 6.8 billion, is expected to double around 2050 [1]. The population increases significantly in several developing countries [2], and it is estimated that in approximately 40 years 90% of planet's population will reside in these countries located in the Southern hemisphere. A great challenge for the future, therefore, lies in global food security, since it will be necessary to double food production in the next 50 years to meet the dietary needs of the increasing population [1]. While demand for food is steadily increasing, biotic and abiotic factors cause considerable damage, including losses of nutritional compounds such as essential amino acids, minerals, and vitamins [3, 4].

In this field, with the goal of finding novel strategies to the food production problem, several highly nutritional seeds have been intensely studied. One of them, known as Indian little millet (*Panicum sumatrense*), has been showing great promise in this area. This Poaceae (Gramineae) is harvested in elevated areas (7000 ft and up), in both plains and hilly regions. Wild cultivars are found in Punjab, Burma, South East Asia, Central China, Ceylon, and subtropical countries, while domestic varieties are cultivated throughout India, Pakistan, and Ceylon. Indian little millet has a short crop cycle. It is mostly cultivated under natural rain-fed irrigation, growing even during relatively dry seasons, but also under artificial irrigation; it is suitable for both dry and waterlogged conditions. It also grows well in adverse environments such as salty and dry soils as well at high temperatures. Most of the improved

Abbreviations: CTAB) cetyltrimethyl ammonium bromide; DEPC) diethyl pyrocarbonate; MOPS) morpholino-1-propanesulfonic acid; RT-PCR) reverse transcription-polymerase chain reaction; SSC) saline sodium citrate; TAE) Tris-acetate EDTA buffer; TBE) Tris-borate EDTA buffer.

* To whom correspondence should be addressed.

cultivars are pure lines from land races located in these regions. Finally, millet grains are rich sources of protein (8.0%), fat (4.7%) carbohydrates (65.0%), and crude fibers (7.6%), apart from minerals and vitamins. These factors are also relevant to human health since they could be utilized in the control of several illnesses such as diabetes, obesity, and hyperlipidemia [5, 6]. Moreover, little millet grains are not attacked by insect pests during prolonged storage due to the presence of proteinaceous defense factors present in the endosperm of the grains [7]. High nutritional value and storage feasibility of little millet grain together have led to considering this kernel as an important food staple by ancient and modern people.

Since little millet has a protein rich kernel, several studies have been carried to identify the proteinaceous compounds. Amongst several plant storage proteins, zeins, coixins, and kafirins, pertaining to the prolamin family, are accumulated in the seed endosperm and compose four different types of polypeptides classified as α-, β-, γ-, and δ-zeins [8]. Accretions of prolamins form spherical protein bodies within the lumen of the endoplasmic reticulum. Collectively, prolamins are rich in glutamine and proline, but they lack lysine and tryptophan. These proteins constitute a large proportion of the total seed protein (60-70%), and the amino acid composition of prolamins is directly related to grain nutritional quality. Efforts to improve the protein quality of cereal kernels have been made, and this report deals with molecular cloning of a gene of a little millet zein-like protein pertaining to the prolamin family. Furthermore, to elucidate the molecular assembly of prolamins, a molecular model has been constructed and analyzed. The occurrence of essential amino acids and certain proteins are considered extremely vital for human growth and development, providing excellent scope for research on gene transfer to improve nutritional content of other food crops such as cereals and legumes.

MATERIALS AND METHODS

Sample collection and DNA isolation. Little millet seeds variety CO 3 were sown at the Millet Breeding Station, Tamil Nadu Agricultural University, Coimbatore, India at different intervals and plants were grown by adopting normal agronomic practices. After 15-20 days of post-anthesis, the developing grains at milky stage were collected in diethyl pyrocarbonate (DEPC)-treated polypropylene tubes in RNase free conditions and frozen immediately in liquid nitrogen. Frozen samples were stored at −70°C for RNA extraction. DNA was isolated from pot grown seedlings using CTAB standard extraction buffer (100 mM Tris-HCl, pH 8.0, containing 1.4 M NaCl, 20 mM EDTA, 2.0% (w/v) CTAB, and 0.1% (w/v) 2-mercaptoethanol) according to the method of Harris et al. [9]. Isolated DNA was dissolved in TE buffer for further PCR reactions. DNA quality was checked in 1.2% agarose gel for intact single band. DNA samples were quantified spectrophotometrically at 260 nm.

Custom synthesis of oligonucleotide for N-terminal primer. Nucleotide sequences for zein-like related sequences were collected from the NCBI database, particularly cereals and millets, to find typical codon usage. The most common codons were identified for each amino acid to design the 5-3′ oligonucleotide sequence. The identified oligonucleotide sequence was checked by Net primer software for oligomer synthesis. The forward primer 5′-GTAGACGAACCACAGAACCTTATAAAT-3′ and reverse oligo$(dT)_{23}$ were used for cDNA synthesis and amplification.

Standardization of PCR conditions. The Gradient PCR device (Eppendorf, Germany) was used for PCR amplification. The template genomic DNA (10 ng) was added to 1.5 µl (10×) PCR buffer, 1.0 µl 2.5 mM dNTPs mix, 1.0 µl containing 100 ng forward and reverse primer, and 0.3 unit *Taq* polymerase, and the volume was made up to 15 µl with distilled water. The annealing temperature of 48°C for the forward and reverse primer was optimized in the gradient PCR device from the tested temperatures ranging from 37 to 60°C.

