Eur Biophys J (2004) 33: 335–343
DOI 10.1007/s00249-003-0354-3

ARTICLE

**Milton Roque Bugs · Lucimara Aparecida Forato
Raquel Kely Bortoleto-Bugs · Hannes Fischer
Yvonne Primerano Mascarenhas · Richard John Ward
Luiz Alberto Colnago**

Spectroscopic characterization and structural modeling of prolamin from maize and pearl millet

Received: 6 June 2003 / Accepted: 18 August 2003 / Published online: 24 September 2003

Abstract Biophysical methods and structural modeling techniques have been used to characterize the prolamins from maize (*Zea mays*) and pearl millet (*Pennisetum americanum*). The alcohol-soluble prolamin from maize, called zein, was extracted using a simple protocol and purified by gel filtration in a 70% ethanol solution. Two protein fractions were purified from seed extracts of pearl millet with molecular weights of 25.5 and 7 kDa, as estimated by SDS-PAGE. The high molecular weight protein corresponds to pennisetin, which has a high α-helical content both in solution and the solid state, as demonstrated by circular dichroism and Fourier transform infrared spectra. Fluorescence spectroscopy of both fractions indicated changes in the tryptophan microenvironments with increasing water content of the buffer. Low-resolution envelopes of both fractions were retrieved by ab initio procedures from small-angle X-ray scattering data, which yielded maximum molecular dimensions of about 14 nm and 1 nm for pennisetin and the low molecular weight protein, respectively, and similar values were observed by dynamic light scattering experiments. Furthermore, ^{1}H nuclear magnetic resonance spectra of zein and pennisetin do not show any signal below 0.9 ppm, which is compatible with more extended solution structures. The molecular models for zein and pennisetin in solution suggest that both proteins have an elongated molecular structure which is approximately a prolate ellipsoid composed of ribbons of folded α-helical segments with a length of about 14 nm, resulting in a structure that permits efficient packing within the seed endosperm.

Keywords Pennisetin molecular model · Zein molecular model · Circular dichroism · Dynamic light scattering · Small-angle X-ray scattering

M. R. Bugs · L. A. Forato · L. A. Colnago (✉)
Embrapa Instrumentação Agropecuária, Rua XV de novembro 1452, 13560-970 São Carlos, SP, Brazil
E-mail: colnago@cnpdia.embrapa.br
Tel.: +55-16-2742477
Fax: +55-16-2725958

R. K. Bortoleto-Bugs · Y. P. Mascarenhas
Grupo de Cristalografia, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador Sãocarlense 400, 13566-590 São Carlos, SP, Brazil

H. Fischer
Instituto de Física, Universidade de São Paulo, Rua do Matão, Travessa R 187, Cidade Universitária, 05315-970 São Paulo, SP, Brazil

R. J. Ward
Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes 3900, 14049-901 Ribeirão Preto, SP, Brazil

Introduction

Seed storage proteins are the most abundant source of proteins in human and animal nutrition, and have being used in the development of new biodegradable materials for engineering and biological applications. However, the nutritional quality of these proteins is limited by deficiencies in the content of some essential amino acids, mainly lysine and tryptophan. The nutritional or technological quality of these proteins can be improved by genetic manipulation and a detailed knowledge of the protein structures is important for designing improved proteins with altered contents of specific amino acids.

A large number of studies have aimed to characterize the structure of prolamins, the alcohol-soluble storage proteins of various cereal kernels. Amino acid composition analyses of many of these proteins reveal a significant proportion of hydrophobic residues, a high α-helix content (Krestchmer 1957; Argos et al. 1982; Garratt et al. 1993) and an elongated molecular structure (Tatham et al. 1993; Matsushima et al. 1997). Zeins are amongst the better characterized groups of storage proteins derived from the prolamin fraction of maize (*Zea mays*) seed, and

are specifically expressed during seed development and act as a reservoir for amino acids. Two groups of zeins, Z19 and Z22, are commonly found with apparent molecular weights as estimated by SDS-PAGE of 19 kDa and 22 kDa, respectively (Argos et al. 1982; Tatham et al. 1993; Matsushima et al. 1997). However, a number of complete amino acid sequences of α-zeins have been derived from their cDNA sequences, which shows that the Z19 and Z22 proteins contain about 210 and 245 residues, with true molecular weights of about 23–24 kDa and 26–27 kDa, respectively (Pedersen et al. 1982; Heidecker et al. 1991). In pearl millet (*Pennisetum americanum*), the prolamin is known as pennisetin and represents the major nitrogenous fraction of the endosperm. In contrast to the prolamins from other members of the *Panicoideae* subfamily, which includes cereals such as corn, sorghum and coix, pennisetin from pearl millet shows no sequence similarity to any other prolamin analysed (Bietz 1982).

Several studies have been performed with the aim of determining the atomic structure of prolamins. The conformation of zeins in alcoholic solutions has been extensively investigated with various spectroscopic techniques, and various zein model structures have been proposed (Argos et al. 1982; Garratt et al. 1993; Tatham et al. 1993; Matsushima et al. 1997). The conformation and stability of pennisetin have previously been analysed using circular dichroism (CD) and ^{13}C nuclear magnetic resonance (NMR) spectroscopy (Sainani et al. 1992). Although all these models incorporate a common theme of packed α-helical repeats, details of the prolamin structure remain unclear.

In the present study we have combined intrinsic tryptophan fluorescence (ITF), CD, Fourier transform infrared (FTIR), dynamic light scattering (DLS), small-angle X-ray scattering (SAXS) and ^{1}H NMR spectroscopy to analyze the structure of zein from maize and pennisetin from millet. These techniques proved applicable for the study of the prolamins in solution, and the combined results obtained from these varied spectroscopic techniques were used to develop a structural model for the maize and millet prolamins. Our analysis indicates that the zeins and pennisetin are elongated structures with a length of about 14 nm derived from a superhelical structure comprised of α-helical coiled coils, which results in an apolar stripe along one side of each helix.

