THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 257, No. 17, Issue of September 10, pp. 9984-9990, 1982
Printed in U.S.A.

A Structural Model for Maize Zein Proteins*

(Received for publication, April 9, 1982)

Patrick Argos‡§, Karl Pedersen¶, M. David Marks¶, and Brian A. Larkins¶‖

From the §Department of Biological Sciences and the ¶Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907

With the knowledge of the amino acid sequences of two maize zein proteins (apparent molecular weights of 19,000 and 22,000), a structural model is proposed for their molecular conformation. The circular dichroic spectrum taken in the 190-240 nm range for a zein protein mixture in methanol solution showed the zein secondary structure to be largely helical. The polar, hydrophobic, and turn characteristics of the zein residues, as well as the homologous repeat units in their primary sequences, suggested a structure with nine adjacent, topologically antiparallel helices clustered within a distorted cylinder. Polar residues distributed along the helical surfaces allowed intra- and intermolecular hydrogen bonding such that the zein molecules could be arranged in planes. The proposed glutamine-rich turns located between the helices and at the cylindrical caps would favor side chain interactions resulting in stacking of the molecular planes. Physical properties observed for the zein proteins are explained by the model.

The zein storage proteins are found in maize endosperm where they are deposited as protein bodies within the rough endoplasmic reticulum (Hurkman *et al.*, 1981; Larkins and Hurkman, 1978). They comprise at least 50% of the total endosperm protein and nearly 40% of the whole grain proteins (for a review, see Wall and Paulis, 1978). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Lee *et al.*, 1976) has resolved four molecular weight species for the zein polypeptides, of which two predominate with apparent weights of 19,000 and 22,000 (referred to as Z19 and Z22, respectively). Isoelectric focusing yields considerable charge heterogeneity within the zein fraction (Righetti *et al.*, 1977). Interest in the zein molecules has centered upon their genetic regulation as well as high lysine mutants which improve the nutritive quality by enhancing the essential amino acid balance (Wall and Paulis, 1978).

The zein proteins display significant hydrophobic properties. They readily self-associate to form protein bodies which are stably retained in membrane vesicles (Hurkman *et al.*, 1981). They are insoluble in water even with low concentrations of salt and require high percentage ethanol aqueous systems to maintain molecular conformation (*cf.* Rees and Singer, 1955 and 1956). Amino acid composition analyses show large amounts of hydrophobic residues such as leucine, proline, alanine, and phenylalanine (Gianazza *et al.*, 1977). Application of genetic cloning techniques has resulted in the nucleotide sequences and primary structures of Z19 (Pedersen *et al.*, 1982; Geraghty *et al.*, 1981) and Z22 (Marks and Larkins, 1982).

In the present work, investigations to determine the molecular conformation of the Z19 and Z22 proteins are discussed. Circular dichroic data gathered in the 190–240 nm range from a 70% methanol solution of all the zein species indicated an average α-helical content of 50–60% with turn or random coil configurations comprising the remaining structure. The Z19 and Z22 amino acid sequences showed homologous, 20-residue spans which are repeated nine times within both primary structures. An analysis of the physical characteristics (hydration potential, polarity, and turn- and helical-forming propensities) of the amino acids comprising the repeated segments indicated them to be α-helices flanked by turn regions. The helical portions consisted of a few polar amino acids and several hydrophobic residues that are often found in membrane-associated signal sequences. The proposed turn segments are rich in glutamine. The suggested structural model for the zein proteins consists of nine topologically antiparallel and adjacent helices clustered within a distorted cylinder of oval cross-section with the helical and cylindrical axes aligned. The top and bottom of the helical cylinder would be populated by polar glutamine residues which largely comprise the turn spans between helices. Polar and hydrophobic residues appropriately distributed along the helical surfaces allow intra- and intermolecular hydrogen bonds and van der Waals interactions among neighboring helices such that the rod-shaped zein molecules could aggregate in molecular planes which would then stack through glutamine interactions at the cylindrical caps. Such a model would explain the dense, membrane-enveloped deposits formed by the proteins within maize seeds.