Isolation of total RNA. A rapid RNA isolation method [10, 11] was followed with minor modification for total RNA extraction from milky-stage grains. This method envisages rapid isolation and minimizes degradation. Extraction buffer containing 100 mM LiCl, 100 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), and 1% SDS was used. The volume was made up with DEPC-treated water. Four grams of milky grains (stored at −70°C) were ground in an ice-cold mortar with liquid nitrogen, until a fine powder was obtained. The powder was transferred to a DEPC-treated polypropylene centrifuge tube. An equal volume of freshly prepared phenol saturated with TE buffer (pH 8.0) and extraction buffer (6.0 ml) were taken and heated at 80°C. Preheated phenol extraction buffer was added onto the sample immediately after grinding. The mixture was thoroughly suspended by inverting the tube several times following an addition of 2.0 ml chloroform. The contents were centrifuged at 10,000 rpm for 10 min and the supernatant was aliquoted into 1.5 ml DEPC-treated Eppendorf centrifuge tubes. An equal volume of 4.0 M LiCl was added to the supernatant to bring the solution at 2.0 M LiCl standard concentration in order to separate RNA from DNA. The contents were stored at −20°C overnight. The sample was then brought to room temperature for 5 min and then centrifuged at 10,000 rpm for 10 min. The pellet was washed in 70% ethanol followed by absolute ethanol washing and air-dried in a totally RNase-free environment. The dried pellet was resuspended in sterile water.

Formaldehyde agarose gel electrophoresis of RNA. The RNA sample was evaluated using 1.2% formaldehyde agarose gel [12]. The denaturing gel (1.2% agarose con-

taining 2.0% formaldehyde) was poured in 1× MOPS buffer (20 mM, pH 7.0) containing 5 mM sodium acetate and 1 mM EDTA. To the RNA sample, a loading buffer (a mixture of 0.72 ml formamide, 0.16 ml of MOPS buffer (10×), 0.26 ml formaldehyde, 0.1 ml of 80% glycerol, 0.18 ml of distilled water, and 0.08 ml of saturated bromophenol blue) was added. Samples were loaded into wells and electrophoresis was performed using 1× MOPS buffer at 8 V/cm. The gel was washed with 0.5 M ammonium acetate and with distilled water before staining with 0.5 μg/ml ethidium bromide.

Quantification of RNA. RNA was quantified spectrophotometrically at 260 nm. One absorbance unit corresponds to 40 μg of RNA. To check RNA purity, absorbance ratio at 260/280 nm (i.e., 1.8-2.0) was evaluated.

Isolation of mRNA. The isolation of mRNA was done through Message Maker Reagent Assembly (Life Technologies, USA) by following the manual. A totally RNase-free environment was maintained around the working bench top. Total RNA (0.5-9.0 ng) was transferred to a DEPC-treated 1.5-ml Eppendorf tube and RNA concentration adjusted to 0.55 mg/ml with sterile Milli-Q water. Total RNA was heated to 65°C for 5 min, chilled on ice, and osmotic concentration was adjusted to 0.5 M NaCl. Oligo(dT)-cellulose suspension (0.5 ml) was added to the RNA sample. The content was mixed well by inverting the tube several times and kept at incubator for 15 min at 37°C. Oligo(dT)-cellulose–RNA mixture was transferred to a filter syringe after removing the plunger on an RNase-free bench top. The plunger was reinserted and pushed to barrel bottom to elute unbound RNA. One and a half milliliter of wash buffer I (0.5 M NaCl, 20 mM Tris-HCl, pH 7.5) was transferred to a disposable beaker. Buffer was drawn into the syringe by gently pulling the plunger up the barrel and mixed well until the oligo(dT)-cellulose was fully resuspended. Pushing the plunger to the barrel bottom, all liquid was expelled. Again, the contents were washed with 1.5 ml wash buffer II (0.1 M NaCl, 20 mM Tris-HCl, pH 7.5) by following the above steps. The mRNA was finally eluted in distilled water (preheated to 65°C). The preheated water was drawn from a disposable beaker by pulling up the plunger and the contents were briefly mixed well. The mRNA was collected onto an RNase-free DEPC treated tube. Oligo(dT)-cellulose fine particles were sedimented by centrifugation at 4°C for 2 min at 3000 rpm, and the supernatant was carefully transferred to a RNase-free tube. The mRNA was quantified at 260 nm and precipitated with 50 μg/ml glycogen, 0.1 volume of 7.5 M ammonium acetate, and two volumes of cold ethanol (–20°C) followed by an overnight incubation at –20°C. Sample was centrifuged at 6000 rpm for 30 min at 4°C and the pellet was washed with 75% ethanol. The mRNAs were air-dried into an RNase-free environment and resuspended in the required volume of sterile Milli-Q water.