Materials and methods

Protein isolation

Zein and pennisetin were extracted using a simplified protocol which is a modification of a previously described method (Forato et al. 2000). Endosperm meal from the maize cultivar BR451 or CO3HS and millet cultivar BRS1501 was extracted with a three-fold (w/v) excess of 70% ethanol containing 25 mM Tris-HCl (pH 9) for 6 h at 4 °C under mild agitation in the absence of salt. After a brief centrifugation, the supernatant was applied at room temperature to a Superdex 75 HR 26/100 (Amersham Biosciences) column equilibrated and eluted with 25 mM Tris-HCl, (pH 9) containing 70% ethanol. Double distilled water was used in all buffers, and all chemicals and reagents were of analytical grade and used without further purification. Protein concentrations were determined by micro-Kjeldahl analysis, and over the concentration range of 0.5–15 mg/mL no evidence of aggregation was observed. Aliquots of the protein solution were analyzed for purity by silver staining of 15% SDS-PAGE gels using the Laemmli discontinuous buffer system (Laemmli 1970).

CD spectroscopy

The CD measurements were made with a Jasco J810 spectropolarimeter in quartz cuvettes of 5 mm optical path length. The sample chamber was purged with dry nitrogen to remove oxygen from the optical path, and CD spectra were collected at 25 °C between 195 and 250 nm with six cumulative wavelength scans. Spectra were obtained at a protein concentration of 25 μM in 25 mM Tris-HCl buffered at pH 9 and the water content of the buffer was varied from 30% to 94%. Protein CD spectra were corrected by subtraction of buffer blank spectra measured under the same conditions. Secondary structure content was estimated from UV circular dichroism spectra using the software package SOMCD (Unneberg et al. 2001). SOMCD is a neural network algorithm that estimates the secondary structure composition of a protein from UV circular dichroism data (http://solea.quim.ucm.es/somcd).

FTIR spectroscopy

FTIR spectra were recorded using a Paragon 1000 spectrometer from Perkin-Elmer, using the IR card technique, which avoids artifacts due to the strong water absorption in the IR region (Bugs and Cornélio 2001). A sample volume of 30 μL was layered over an IR screen cell sampling card and afterwards maintained enclosed for ~3 h over silica gel, resulting in a dried thin film of protein. Spectra from the IR card samples were collected with 128 scans at a resolution of 4 cm^{-1} over the wavenumber range 4000 to 400 cm^{-1} at 25 °C. The spectrometer was continuously purged with nitrogen to eliminate water vapor and CO_2 from the sample compartment, and the spectrum of the IR card in the absence of protein was recorded under conditions identical to the sample spectra and subtracted digitally from the protein spectra. Spectra were analyzed by taking the second derivative and the secondary structure content estimations were performed by Gaussian curve-fitting of the pennisetin spectra in the amide I band in the region 1600–1700 cm^{-1} using the Origin program (Microcal software) on the original (non-smoothed) protein vibrational spectra. Fitting was judged acceptable when the value of χ^2 was less than 10^{-6}. The fractional secondary structure content was calculated from the final fitted band areas in the amide I region.

Fluorescence spectroscopy

Steady-state ITF emission spectra were recorded with a SLM-AMINCO 8100 spectrofluorimeter. Protein samples at a concentration of 25 μM in 25 mM Tris-HCl buffered at pH 9 containing 30–63% water were placed in a quartz cuvette of 10 mm optical path length. An excitation wavelength of 295 nm was used and emission spectra were recorded from 310 to 450 nm with excitation and emission slit widths fixed at 4 nm. All experiments were performed at 25 °C, and all spectra were corrected by subtraction of a buffer blank.

Dynamic light scattering

DLS measurements were performed at 774.7 nm using a fixed scattering angle (90°) DynaPro-801 DLS instrument with sets of 50–150 measurements at 20 °C. The samples were centrifuged at 13,400×*g* for 10 min to remove air bubbles from the sample and

subsequently injected into a quartz cuvette with a volume of 12 μL. Data were collected and analyzed using the control software for molecular research DYNAMICS version 5.26.39 (Protein Solutions). Several measurements were taken, and the molecular weights of the proteins were calculated from the hydrodynamic radii using the standard curve model $MW = [R_{\text{h factor}} \times R_{\text{h}}]^{\text{power}}$, with "factor" and "power" derived from the molecular dimensions of globular proteins and pullulans families.

SAXS measurements and data analysis

SAXS data were collected at the SAS beamline of the National Synchrotron Light Laboratory (LNLS, Campinas, Brazil) using a 1D position-sensitive detector (Kellermann et al. 1997). The X-ray scatter of millet proteins at a concentration of 5 mg/mL prepared in 70% ethanol with 25 mM Tris-HCl (pH 9) was measured at 4 °C in a sample cell (volume of 100 μL) containing a 1 mm Teflon ring between two mica windows. A wavelength of $\lambda = 0.1488$ nm was used with a sample-to-detector distance of 843 mm covering the momentum transfer q interval of 0.1–4.5 nm^{-1} ($q = 4\pi\sin\theta/\lambda$, where θ is half the scattering angle). The scattering curves of the protein solutions and the corresponding buffer were collected in one- to two-minute frames in order to monitor beam stability and radiation damage to the sample. The data were normalized to the intensity of the incident beam, corrected for non-homogeneous detector response and desmeared for the 8 mm height entrance window of the position-sensitive detector. The scattering of the buffer was subtracted and the difference curves were scaled to equivalent concentrations. The experimental scattering data were analyzed using the indirect Fourier transform method as implemented in the program GNOM (Svergun 1992). The intraparticle distance distribution functions, $p(r)$, were calculated together with the radius of gyration of the protein (R_g) derived from the Guinier approximation: $I(q) = I(0)\exp(-q^2R_g^2/3)$ (Guinier and Fournet 1955), where $I(q)$ is the scattered intensity and $I(0)$ is the extrapolated forward scattering intensity [$I(q = 0)$ on a relative scale].