METHODS

The circular dichroic spectrum was measured for a mixture of zein proteins in a 70% methanol solution with relative amounts of zein species proportioned according to their natural occurrence. The NH_2-terminal signal sequences were removed. The CD measurements were made in a wavelength range of 190–240 nm with a Cary 60 spectropolarimeter with the 6002 CD attachment under constant nitrogen flush. Background resulting from the alcohol solution was subtracted from the CD signal by mechanical adjustment in 5 nm steps. Specimen cells of path length 0.1 cm were used.

The fractional composition of helical (f_α), β-strand (f_β), and irregular or turn (f_t) structures were determined by the following equation:

$$[\theta]_\lambda = f_\alpha[\theta]_{\lambda\alpha} + f_\beta[\theta]_{\lambda\beta} + f_t[\theta]_{\lambda t} \quad (1)$$

where $[\theta]_\lambda$ is the measured molar ellipticity of the proteins at the wavelength λ while $[\theta]_{\lambda\alpha}$, $[\theta]_{\lambda\beta}$, and $[\theta]_{\lambda t}$ are the reference CD spectra for the three secondary structural types. The basis spectra were taken from Chen *et al.* (1974) who used eight reference proteins with known secondary structural compositions and from Greenfield and Fasman

* This is Journal Paper No. 9000 of the Purdue Agriculture Experiment Station. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of National Institutes of Health Grant GM27682 and American Cancer Society Faculty Research Award FRA173. To whom correspondence should be addressed.

‖ Recipient of National Science Foundation Grant PCM-8003757.

(1969) who utilized reference spectra from synthetic polypeptides in a given structural state. Since the observed and reference spectra are known, a linear least squares procedure was utilized to determine the best f values that satisfy Equation 1 for various wavelengths. Approximately 20 readings of the zein ellipticity in 2–3 nm intervals were used to determine the zein fractional compositions under the constraints that

$$f_\alpha + f_\beta + f_t = 1 \tag{2}$$

and

$$f_{\alpha,\beta,\text{ or } t} \geq 0 \tag{3}$$

Since the CD analysis indicated a large helical content for the zein proteins, certain physical characteristics of the amino acids in Z19 and Z22 were plotted against sequence number to delineate the secondary structural regions. The parameters included 1) the experimental hydration potential of Wolfenden *et al.* (1979), 2) the Chou-Fasman conformational preference parameters for α-helix (Chou and Fasman, 1974, a and b) as calculated by Levitt (1978), 3) the Chou-Fasman conformational propensity values for reverse turn configurations (Chou and Fasman, 1974, a and b) as calculated by Levitt (1978), and 4) the normalized propensity for a residue to be in a helical conformation within a membrane.[1] The Wolfenden values result from measured vapor-water partition coefficients for model compounds identical with each of the amino acids. As Wolfenden *et al.* (1979) do not list the hydration potentials for glycine and arginine, these values were assigned according to the method of Moews *et al.* (1981) who proportioned the residue hydration states to their degree of buriedness in soluble proteins as determined by Chothia (1976). The conformational preference parameters (P_C) are calculated for the 20 amino acids to indicate their likelihood to be in a particular secondary structural state (α, β, or t). They were first defined by Chou and Fasman (1974, a and b) as

$$P_{C,i} = \frac{N_{C,i} \Big/ \sum_{i=1}^{20} N_{C,i}}{N_{D,i} \Big/ \sum_{i=1}^{20} N_{D,i}} \tag{4}$$

where $N_{C,i}$ is the number of times the ith amino acid type appears within a given secondary structure in known protein structures and $N_{D,i}$ is the number of times the ith amino acid type appears within sequences of the protein data base. The denominator which is the fractional occurrence of a residue within the observed protein structures normalizes the P_C values to 1.0. The parameters for helical and turn configurations used in this work were calculated by Levitt (1978) and based on the conformation of nearly 11,000 residues in 60 known protein structures. The propensities for amino acids to be in a helical configuration within the hydrophobic environment of a membrane were determined by Argos *et al.*[1] with a 1125-residue data base taken from the seven helical membrane-buried spans in bacteriorhodopsin as proposed by Engelman *et al.* (1980), signal sequence segments listed by von Heijne (1981) and Austen (1979), and various other transmembrane spans suggested for such proteins as glycophorin (Tomita *et al.*, 1978) and cytochrome b_5 (Dailey and Strittmatter, 1981). Since the preference parameters are all positive and normalized to a value of 1.0 by definition, the hydration potentials were similarly scaled by adding the largest negative potential to each of the values listed by Wolfenden *et al.* (1979) and then dividing by their mean. Normalized hydration states greater than 1.0 would thus correspond to the more hydrophobic amino acids. The four parameters are given for each of the 20 amino acids in Table I.