First strand cDNA synthesis. The enhanced avian HSRT-PCR$^+$ Kit (Sigma, USA) was used to convert mRNA transcripts into cDNA in DEPC-treated 0.5-ml PCR tube using 2 μl (100-250 ng) mRNA, 1 μl of 10 mM dNTPs, 1 μl (0.5 μg) anchored oligo(dT)$_{23}$, and water to bring total volume to 10 μl (PCR reagent). The contents were briefly centrifuged to collect all components at the bottom and the tube was incubated at 65°C for 10 min and chilled on ice. Two microliters of 10× AMV-RT buffer (500 mM Tris-HCl, pH 8.3, 400 mM KCl, 80 mM $MgCl_2$, 10 mM DTT), 1 μl of 20 U/μl RNase inhibitor, and 1.0 μl of 20 U/liter enhanced avian reverse transcriptase were added to the contents. Reaction was performed at 45°C for 60 min.

PCR amplification of target cDNA using gene-specific primer. First strand cDNA was amplified with anchored oligo(dT)$_{23}$ as reverse primer and forward gene specific primer. The standardized PCR condition for respective gene-specific primer was adopted for second strand amplification using an HSRT-PCR kit (Fig. 1). The 50 μl reaction volume contained 5 μl of Accu *Taq* buffer (10×), 1 μl of dNTPs (each 10 mM), 1.0 μl (0.5 μg) gene-specific primer, 1 μl (0.5 μg) anchored oligo(dT)$_{23}$ primer, 100 ng of template cDNA, 1.0 μl (25 units) of Jumpstart Accu *Taq* LA DNA polymerase mix and the volume was made up to 50 μl with water. The standardized thermal cycle program used an initial denaturing 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 1 min,

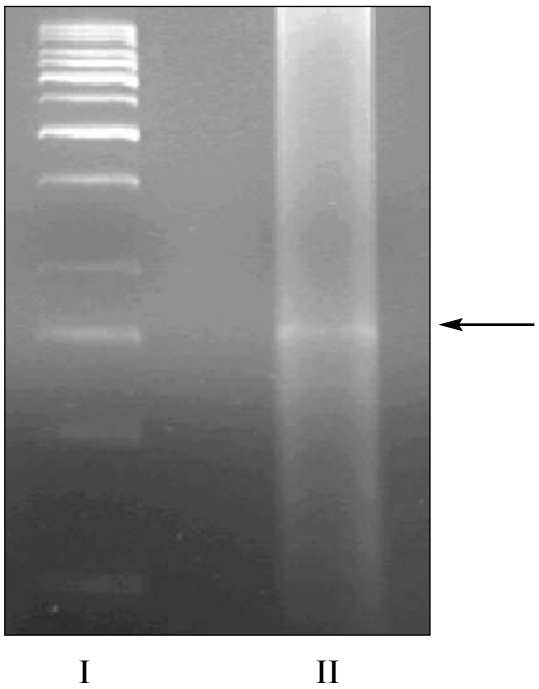

Fig. 1. Agarose gel of RT-PCR product from little millet cDNA (arrow). Lane I corresponds to molecular marker (1000 bp) and lane II corresponds to cloned cDNA (750 bp).

annealing temperature 48°C for 1 min, extension temperature of 55°C for 1 min, and final extension of 78°C for 7 min. After PCR reaction, one-fifth volume of mixture assay was used for agarose electrophoresis on TBE buffer (1×) following the standard procedure [11].

Agarose gel electrophoresis of DNA. The quality of DNA was analyzed on 1.2% agarose gel. Agarose was dissolved by heating using TBE buffer (90 mM Tris-borate and 2 mM EDTA, pH 8.0). The gel mixture was cooled to 50°C and 0.5 μg/ml ethidium bromide was added. The mixture was poured onto an alcohol-sterilized preset template with a comb. After solidification, the comb and sealing tapes were removed and the gel was mounted in an electrophoresis tank. For PCR products of genomic DNA and cDNA, the loading buffer (5×) was prepared with 50% glycerol and 0.25% bromophenol blue in TBE. Electrophoresis was performed at 8 V/cm using TBE as the running buffer. After running, bands were visualized on a UV trans-illuminator.

Elution of cDNA. The PCR product was resolved on 1.0% agarose gel in TAE buffer (1×) and visualized on the UV trans-illuminator. Target cDNA was carefully excised with a clean sharp scalpel. The trimmed gel piece containing cDNA was transferred to a 1.5-ml Eppendorf tube and eluted with a GenElute Gel extraction kit (Sigma). The gel piece was weighed and three gel volumes of solubilization buffer were added. Samples were kept in a water bath at 60°C for 10 min with frequent mixing. A GenElute mini preparation-binding column was made ready by adding 500 μl of column preparation solution. To dissolve the gel, one gel volume of isopropanol was added with and mixed thoroughly. The sample was loaded onto the binding column and centrifuged at 12,500 rpm for 1 min, and the column was washed with 600 μl of wash solution, incubated for 2-3 min, centrifuged for 1 min, and finally, centrifuged for an additional 2 min to remove traces of wash solution. The binding column was transferred to a fresh tube and 50 μl of elution buffer was added. The column was centrifuged for 1 min, and the cDNA collected was checked using agarose electrophoresis.