Ab initio molecular shape determination

The scattering intensity of the q interval did not vary over the time course of the data collection; therefore data from all time frames were used to construct the scattering intensity function, and the resolution of the resulting solution X-ray scattering curve extended to 20 Å. The low-resolution protein shape was restored using the ab initio procedure described by Svergun as implemented in the program GASBOR (Svergun et al. 2001). In this method, a dummy residue (DR) model is generated by a random-walk C_α chain and is folded to minimize a discrepancy between the calculated scattering curve from the model and the experimental data. The program simulates the protein internal structure, which makes it unnecessary to subtract a constant background from the experimental data to ensure the asymptotic Porod's behavior (Porod 1982), which yields the protein envelope at low resolution. The exact model is transformed into an assembly of spheres and adjusted to the estimated volume of the protein derived from an arbitrary ab initio amino acid sequence using tabulated specific amino acid volumes. Several runs of ab initio shape determination with different starting conditions led to consistent results, as judged by the structural similarity of the output models, and yielded essentially identical scattering patterns and fitting statistics in a stable and self-consistent process. The final shape restoration for pennisetin was performed using 248 dummy residues and 234 water molecules and no molecular symmetry was assumed. For the low molecular weight protein, an arbitrary model was proposed which was compatible with the calculated envelope.

NMR spectroscopy

The NMR spectra were acquired using a Varian Inova 400 spectrometer. The ^{1}H spectra of the zein and pennisetin solutions were recorded at a concentration between 250–500 μM in 70% deuterated ethanol-d_6 and 30% D_2O at 27 °C, using 32 scans and a $\pi/2$ pulse of 10 μs with a recycle time of 1 s and 3.7 s of acquisition time.

Results and discussion

Although prolamins have previously been studied by spectroscopic techniques and molecular model building based on the amino acid sequence arrangement, a definitive model for the structure of these proteins remains to be established. We have attempted to elucidate the prolamin solution structure using a series of complementary biophysical measurements, each of which characterizes different aspects of protein architecture and structural organization.

Purification of prolamins

A simple method was used for the purification of prolamins in alcoholic solution, and Fig. 1 presents the elution profile of the gel filtration of seed extract with Superdex 75 of total protein measured by the absorbance at 280 nm. In Fig. 1A a smaller protein peak at an elution volume of 135 mL (peak a) is assigned to pennisetin, and the second peak at an elution volume of 170 mL (peak b) is the major constituent of the extract, which has a predominant yellow color and contains a low molecular weight protein. The insert to Fig. 1A shows the silver-stained SDS-PAGE gel of the crude millet prolamin (lane 1), which exhibits bands of ~7, 25.5 and 51 kDa. The molecular weight of purified and concentrated pennisetin (lane 2), which corresponds to peak a in the gel filtration chromatogram, exhibits bands of 24.5, 22 and 49 kDa. The trace amount of the 22 kDa component is not seen in the crude extract, and the 49 kDa band is a residual amount of protein in dimeric form. In the samples from peak b (lane 3) a single band at ~7 kDa is observed. Figure 1B shows a smaller protein peak at an elution volume of 135 mL (peak a), which is assigned to the zein; as observed with pennisetin, the second peak at an elution volume of 170 mL (peak b) has a predominant yellow color. The insert to Fig. 1B shows the silver-stained SDS-PAGE of purified maize prolamins, which corresponds to peak a in the gel filtration chromatogram. The purified maize (CO3HS) (lane 1) exhibits bands of 22, 24.5 and 42 kDa, and the maize (BR451) (lane 2) shows bands of 25.4 and 50.8 kDa. The trace amount of the 42 and 50.8 kDa components is the residual amount of protein in dimeric form. Many maize endosperm storage proteins present a large disparity between the observed electrophoretic mobility in SDS-PAGE and the predicted mobility based on calculated molecular weights (Woo et al. 2001). However, the results of the present study demonstrate only minor differences in molecular weights, and in both cases a low molecular weight protein was present in the crude seed extract.