Plots of the amino acid sequence number *versus* a given parametric value for a particular residue in the primary sequence were calculated. The curves were then "smoothened" through several cycles according to the methodology of Rose and Siddhartha (1980) who determine a least squares line for all successive seven-point groups to determine points for the smoothened graph. For example, the best line is found through seven points on the plot corresponding to residue numbers (i) to (i + 6); the line is then used to calculate the new parametric value for residue (i + 3). The procedure is then repeated for amino acids (i + 1) to (i + 7), (i + 2) to (i + 8), and so forth, utilizing the unsmoothened parametric values. One cycle of smoothening would be complete when reaching the residue stretch (i_C − 6) to (i_C), where (i_C) is the COOH-terminal amino acid number. With the newly calculated values, the entire process is repeated, constituting the second cycle of smoothening. The plots for the four parameters shown in the present work were translated such that their average value was set to 0.0. The first and last 3 residues in the primary sequence were arbitrarily assigned 0.0 parametric values due to the end effects of smoothening.

Helical "wheels" are used to show the angular and projected spatial distribution of polar and hydrophobic residues about the helix axis. Successive C_α-atoms in a helical peptide backbone would, in projection down the helical axis, lie on a circle with successive positions rotated by 100°. Lines emanating from the circle center and passing through the C_α positions would illustrate the projected direction of the associated amino acid side groups giving the appearance of a helical wheel (Schiffer and Edmundson, 1967). Only 18 unique C_α points separated by 20° would exist on the wheel regardless of the helical length as there are generally 3.6 residues per turn of an α-helix.

The amino acid numbering scheme includes signal sequences. Thus, the total residue count for Z19 and Z22 are, respectively, 235 and 266 amino acids.

RESULTS AND DISCUSSION

Circular Dichroism—The CD spectrum was measured for a mixture of zein proteins in 70% methanol solution in the 190–240 nm wavelength range (Fig. 1). The fraction of helix, sheet, and turn secondary structures was determined by fitting the observed CD data to that calculated from the basis spectra for the three secondary structural types. A linear least squares procedure yielded the best fractional values (f_α, f_β, and f_t) to satisfy Equation 1 (under the constraints of Equations 2 and 3) for approximately 20 experimental ellipticity readings at different wavelengths spaced by about 2 nm. The reference spectra were taken from Chen *et al.* (1974) who used eight known protein structures with helices averaging about 10 residues in length and Greenfield and Fasman (1969) who based their reference spectra on CD measurements of poly-L-lysine synthetic peptides in various configurational states. The Chen *et al.* fit yielded 44% α-helix, 5% β-strand, and 51% turn while the Greenfield-Fasman application resulted in 59% α-helix, 0% β-structure, and 41% reverse turn. Given the suggested zein structural model (see below), which consists of stacked molecules with aligned helical axes, the higher helical content from the Greenfield-Fasman analysis would be ex-

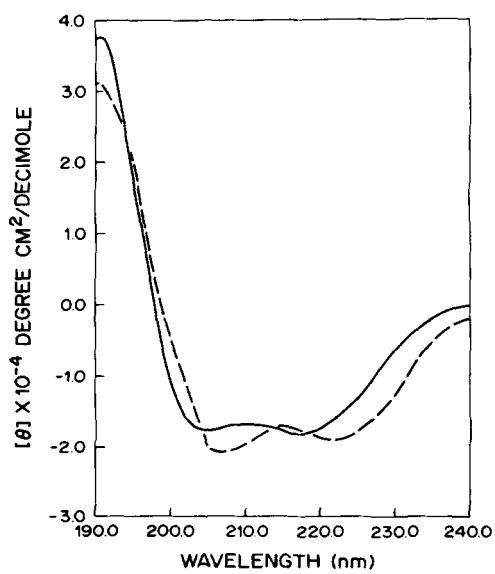

FIG. 1. **Circular dichroic spectrum (*solid*) for the zein protein mixture in 70% methanol.** The calculated curve (*dashed line*) is based on the reference spectra determined by Greenfield and Fasman (1969).