Northern blot analysis. To confirm cDNA expression in developing grains, a Northern hybridization analysis was performed [11]. Total RNA was isolated from milky stage grains of little millet as described earlier. The RNA sample was resolved on a 1.2% formaldehyde agarose gel [12]. After resolving on the denaturing gel, the formaldehyde was removed by soaking the gel in large volume of water three times for 30 min each time. The gel was then soaked in SSC (10×) for 30 min and placed carefully on a "transfer platform" covered with a Whatman 3 mm filter paper wick with its margins hanging in SSC (10×) in a glass tray. A nylon filter piece was marginally cut, placed carefully over the gel, covered immediately with five sheets of Whatman 3 filter paper, and stacked with crude filter paper cut exactly according to gel size. The stack of filter papers was covered by a glass plate, over which a weight of 500 g was placed. The level of 10× SSC solution in the tank was maintained below the platform. The upward diffusion of the buffer was allowed for 18 h. After transfer, the nylon filter was carefully removed, cross-linked with UV light for 50 sec, rinsed with 2× SSC, air-dried, and maintained at 4°C. The PCR products of little millet were also gel purified and used for labeling. About 25 μg of purified cDNA were taken and denatured in a boiling water bath for 5 min and chilled immediately on ice. Denatured cDNA was labeled using a random primer labeling kit (Bangalore Genei Pvt. Ltd, India). Reaction was performed with 50 μCi of [α-^{32}P]dCTP in place of dCTP and incubated at 37°C for 2 h. After the reaction, the sample was heat-denatured and introduced in the hybridization reaction as described by Sambrook et al. [11], and nylon filters with total RNA pattern were pre-hybridized in preparation solution (50% formamide, 5× SSC, 5× Denhardt's solution, 10% dextran sulfate, and denatured salmon sperm DNA (1 mg/ml)) at 42°C for 6 h. Hybridization solution (pre-hybridization solution + radiolabeled samples at standard concentration of 10^6 cpm/ml) were added to pre-hybridized filters at 42°C for 18 h. After that, the nylon filters were washed twice with 2× SSC + 0.1% SDS for 20 min each at room temperature. Then, the filters were dried at room temperature and exposed to Hindustan X-ray film for autoradiography.

Sequence analysis. The presence of the little millet cDNA fragment was confirmed by Northern hybridization and also by direct nucleotide sequencing by the Big Dye method (ABI Prism, Model 1000) using an automated fluorescent sequencer (DNA sequencing services; Microsynth, Switzerland). One single pass analysis identified part of 444 nucleotides of approximately 750 bp cDNA fragments (NCBI accession was AY 839104 gi: 58760522). Database searches for nucleotide sequence were carried out to localize homolog sequences and thereby infer protein functions. First, the sequence was submitted to BLAST X [13], BLAST P, and PSI-BLAST [14], which were used to localize distant sequence homologs in the sequence databases. Alignment manipulations were carried out with Jalview [15]. Searches for domains were performed with CDD [16, 17], Pfam [18], Prosite [19], Secondary structure prediction [20], Protoparm [21], and Phylogram [22]. Protein translation was performed with the translation tool at Expasy (http://ca.expasy.org/tools/dna.html).

Molecular modeling. For molecular modeling, little millet zein-like amino acid residue sequence was divided in two and these sequences were submitted to BioInfo Meta Server [23] to find the best three-dimensional structural template by threading analyses. 3D-PSSM alignments resulted in antifreeze protein isoform (PDB code: 1WFB) (59.2%) [24] and Rho-kinase protein (PDB code: 1UIX) (16.4%) [25] due to their best sequence analogy. A chimerical alignment of both structures against our sequence was manually adjusted and Modeller v8 [26] was

used for atomic coordinates transference. Energy minimization was performed in 2000 steps of steepest descent with temperature and pressure coupling in a water box using Gromacs packages [27]. GROMOS96 [28] was utilized as the force field. A molecular dynamics simulation was also performed setting configuration file for 90,000 steps at 300°K temperature and 1 bar of isotropic pressure couplings in the Berendsen algorithm to analyze model stability. For modeling and Gromacs simulations, we used a Sun 2.4 Ghz double processed opteron workstation taking about 4 h of process. Deep-View Swiss-PDB Viewer 3.7 [29] was used for protein visualization.

RESULTS AND DISCUSSION

Cloning and Northern blot analyses of little millet zein-like protein. The little millet zein-like protein gene contained 444 bp of the 750 bp open reading frame that encoded a predicted protein of 118 amino acid residues. The molecular mass of the predicted translated gene was 15,466 daltons, with a theoretical p*I* of 8.14 ([30], http://ca.expasy.org/tools/pi_tool.html). Furthermore, the YinOYang program (http://www.cbs.dtu.dk/services/YinOYang/) predicted nine possible O-glycosylation sites at threonine and serine residues as indicated further in Fig. 3 (bold residues). These residues showed a high conservation rate suggesting their importance for zein-like protein assembly. Additionally, to confirm expression, Northern hybridization [11] was carried out by using little millet cDNA probe against total RNA isolated from milky/developing grains. Northern hybridization signal confirmed the expression of zein-like transcripts in developing grains as also evidenced mRNA accumulation of zein-α precursors (Fig. 2).