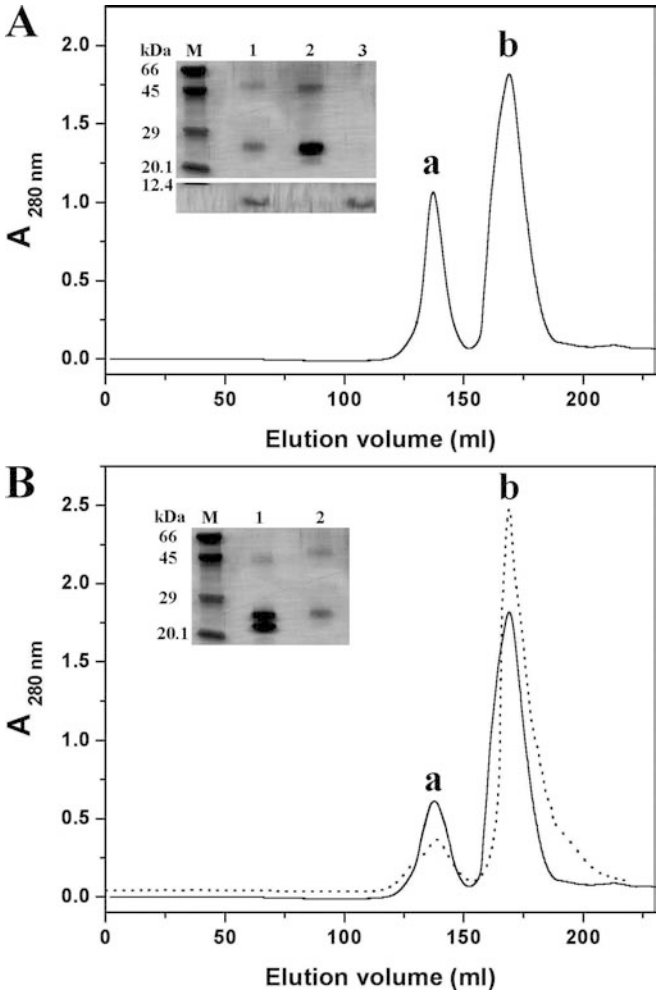

Fig. 1A, B Gel filtration by FPLC of millet and maize prolamins with Superdex 75 column chromatography. **A** (*a*) pennisetin and (*b*) the low molecular mass. **B** *Solid line*: CO3HS; *dotted line*: BR451; (*a*) zeins and (*b*) low molecular weight fraction. The *insert* in **A** shows the silver-stained SDS-PAGE gel of the millet prolamins before and after purification process (*M*, molecular marker masses; *1*, crude millet; *2*, peak of millet from *a*; *3*, peak of millet from *b*. The *insert* in **B** shows the silver-stained SDS-PAGE gel of the maize prolamins after the purification process: *1*, purified maize CO3HS; *2*, purified maize BR451

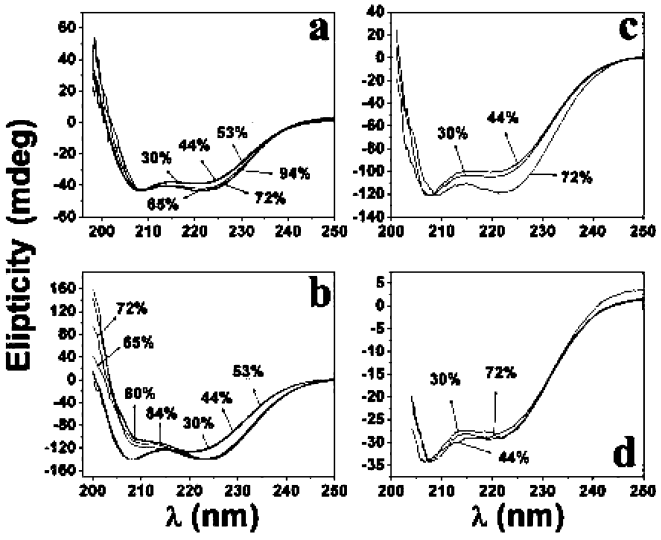

Fig. 2a–d The far-UV CD spectra of prolamins in alcoholic solution. The *arrows* indicate the spectra obtained by variation of the water content from 30% to 94%. The CD spectra were normalized to the minimum intensity of the first spectrum in order to compare the spectral changes as a function of water content of the sample. Peak A of purified millet before (**a**), and after lyophilization (**b**), and peak A of purified zein of CO311S (**c**) and BR451 (**d**). See Material and methods for details of the CD measurements

CD spectroscopy

CD measurements were used to monitor the influence of hydration on the secondary structures of zein and pennisetin. CD spectra were normalized to the intensity minimum of the spectra in 30% water in order to analyze the structural change as a function of the water content of the sample. As shown in Fig. 2a, CD spectra of the millet prolamin in alcoholic solution present profiles with characteristic minima at 208 and 222 nm that are typical of the α-helical conformation. Analysis of the secondary structure using the SOMCD software indicates that the subtle changes in the spectral profile are largely a consequence of the reduction in helical content from 89 ± 7% α-helix in 30% water to 75 ± 18% in 72% water. In comparison, Fig. 2b shows that, after lyophilization and subsequent rehydration of the purified sample, the spectrum of the same protein in greater than 65% water shows a reduced signal at 208 nm, and the spectral analysis suggests that this is due to a reduction in the helical content from 96 ± 2% in 30% water to 18 ± 9% in 94% water. These changes are concomitant with an increase in the fraction of β-sheet and random structures. Figure 2c shows the spectra of the zein from the maize cultivar CO3HS, which are also typical of α-helices, and shows a small change in conformation when the aqueous content of the media is increased from 30% to 72%, which is due to an increase in helical content from approximately 82 ± 10% to 96 ± 2%. The effect of water content is not observed in the zein from maize (BR451) (Fig. 2d), which has a constant helical content of around 89%.

In CD spectra of α-helical structures the band near 222 nm is observed due to the strong hydrogen-bonded environment of this conformation, and is largely independent of the length of the helix (Pelton and McLean 2000). The CD results presented here support the assumption that the secondary structure of the proteins will be affected by the stabilizing effect of alcohols on the helical structure (Nelson and Kallenbach 1986; Fasman 1996). The electrically forbidden nπ* transition near 220 nm is the lowest energy transition in the amide group, but has a large magnetic dipole transition moment directed along the carbonyl bond. This transition is sensitive to the polarity of the solvent, and may shift from ~222 nm in apolar solvents to ~215 nm in solutes

which provide strong H-bond donors. The $\pi\pi^*$ transition is electrically permitted, and the transition moment is directed approximately along the NO direction; however, the solvent dependence of this transition is much less than that of the $n\pi^*$ transition, showing only a marginal red shift when an amide is transferred from chloroform to water (Fasman 1996). These spectral properties are in broad agreement with the changes in the CD spectra observed for the prolamins, and lend support to the interpretation that when the aqueous content of the buffer increases, the α-helical content of the pennisetin is reduced owing to H-bonding interactions with the solvent. No effect of solvent on the CD spectra of the zein was observed, and it might therefore be predicted that the main-chain hydrogen bonds in the helices are more protected from solvent in these proteins. It should be noted, however, that the modeling of the structures was made in vacuum, and therefore a reliable test of this prediction awaits the results of more detailed modeling studies made in the presence of water.