[1] P. Argos, J. K. Mohana Rao, and P. A. Hargrave (1982) *Eur. J. Biochem.*, in press.

pected as the reference spectra are derived from long synthetic polymers. This correspondence would lend support to the helical structural model proposed here. Optical rotary dispersion studies also suggest that zein proteins have a large helical content. Kretschmer (1957) found that zein in 80% ethanol displayed 50% helix while Danzer *et al.* (1975) noted 42% helix in 100% *N*-methylacetamide.

Considerable controversy has recently erupted regarding precision of the fractional compositions resulting from basis spectra fitting (*cf.* Brahms and Brahms, 1980; Provencher and Glockner, 1981). It has been suggested that measurements at wavelengths lower than 210 nm are significant for proper analysis. However, it is generally agreed that helical estimates based on longer wavelength data are accurate (Chang *et al.*, 1978; Siegel *et al.*, 1980).

Repeat Peptides in Z19 and Z22—The amino acid sequences of Z19 and Z22 and their homologous arrangements are shown in Fig. 2. The primary structures were aligned by visual inspection to achieve the largest number of amino acid identities in the two sequences. The Z19 and Z22 structures can be divided into four basic segments: an NH_2-terminal signal sequence, an NH_2-terminal turn region, nine repeating sequences flanked by glutamines, and a COOH-terminal region. Z22 displays a 19-residue insertion in the COOH-terminal span relative to Z19. The nine repeat units in Z19 are in agreement with those suggested by Pedersen *et al.* (1982) who compared the nucleotide sequences. Geraghty *et al.* (1981), who determined the nucleotide sequence of a similar maize Z19 protein, also identified a repeating unit; however, their analysis placed the successive glutamine residues in the middle of each repeat. This alignment would provide only seven, instead of nine, repeat units. Repeats in the physical characteristics of the amino acids comprising the unit structures also support the nine-unit interpretation (*see below*).

Amino Acid Characteristics in Z19 and Z22—The homologous repeat polypeptides would be expected to adopt a similar secondary structure. Since the CD and ORD data indicate a large helical content, it is reasonable to propose that the repeat units represent generally hydrophobic α-helices flanked by polar glutamine-rich turn regions. Turns exposed to the solvent are typically composed of charged and polar residues in known protein structures[1] (Chou and Fasman, 1978). A smoothened plot of the amino acid physical characteristics with sequence number should display a peak-trough pattern corresponding to the homologous repeat units.

A plot of the Chou-Fasman helical potential after three cycles of smoothening is shown in Fig. 3 for Z19. The dashed line peaks in the graph correspond to the repeating peptides and signal sequence as delineated in Fig. 2, while the dashed line troughs correspond to the NH_2- and COOH-terminal turn regions. It is evident that there is no clear phase relationship between the helical potential peaks and the peptide repeats. However, a plot of the hydration potential (hydrophobicity) for the Z19 amino acids shows an obvious correspondence