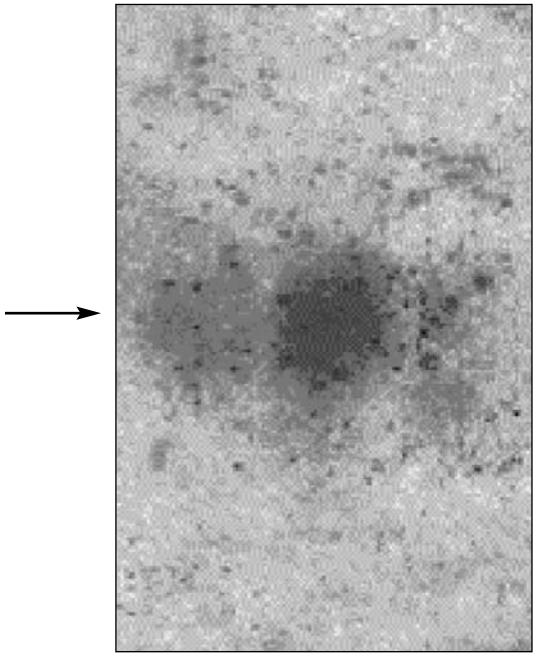

Fig. 2. Little millet zein-like cDNA Northern blot analysis for 750 bp little millet cDNA. Arrow indicates the respective band.

Homology searches. Millets are commonly classified with maize, sorghum, and Job's tears into the grass subfamily Panicoideae. In all species pertaining to this group, the major storage proteins are alcohol-soluble prolamins, which could be classified into α, β, γ, and δ according to their solubility, being deposited in organelles known as protein bodies located in cereal endosperm [31, 32]. A common reason for performing a database search with a query sequence is to find a related gene in another organism. For a query sequence of an unknown function, a matched gene may provide a functional clue. One very important principle for database searches is DNA sequence translations that encode proteins into protein sequences before performing a database search. Due to fivefold larger variety of sequence characters in proteins and peptides, it is much easier to detect sequence similarity patterns between protein sequences than between DNA sequences. It was proven [33-35] that searches with DNA sequence against a nucleic acid sequence database yield far fewer significant matches than searches using corresponding protein sequence. The mRNA sequence was translated to the corresponding protein primary structure using the Expasy translate tool.

Bioinformatics analyses were carried out in several forms. Similarity search was first done using BlastX, and all six possible reading frames were compared to protein sequence. Data bank searches were done against all databases by using the SwissProt data bank. As for the results, finger millet zein-like sequence showed maximum similarity to 22 kD zein-α precursor, a seed storage protein from *Z. mays* with 39% identity. The next best hit matched another seed storage protein named kafirin PSKR2-precursor from *S. bicolor*, with 38% identity, which also belongs to the zein-like family. The third hit matched to zein-α precursor (19 kD) with 41% identity. If the search is limited to *Z. mays* genes, all obtained hits showed identity to 22-kD as well as 19-kD zein precursor. Similar results were also obtained using protein blast (BlastP) analysis.

Search for conserved domains and protein family. Domains can be denominated according to distinct functional and/or structural units of proteins. Domains are often identified as recurring (sequence or structure) units in molecular evolution, whose extension can be determined by sequence and structure analysis. Molecular evolution may have utilized such domains as building blocks, recombined in different arrangements to modulate protein function. NCBI conserved domain search produced significant alignment to zein-like seed storage protein with a 1e-05e value (Fig. 3). Other top listed sequences in this family were closely related to kafirin PGK1 precursor (*S. bicolor*), zein-α precursor 22 kD (*Z.*