FTIR spectroscopy

The spectrum of pennisetin in alcoholic solution is shown in Fig. 3, which show the typical amide I and II protein absorption bands from 1750 to 1450 cm^{-1}. The pennisetin spectrum has four peaks at 1642, 1652, 1660 and 1670 cm^{-1}, and the absorbance band centered at 1670 cm^{-1} has been assigned as arising from β-turns (Pelton and McLean 2000), which are often found on the surface of proteins where hydrogen bonding with the solvent is favorable. The broadening of the band in the amide I region in the FTIR spectra is due mainly to the absorbance contribution at 1670 cm^{-1}, which suggests a significant percentage of turns in the pennisetin molecule. The amide I spectral region of the pure pennisetin in alcoholic solution was resolved into four bands by using the peak positions (1624, 1642, 1652 and 1690 cm^{-1}) obtained from the second derivative of the original spectrum as the input parameters for curve-fitting analysis of the amide I band contour (Byler and Susi 1986). The Gaussian band profiles positions in the amide I region are assigned as follows: the 1652 cm^{-1} is assigned to α-helical structure, 1675 cm^{-1} to turns, 1691 cm^{-1} to β-sheet structures and the 1618 cm^{-1} component (4%) arises from amino acid side-chain vibrations (Byler and Susi 1986; Hadden et al. 1995). The results of this analysis reveal that the secondary structure content of the pennisetin is 72% α-helix, 15% turn and 9% β-sheet. The IR card spectra from two other well-characterized proteins, myoglobin and concanavalin A, with a defined content of α-helical and β-sheet, are also shown in Fig. 3 for comparative purposes. A wealth of structural information can be derived by analysis of the shape and position of bands in the amide I region of the IR card spectrum, and it is now well accepted that absorbance in the range from 1650 to 1658 cm^{-1} is generally associated with the presence of α-helices in aqueous environments. The spectrum of myoglobin is centered at 1653 cm^{-1} with a shoulder at 1644 cm^{-1}, and that of concanavalin A at 1635 cm^{-1}, demonstrating that the IR card spectra of these proteins have the same characteristics as those in aqueous solution (Haris and Chapman 1996; Pelton and McLean 2000). Nevertheless, the protein secondary structure assessed by CD is in close agreement with the secondary structure composition estimated by FTIR.

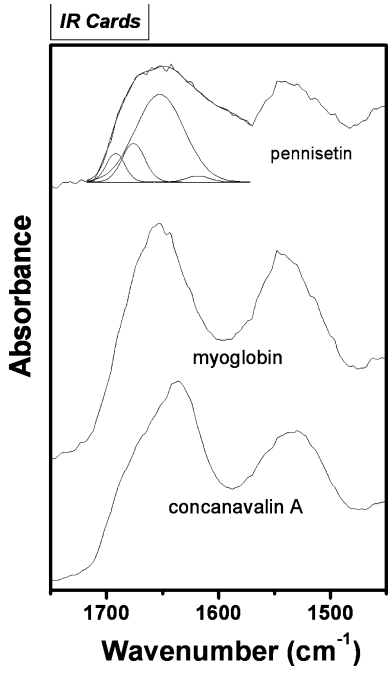

Fig. 3 FTIR spectra in spectral region 1750–1450 cm^{-1}. *IR cards*: pennisetin, myoglobin and concanavalin A. The component bands of pennisetin in the amide I region (1700–1600 cm^{-1}) were obtained by second-derivative and Gaussian curve-fitting analysis. Measurements were performed as described in Materials and methods at 25 °C with 4 cm^{-1} spectral resolution

Fluorescence spectroscopy

Fluorescence techniques were employed to study pennisetin in aqueous solvents. The crude millet prolamin does not demonstrate changes in the ITF spectra on increasing the water content of the buffer (Fig. 4); however, the ITF of both fractions (a and b) of the purified millet proteins show changes under these conditions. The emission maximum of the pennisetin (peak a) spectrum is blue-shifted by about 5 nm in buffer with increased water content, whereas the emission maximum of the low molecular weight sample from peak b is red-shifted under the same conditions. These changes indicate that pennisetin has one or more tryptophan residues that are in a partially hydrophobic environment and that these tryptophans become buried in a more hydrophobic environment with an increase in water content of the solvent. This is strong evidence of a conformational

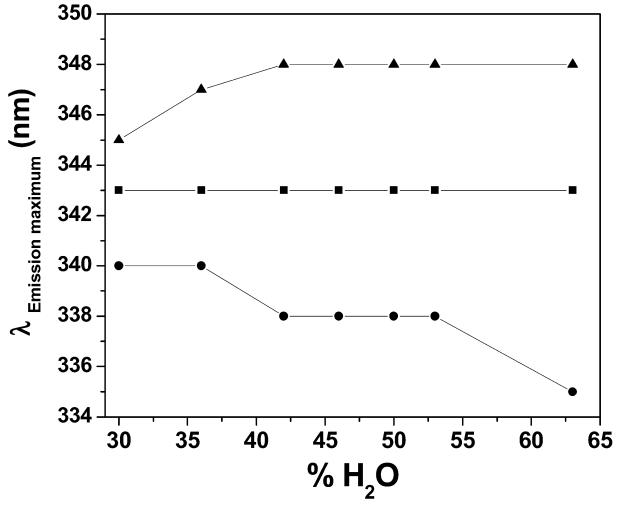

Fig. 4 Fluorescence emission maximum ($\lambda_{\text{Emission maximum}}$) in alcoholic solutions for millet prolamin as a function of the percentage variation of the water content (% H_2O) for (*squares*) crude millet, (*circles*) purified millet (*a* peak), and (*triangles*) the low molecular mass (*b* peak)

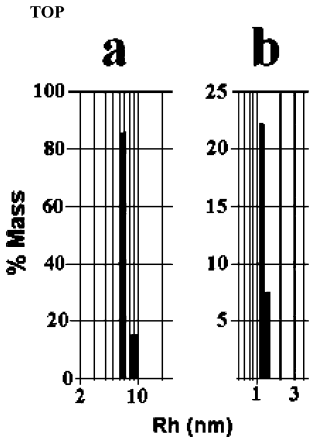

Fig. 5a, b DLS experiments of the purified millet. (**a**) Pennisetin (*a* peak); (**b**) corresponds to the low molecular mass (*b* peak). The histogram of size distribution for the samples exhibited a monomodal distribution with a hydrodynamic radius (R_h) of 6.70 nm and 8.90 nm for pennisetin (peak *a*) and 1.16 and 1.40 nm for the low molecular weight protein (peak *b*)

change of pennisetin in a more polar solvent. The emission maximum of the tryptophan in the low molecular weight protein from peak b is ~348 nm in 45% water, and is unchanged in more polar solvents, which suggests that in this protein the tryptophan is exposed at the protein surface, and the emission maximum reflects the polarity of the solvent. It should be noted that a more complete interpretation of the changes in the intrinsic fluorescence emission is hampered by the absence of the complete amino acid sequence of both pennisetin and the low molecular weight protein from peak b.