Zein protein		Sequence positions
	SIGNAL SEQUENCE	
Z22	M A T K I L S L L A L L A L F A S A T N A	1-21
Z19	M A A K I F C L I M L L G L S A S A A T A	1-21
	N-TERMINAL TURN	
Z22	S I I P Q C S L A P . S S I I P Q F L P P V T S M A F E H P A V Q A Y R	22-56
Z19	S I F P Q C S Q A P I A S L L P P Y L S P A M S S V C E N P I L L P Y R	22-57
	REPEAT SEQUENCES	
Z22	. L Q Q A I A A S V L . . Q Q P I . A Q L Q . . .	57-74
Z22	. . Q Q S L A . . . H L T I Q T I . A T Q Q . . .	75-90
Z22	. Q Q Q F L P A L S H L A M V N P V A Y L Q . . .	91-111
Z22	. . Q Q L L A S N P L A L A N V . V A N Q Q Q Q .	112-132
Z22	Q L Q Q F L P A L S Q L A M V N P A A Y L Q Q Q .	133-156
Z22	Q L L S S S P . . . L A V A N A P . T Y L Q Q E L	157-177
Z22	L Q Q I V P A L T Q L A V A N . P V A Y L Q . . .	178-198
Z22	. . . Q L L P F N Q L T M S N S . V A Y L Q Q R .	199-218
Z22	. Q Q L L N P . . . L A V A N P L V A A F L Q . .	219-237
Z19	. I Q Q A I A A G I L . . . . P L S P L F L Q . .	58-75
Z19	. Q S S A L L Q Q L P L V H L L A . Q N I R A Q Q	76-98
Z19	. L Q Q . . . . . L V L A N L A . A Y S Q Q . .	99-113
Z19	. . Q Q F L P F N Q L A A L N S A . A Y L Q Q . .	114-133
Z19	. . Q Q L L P F S Q L A . . . . A . A Y P R . . .	134-148
Z19	. . . Q F L P F N Q L A A L N S H . A Y V Q Q . .	149-167
Z19	. . Q Q L L P F S Q L A A V S P A . A F L T Q . .	168-187
Z19	. . Q Q L L P F Y L H T A P N V G . T L L Q . . .	188-206
Z19	. L Q Q L L P F D Q L A L T N P A . A F Y	207-225
	C-TERMINAL TURN	
Z22	Q Q Q L L P Y N R F S L M N P V L S R Q Q P I V G G A I F	238-266
Z19	. . . . . Q Q P I I G G A L F	226-235

FIG. 2. **The amino acid sequences for the zein proteins with apparent molecular weights of 19,000 (Z19) and 22,000 (Z22).** The residues are divided according to the four structural segments discussed in the text. The amino acids are aligned to achieve the greatest number of identities between Z19 and Z22. The sequence numbers of the residue spans are shown in the *rightmost column*; they correspond to a residue count which includes signal sequences.