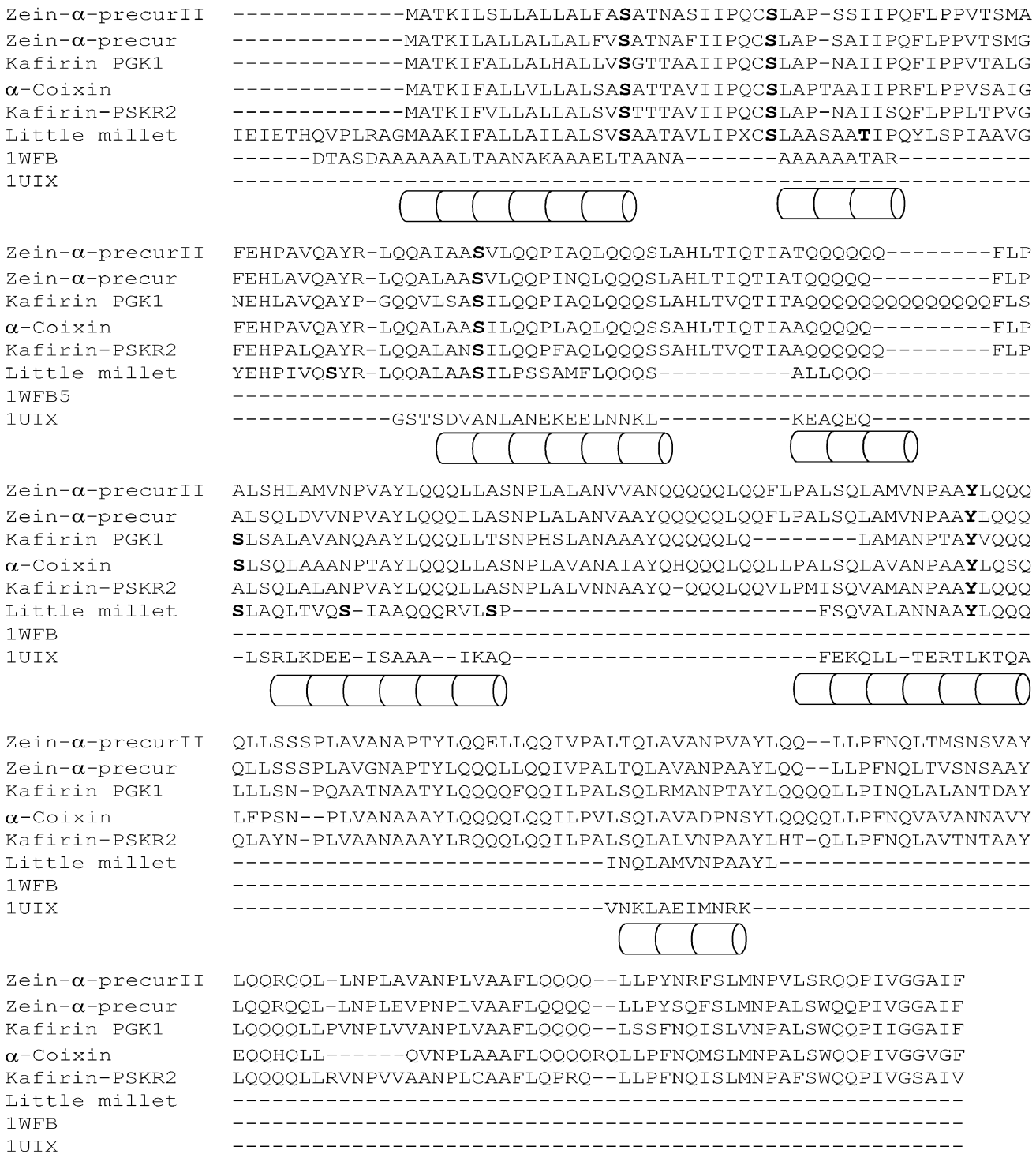

Fig. 3. Primary and tertiary structural alignments of little millet zein-like proteins against several zein-like family sequences (ClustalW) and also against two three-dimensional structures (BioInfo Meta Server): an antifreeze protein isoform (PDB code: 1WFB) and a Rho-kinase protein (PDB code: 1UIX). Bold residues correspond to possible glycosylation sites. Cylinders correspond to secondary structure of α-helices.

mays), α-coixin 27 kD (*C. lachryma-jobi*), and kafirin PSKR2 precursor (*S. bicolor*). All these proteins belong to zein-like family. Protein family (Pfam) search results generated with the domain starting from 14th residue showed that this protein belongs to the family of zein-like seed storage proteins with a 6.3e-44e value. Top scoring sequences were obtained for 22-kD zein-α precursor, 27-kD α-coixin, kafirin PSKR2, and kafirin PGK1.

As little millet zein-like protein contains only one cysteine residue, it could probably be classified into the α-prolamins' group. α-Prolamin primary structures contain peptide repeat motifs of about 20 residues, rich in leucine, alanine, and glutamine. A search in Prosite for motifs generated a score hit of 9.652 for sequence $_{88}$QQQSALLQQQSLAQLTVQSIAAQQQRVLSPFSQVA LANNAAYLQQQINQ$_{136}$. This motif consists in a typical glutamine-rich region from the prolamin group. Zeins are seed storage proteins, which are unusually rich in glutamine, proline (hence the name prolamins), alanine, and leucine; their primary function is related to nitrogen and sulfur accumulation, which are extremely important for seed germination [36]. Tsai et al. [37] hypothesized that prolamin and glutelin fractions in maize could act as nitrogen depots. Furthermore, it was also found that starch synthesis appears to be dependent on the presence of prolamin in maize [38].

It has also been proposed that repeat units present in prolamins form α-helices, which are arranged in anti-parallel fashion to give a compact circular or hexagonal structure, but direct evidence for this structure is lacking and will be suggested further in this manuscript by molecular modeling. Secondary structure prediction (SOPMA) of little millet zein-like protein shows that 67.81% of the residues form α-helices, which support the argument that this protein could be classified as prolamin-like [31].

Molecular modeling. Multiple sequence alignment generates a concise information-rich summary of sequence data (Fig. 3). It derives a quantitative evaluation of relatedness of sequence pairs and gives some information about evolutionary distances between aligned proteins. Significant hits obtained by various similarity searches mentioned elsewhere were used for multiple alignment construction. By using alignment a phylogram, which is a branching phylogenetic diagram (tree), where branch lengths are proportional to inferred evolutionary change amount, was constructed (data not shown). The phylogram produced by ClustalW [39] shows that on one hand α-coixin 27 kD (*C. lachryma-jobi*), 22 kD zein (*Z. mays*), and α-kafirin (*S. bicolor*) are closely related to little millet protein, and on the other hand 19 kD zein (*Z. mays*) was more evolutionary distant. Branch length also shows that zein is the latest in the evolutionary ladder, and the protein from little millet comes before that. Furthermore, this cladogram shows a common ancestry for zein-like proteins. The present analysis revealed that the structural genes involved in the expression of 22 kD α-prolamin-like genes of *C. lachryma-jobi*, sorghum, maize, and little millet might have originated from a common ancestor in Poaceae species and that variations were introduced in the structural and regulatory sequences after species separation.