Dynamic light scattering

Figure 5 presents the result of DLS experiments with the purified millet prolamins in 70% ethanol, which show a narrow monomodal distribution with a hydrodynamic radius (R_h) of 6.70 nm and 8.90 nm for pennisetin (Fig. 5a) and 1.16 nm and 1.40 nm for the protein of low molecular weight in peak b (Fig. 5b). Factors which influence the DLS signal include effects arising from both shape and hydration of the protein, and the R_h of 6.70 nm indicates a molecule with a length of 13.4 nm. The molecular weight of the low molecular mass protein in peak b was estimated to be between 4.75 kDa and 7.40 kDa from a calibration curve of globular proteins of known molecular weight and size. However, the estimated molecular mass for pennisetin (R_h of 6.70 nm), calculated using the factor and power parameters derived from globular (286 kDa) and non-globular (64 kDa) proteins, resulted in predicted molecular weights that were different from the expected values in spite of the experimentally observed narrow monomodal distribution. This is an indication that the molecules are non-spherical and probably have a rod-like shape.

Small-angle X-ray scattering

SAXS measurements were performed using solutions of the purified millet prolamins from peaks a and b, and the scattering curves obtained are presented in Fig. 6. The pair distribution function $p(r)$ computed for the SAXS is a measure of the frequency of interatomic vector lengths within the protein, and this function is very sensitive to the overall asymmetry and domain structure within the particle. The Guinier plots exhibit a linear behavior and the estimated R_g of pennisetin and of the low molecular weight protein from peak b are 46(1) Å and 7.7 Å, respectively. The function obtained from the SAXS low-resolution model indicates maximum dimensions of the molecules of 145(20) Å and 20(1) Å, respectively. The pennisetin protein proved to be an elongated molecule having a prolate shape, as shown in Fig. 6a. The restored DR models of the proteins are presented as insets to Fig. 6, and reveal that pennisetin has an extended structure with an R_g value of 46 Å. For a special scattering function of a sphere with radius of gyration $R_g = \sqrt{3/5}R$, where R is the radius, the corrected R_g value of 46 Å corresponds to a radius of 60 Å (Glatter and Kratky 1982) for pennisetin. The maximum dimension of 145 Å and the radius of 6 nm for pennisetin is comparable with the R_h determined from the DLS measurements, and the extended asymmetric structure with a length of ~140 Å is in good agreement with previous estimates (Tatham et al. 1993; Matsushima et al. 1997). The experimental average size of the peptide as shown in Fig. 6b indicates that the low molecular weight protein in peak b has an R_g of 7.7 Å and maximum dimensions of 20 Å. Using a special

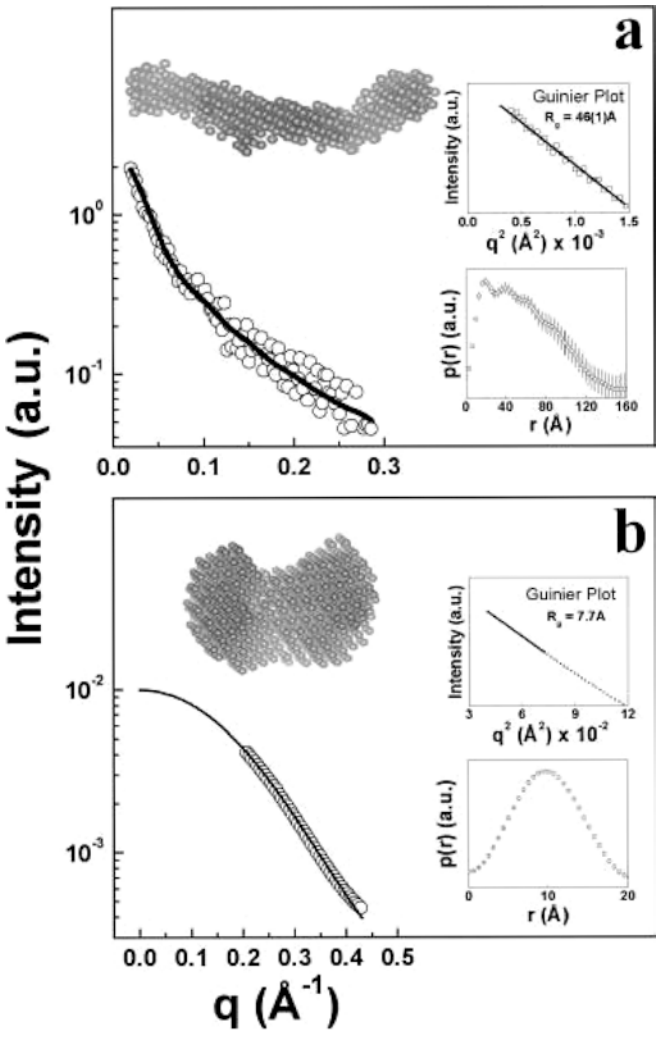

Fig. 6 Final fits for the solution scattering curves of (**a**) pennisetin and (**b**) the low molecular mass. The modeled curves (*continuous line*) of the experimental scattering data (*open circles*), together with the corresponding sphere-shaped models. The *insets* present the Guinier plots which show the monodisperse behavior of the protein in solution, and the pair distribution function $p(r)$ showing the maximum dimensions of the proteins. Experimental error bars are not shown. The protein models were generated up to 20 times, yielding self-consistent results

scattering function of a sphere, the R_g of 7.7 Å corresponds a radius of 10 Å, which is consistent with the value determined from DLS (1.16–1.35 nm).