To predict a tertiary structure of little millet zein-like protein, a three-dimensional model was constructed, showing enhanced stability when submitted to molecular dynamics simulation. Trajectory analysis demonstrated that the peptide did not break or modify its three dimensional conformation by the electrical forces of the water box. The final model shows a six α-helix folding in an anti-parallel conformation (Fig. 4). The structural validity of the model was checked using the program PROCHECK v3.5 [40], which showed that 98% of the residues occupied the most favored regions of the Ramachandran plot with permitted side-chain conformation. Previous structure predictions of two zeins suggested, as observed in this report, a repeated hairpin α-helical motif, which is arranged in short segments in an extended way connected by loops and turns [31]. The model presented here (Fig. 4) also agrees with the large amount of helix found by FTIR spectroscopy and fast N-H to N-D exchange measured by NMR [31]. On the other hand, our model clearly differs from the theoretical structure proposed by Matsushima et al. [41] for α-zeins, which showed helical structures arranged in a way that they are connected by several structures as loops and sheets, but not as anti-parallel helices as observed in our model. The model of little millet zein-like protein (Fig. 4) also differs from the structure of α-zein 3D structure from maize, which showed a α-helix folded back on itself, either partially or in half, forming a helix bundle [42] suggesting that the C-terminal domain stabilizes the protein fold by intramolecular hydrogen bonds. Otherwise, our

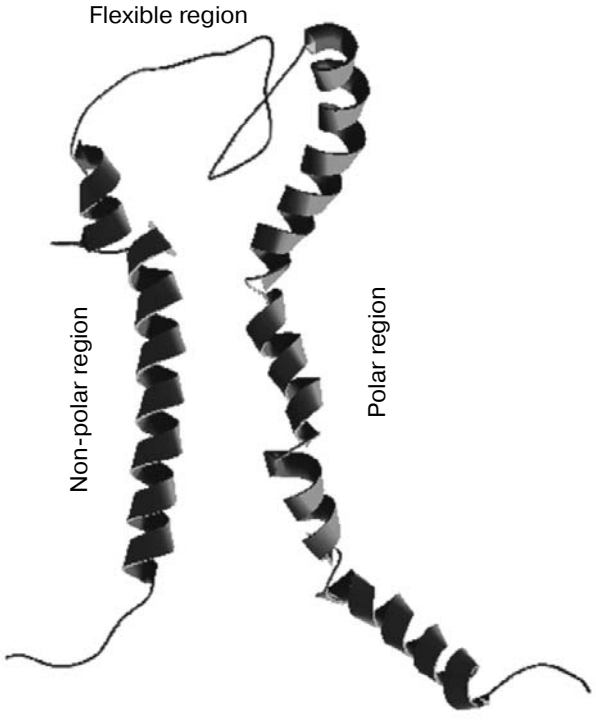

Fig. 4. Ribbon structure of three-dimensional molecular model of little millet zein-like protein. Light gray color indicates α-helices. Polar, nonpolar, and flexible regions are indicated.

model agrees with other proposed structures in relation to amphipathic properties [41-44].

Amphipathic secondary structures play an important role in protein folding, protein–protein interactions, and in self-assembly [43]. Since α-zeins are extremely involved in protein body formation, polar residues were observed on the α-helices indicating the involvement of intermolecular hydrophobic interactions, suggesting their importance in protein aggregation [42, 43]. It was also observed that Z19 prolamins could directly bind to free fatty acids, maintaining this interaction by van der Waals forces supplied by hydrophobic residues [31]. These interactions, suggested in our model (Fig. 4), could facilitate protein body formation, since protein–lipid interactions are normally required [45]. Furthermore, the hydrophobic and elongated molecular structure appears to remain stable in low water content environments, as simulated by molecular dynamics. This is an essential property for proteins in desiccated tissues as seeds and kernels, as observed for pennisetin and zeins [44].

In summary, data presented here indicate that little millet zein-like protein is closely related to a single domain zein family. Additionally, the micelle model described by Kogan et al. [43] is compatible with the known dimensions of zein-like little millet and probably could be involved in a self-assembly procedure. Several studies have demonstrated the potential of zeins as storage proteins [8]. The little millet zein-like proteins described here could be potentially used for production of transgenic plants with enhanced essential amino acid and protein contents, improving the nutritional quality of grains. These advances, along with our report, attest to a promising future for the use of biotechnology in genetic study and agronomic improvement of several crops.

The authors are thankful to Mr. K. Devakumar, Scientist (Central Plantation Crops Research Institute, Kasaragod, Kerala) for his valuable work on computational analysis.

This work was partially financed by Universidade Catolica de Brasilia.

REFERENCES

1. James, C. (1998) *Biotech. Develop. Mon.*, **35**, 9-12.
2. Swaminathan, M. S. (1995) *Issue Agric. 7*, CGIAR, Washington, DC, USA.
3. Gatehouse, A. M., Boulter, R. D., and Hilder, V. A. (1992) in *Biotechnology in Agriculture*, *Plant Genetic Manipulation for Crop Protection*, CAB International, 7, Wallingford, UK, pp. 155-181.
4. Franco, O. L., and Melo, F. R. (2000) *Rus. J. Plant Physiol.*, **47**, 137-144.
5. Layer, P., Carlson, G. L., and di Magno, E. P. (1985) *Gastrotech.*, **88**, 1875-1902.
6. Turcotte, G. E., Nadeau, L., Forest, J. C., Douville, P., Leclerc, P., Bergerm, J., and Laclos, B. F. (1994) *Clin. Biochem.*, **27**, 133-139.
7. Rajendran, P., and Thayumanavan, B. (2000) *J. Plant Biochem. Biotechnol.*, **9**, 89-94.
8. Coleman, G. E., Clore, A. M., Ranch, J. P., Higgins, R., Lopes, M. P., and Larkins, B. A. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 7094-7097.
9. Harris, E., Detmer, J., Dungan, J., Doua, F., White, T., Kolberg, J. A., Urdea, M. S., and Agabian, N. (1996) *J. Clin. Microbiol.*, **34**, 2401-2407.
10. Powlowski, K., Kunze, R., de Vries, S., and Bisseling, T. (1994) *Plant Mol. Biol.*, **5**, 1-13.
11. Sambrook, J., Frotsch, E. F., and Maniatis, F. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd Edn., Cold Spring Harbor, NY.
12. Davis, L. G., Dibener, M. D., and Battey, J. F. (1986) *Basic Methods in Molecular Biology*, Elsevier, NY.
13. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.*, **215**, 403-410.
14. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.*, **25**, 3389-3402.
15. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) *Bioinformatics*, **12**, 426-427.
16. Marchler-Bauer, A., Anderson, J. B., de Weese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., Liebert, C. A., Liu, C., Madej, T., Marchler, G. H., Mazumder, R., Nikolskaya, A. N., Panchenko, A. R., Rao, B. S., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Vasudevan, S., Wang, Y., Yamashita, R. A., Yin, J. J., and Bryant, S. H. (2003) *Nucleic Acids. Res.*, **31**, 383-387.
17. Marchler-Bauer, A., and Bryant, S. H. (2004) *Nucleic Acids Res.*, **32**, 327-331.
18. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiler, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. L. (2002) *Nucleic Acids Res.*, **30**, 276-280.
19. Hulo, N., Sigrist, C. J. A., Le-Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., de Castro, E., Bucher, P., and Bairoch, A. (2004) *Nucleic Acids Res.*, **32**, 134-137.
20. Geourjon, C., and Deleage, G. (1995) *Comput. Appl. Biosci.*, **11**, 681-684.
21. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) in *The Proteomics Protocols Handbook* (Walker, J. M., ed.), Humana Press, pp. 571-577.
22. Pearson, W. R. (1990) *Meth. Enzymol.*, **183**, 63-98.
23. Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) *Bioinformatics*, **19**, 1015-1018.
24. Sicheri, F., and Yang, D. S. (1995) *Nature*, **375**, 427-431.
25. Shimizu, T., Ihara, K., Maesaki, R., Amano, M., Kaibuchi, K., and Hakoshima, T. (2003) *J. Biol. Chem.*, **278**, 46046-46051.
26. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol.*, **234**, 779-815.
27. Lindahl, E., Hess, B., and van der Spoel, D. (2001) *J. Mol. Model.*, **7**, 306-317.
28. Van Gunsteren, W. F., Hunenberger, P. H., Mark, A. E., Smith, P. E., and Tironi, I. G. (1995) *Comp. Phys. Commun.*, **91**, 305-319.

29. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis*, **18**, 2714-2723.
30. Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., and Hochstrasser, D. F. (1993) *Electrophoresis*, **14**, 1023-1031.
31. Forato, L. A., Yushmanov, V. E., and Colnago, L. A. (2004) *Biochemistry*, **43**, 7121-7126.
32. Bicudo, T. C., Forato, L. A., Batista, L. A., and Colnago, L. A. (2005) *Anal. Bioanal. Chem.*, **16**, 1-6.
33. Pearson, W. R. (1995) *Protein Sci.*, **4**, 1145-1160.
34. Pearson, W. R. (1996) *Meth. Enzymol.*, **266**, 227-258.
35. Pearson, W. R. (2000) *Meth. Mol. Biol.*, **132**, 185-219.
36. Zhou, X., and Fan, Y. L. (1993) *Trans Res.*, **2**, 141-146.
37. Tsai, C. Y., Huber, D. M., and Warren, H. L. (1980) *Plant Physiol.*, **66**, 330-333.
38. Bhatia, C. R., and Rabson, R. (1987) in *Nutritional Quality of Cereal Grains: Genetic and Agronomic Improvements*, Vol. 28, Madison WI, USA, pp. 111-131.
39. Li, K. B. (2003) *Bioinformatics*, **19**, 1585-1586.
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.*, **26**, 283-291.
41. Matsushima, N., Danno, G. I., Takezawa, H., and Izumi, Y. (1997) *Biochim. Biophys. Acta*, **1339**, 14-22.
42. Tatham, A. S., Field, J. M., Morris, V. J., I'Anson, K. J., Cardle, L., Dufton, M. J., and Shewry, P. R. (1993) *J. Biol. Chem.*, **268**, 26253-26259.
43. Kogan, M. J., Dalcol, I., Gorostiza, P., Lopez-Iglesias, C., Pons, M., Sanz, F., Ludevid, D., and Giralt, E. (2001) *J. Mol. Biol.*, **312**, 907-913.
44. Bugs, M. R., Forato, L. A., Bortoleto-Bugs, R. K., Fischer, H., Mascarenhas, Y. P., Ward, R. J., and Colnago, L. A. (2004) *Eur. Biophys. J.*, **33**, 335-343.
45. Forato, L. A., Doriguetto, A. C., Fischer, H., Mascarenhas, Y. P., Craievich, A. F., and Colnago, L. A. (2004) *J. Agric. Food Chem.*, **52**, 2382-2385.