NMR spectroscopy

Pennisetin and the zeins have molecular weights greater than 20 kDa, and present ^{1}H NMR spectra with low dispersion of the N–H and Cα–H chemical shifts which complicate the interpretation of NOE and spin–spin coupling data. Nevertheless, useful information can be derived from a qualitative analysis of the NMR spectra. Aromatic amino acids account for about 10% of the total amino acid content of pennisetin and the zeins, and in the previously published prolamin models (Argos et al. 1982; Garratt et al. 1993) it would be expected that shielding due to ring current effects arising from these aromatic side chains would be observed. However, the NMR spectra of these proteins do not show any signal below 0.9 ppm (data not shown), indicating that the solution structure is not compatible with a compact structure, which is contrary to the models previously proposed. Furthermore, the NMR spectra show a rapid N–H to N–D exchange in deuterated alcoholic solutions, which is completed during sample preparation and NMR calibration (less than 10 min). The access of N–H indicates they are exposed to solvent in an open structure, since if pennisetin and the zeins were a compact structure the N–H would be retained for several hours or days.

Molecular modeling

For structure prediction and modeling, two maize zein sequences (accession nos. PO2859/PO4699) and one pennisetin amino acid sequence (Q9M6M2) deposited in the SWISS-PROT and TrEMBL databases were used. The structure prediction results from Tatham et al. (1993) suggest that a single α-helix is folded to form a two-helix bundle; thus the initial conformation for the model was constructed by dividing the amino acid sequence between two anti-parallel α-helices to form an elongated molecular structure. Molecular models of these proteins were computed using the HyperChem 7 software packages. Several cycles of energy minimization were performed to optimize the geometry by calculating the structure for a molecule with the minimum energy and minimal atomic forces in vacuum. These studies showed that an extended conformation for the prolamins is preferred with a predominant fraction of α-helical conformation, which is in good agreement with the measurements obtained by CD and FTIR studies. The α-helical conformation is favored in nonpolar solvents due to the formation of an extensive H-bonded network, which decreases the number of polar peptide backbone atoms that form H-bonded contacts with the solvent. This is a common protein motif that is often associated with contacts between nonpolar side chains on adjacent helices. The preferred conformation with the minimum energy leads to the formation of coiled coils in which the two antiparallel α-helices form a superhelical conformation. The structural validity of the models was checked using the program PROCHECK v3.5 (Laskowski et al. 1993), which showed that >96% of the residues occupied the most favored regions of a Ramachandran plot with permitted side-chain conformations.

The molecular models for pennisetin and zein are illustrated in Fig. 7, in which the polar, charged and hydrophobic amino acids are distributed along the helical surfaces that pack anti-parallel against one another, and an apolar stripe is displayed along one side of each helix. The models shown in Fig. 7a (pennisetin) and

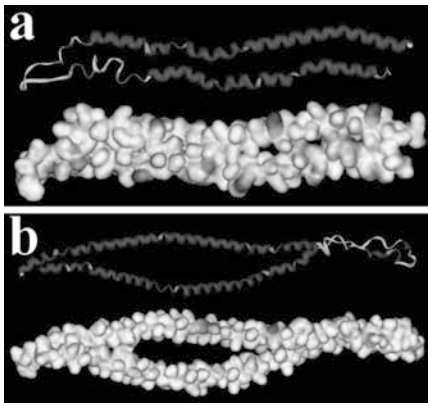

Fig. 7 The models of (**a**) pennisetin and (**b**) zein in solution. The HyperChem 7 software package was used to generate the structure by repeated cycles of energy minimization to find the optimum stereochemical conformation of pennisetin and zein in the monomeric form. The molecules are drawn showing α-helix, turn and random coils, and in surface representations (*lower structure*). The figures were produced with the DS ViewerPro 5.0 software package (Accelrys)

Fig. 7b (zein) have a percentage of secondary structure elements that are compatible with the results obtained from the spectroscopic techniques. The pennisetin model (Fig. 7a) contains 68% α-helix, 15% turn and the remainder in a random coil conformation. This α-helical packing gives rise to a rod-like structure for pennisetin with a length of about 136 Å and diameter of about 20 Å, with an axial ratio of 6.8:1. The zein model (Fig. 7b) contains 70% of the residues in α-helix, 13% in turn and the remainder in a loosely organized extended random coil conformation. This α-helical packing gives rise to an approximate prolate ellipsoid conformation with the semi-axes $a = 13.8$ nm and $b = 3.4$ nm, with an axial ratio of 4:1. The hydrophobic and elongated molecular structure appears to remain stable in low water content environments, which is an essential property for proteins in desiccated tissues such as seeds. The results clearly indicate that the prolamins have a similar extended and asymmetric structure with a length of about 14 nm, which is in good agreement with values previously reported (Tatham et al. 1993; Matsushima et al. 1997). Furthermore, the surface properties of the molecular model in solution suggest that a structural unit composed of ribbons of folded α-helical segments allows a more efficient packing of the protein, again a desirable property for a storage molecule.

Conclusions

We have applied a combination of theoretical and experimental techniques to characterize the conformational of pennisetin and zein in alcoholic solutions. The biophysical measurements reveal an elongated molecular structure for prolamins, with a length of about 140 Å. The helix packing presented in the proposed model involves the interaction between two antiparallel amphipathic α-helices. The overall molecular dimensions are in excellent agreement with those reported previously (Tatham et al. 1993; Matsushima et al. 1997); however, the helical packing in the model presented here differs from those previously reported. The elongated topology of prolamins is in agreement with the ^{1}H NMR data, which indicate these proteins are not globular in alcoholic solution, and that the H atoms of the N–H groups are exposed to solvent. Previous physicochemical studies have indicated that the α-zeins have an elongated rod or prolate ellipsoid structure in solution (Williams and Watson 1938; Elliot and Williams 1939; Foster and Edsall 1945), and a similar conclusion was drawn from CD, SAXS and viscosimetry measurements of α-zeins (Tatham et al. 1993; Matsushima et al. 1997), which are consistent with an elongated structure. This elongated structure has been observed for other proteins from the prolamin family, which includes gliadin and the glutelins from wheat, secalin from rye and hordein from barley. In the absence of high-resolution experimental data, based on X-ray crystallography or NMR in solution, we believe that the prolamin model is consistent with the spectroscopic data presented. Much remains to be understood with respect to the structure, function and biological mechanisms of prolamins, and it is interesting to note that although pennisetin shows no sequence similarity to other seed prolamins, the antiparallel α-helical motif and overall protein topology are clearly conserved.

Acknowledgements This work was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), grant nos. 01/08779-0, 97/13449-1 and 98/14526-2, and by the Brazilian National Research Council (CNPq), grant no. 301274/02-0. We are grateful to EMBRAPA's Maize and Sorghum Research Center (Sete Lagoas, MG, Brazil) for supplying samples of the millet BRS1501 and maize BR451 cultivars, to Gustavo Frederico Guilherme Kudiess sementes (Chiapeta, RS, Brazil) for samples of the maize cultivar BR451, and to the Instituto Agronômico de Campinas (IAC) (Campinas, SP, Brazil) for samples of the maize cultivar CO3HS. We are grateful to Prof. Dr. Igor Polikarpov (IFSC-USP, São Carlos, SP, Brazil) for access to the DLS equipment.

References

Argos P, Pedersen K, Marks MD, Larkins BA (1982) A structural model for maize proteins. J Biol Chem 257:9984–9990

Bietz JA (1982) Cereal prolamin evolution and homology revealed by sequence analysis. Biochem Genet 20:1039–1053

Bugs MR, Cornélio ML (2001) Analysis of the ethidium bromide bound to DNA by photoacoustic and FTIR spectroscopy. Photochem Photobiol 74:512–520

Byler DM, Susi H (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers 25:469–487

Elliot MA, Williams JW (1939) The dielectric behavior of solutions of the protein zein. J Am Chem Soc 61:718–725

Fasman GD (1996) Circular dichroism and the conformational analysis of biomolecules. Plenum Press, New York

Forato LA, Colnago LA, Garratt RC, Lopes-Filho MA (2000) Identification of free fatty acids in maize protein bodies and purified α zeins by ^{13}C and ^{1}H nuclear magnetic resonance. Biochim Biophys Acta 1543:106–114

Foster JF, Edsall JT (1945) Studies on double refraction of flow II. The molecular dimensions of zeins. J Am Chem Soc 67:617–625
Garratt R, Oliva G, Caracelli I, Leite A, Arruda P (1993) Studies of the zein-like α-prolamins based on an analysis of amino acid sequences: implications for their evolution and three-dimensional structure. Proteins Struct Funct Genet 15:88–99
Glatter O, Kratky O (1982) Small-angle X-ray scattering. Academic Press, London
Guinier A, Fournet G (1955) Small angle scattering of X-ray. Wiley, New York
Hadden JM, Chapman D, Lee DC (1995) A comparison of infrared spectra of proteins in solution and crystalline forms. Biochim Biophys Acta 1248:115–122
Haris PI, Chapman D (1996) Fourier transform infrared spectroscopic studies of biomembrane systems. In: Mantsch HH, Chapman D (eds) Infrared spectroscopy of biomolecules. Wiley-Liss, New York, pp 239–278
Heidecker G, Chaudhuri S, Messing J (1991) Highly clustered zein gene sequences reveal evolutionary history of the multigene family. Genomics 10:719–732
Kellermann G, Vicentin F, Tamura E, Rocha M, Tolentino H, Barbosa A, Craievich AF, Torriani I (1997) The small-angle X-ray scattering beamline of the Brazilian synchrotron light laboratory. J Appl Crystallogr 30:880–883
Krestchmer CB (1957) Infrared spectroscopy and optical rotatory dispersion of zein, wheat and gliadin. J Phys Chem 61:1627–1631
Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bactriophage T4. Nature 227:680–685
Laskowski RA, MacArthuer MW, Moss D, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structure. J Appl Crystallogr 26:283–291
Matsushima N, Danno G, Takezawa H, Izumi Y (1997) Three-dimensional structure of maize α-zein proteins studied by small-angle X-ray scattering. Biochim Biophys Acta 1339:14–22
Nelson JW, Kallenbach NR (1986) Stabilization of ribonuclease S-peptide α-helix by trifluoroethanol. Proteins Struct Funct Genet 1:211–217
Pedersen K, Devereux J, Wilson DR, Sheldon E, Larkins BA (1982) Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell 29:1015–1026
Pelton JT, McLean LR (2000) Spectroscopic methods for analysis of protein secondary structure. Anal Biochem 277:167–176
Porod G (1982) General theory. In: Glatter O, Kratky O (eds) Small-angle X-ray scattering. Academic Press, London, pp 17–51
Sainani MN, Mishra VK, Gupta VS, Ranjekar PK (1992) Circular dichroism and ^{13}C nuclear magnetic resonance spectroscopy of pennisetin from pearl millet. Plant Sci 83:15–22
Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25:495–503
Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80:2946–2953
Tatham AS, Field JM, Morris VJ, I'Anson KJ, Cardle L, Dufton MJ, Shewry PR (1993) Solution conformational analysis of the α-zein proteins of maize. J Biol Chem 268:26253–26259
Unneberg P, Merelo JJ, Chacón P, Morán F (2001) SOMCD: method for evaluating protein secondary structure from UV circular dichroism spectra. Proteins Struct Funct Genet 42:460–470
Williams JW, Watson CC (1938) The physical chemistry of the prolamins. Cold Spring Harbor Symp Quant Biol 6:208–214
Woo YM, Hu DWN, Larkins BA, Jung R (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. Plant Cell 13:2297–317