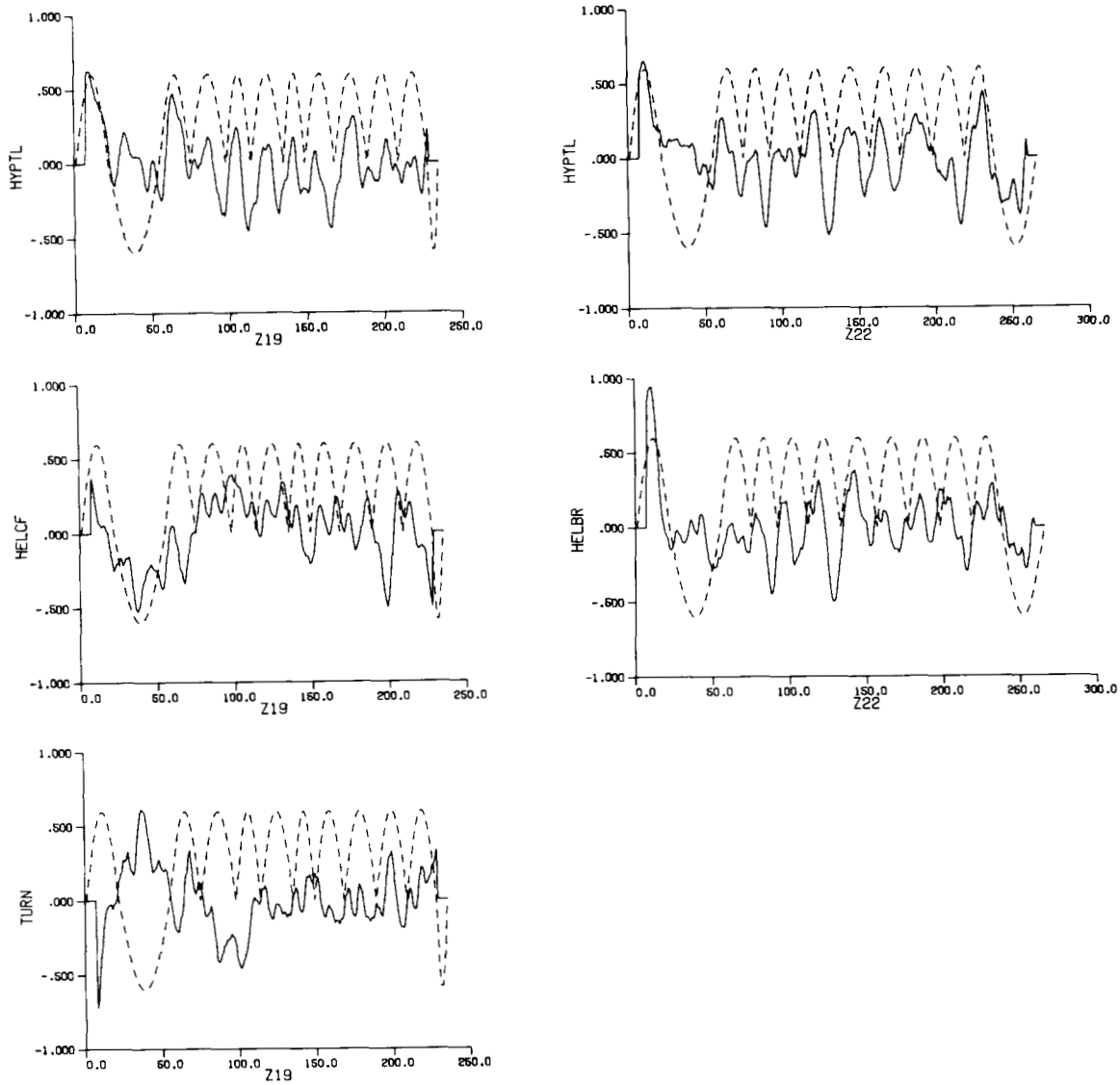

FIG. 3. **Plots of the Z19 or Z22 amino acid sequence number *versus* various physical parameters for the respective residues.** The sequence numbering scheme includes the signal sequence spans. The curve has undergone three cycles of smoothening according to Rose and Siddhartha (1969). The *dashed line peaks* correspond to the homologous repeat units in Z19 and Z22 (Fig. 2), while the *troughs* refer to the predicted turn regions. The signal sequences at the Z19 and Z22 NH_2 termini are also indicated by *dashed line peaks*. The abbreviations used for the physical parameters are: *HYPTL*, hydration potential (Wolfenden *et al.*, 1979); *TURN*, Chou-Fasman reverse turn conformational preference values (Levitt, 1978); *HELCF*, Chou-Fasman helical potentials (Levitt, 1978); and *HELBR*, membrane-buried helical propensities.[1]

(Fig. 3). Use of the membrane-buried helical propensities results in a similar consistency (Fig. 3). The Chou-Fasman helical preference parameters are derived from known soluble protein structures whose helices typically display hydrophobic and hydrophilic faces which result from the residue side groups that interact, respectively, with the protein interior and the aqueous environment. Since zein molecules are self-aggregating and largely hydrophobic, the soluble protein potentials would not be expected to recognize their helical segments. However, the hydration potential and membrane-buried helical propensities (based on interaction between amino acid side chains and the uniquely apolar membrane environment) would be expected to be sensitive to the repeating helical secondary structures in a hydrophobic protein. Fig. 3 also shows the hydration potential plot for Z22; the strong relationship with the nine homologous repeating units is once again obvious. The membrane-bound signal sequence helix in both Z19 and Z22 is delineated by the largest hydrophobicity peak in the graphs. A plot of the reverse turn conformational preference with sequence number is also illustrated for Z19 in Fig. 3. For this case, troughs in the dashed line curve should align with peaks in the turn potential curve; there appears to be some correspondence. However, since turn regions are characterized by polar and charged residues, the glutamine-rich regions are likely candidates as emphasized in the troughs of the hydration potential plots. The highest turn peak in the Z19 plot corresponds to residues 26–57 at the NH_2 terminus; a short turn span is also indicated at the COOH terminus. The Z22 turn plot displayed a similar NH_2-terminal region, as well as a longer COOH-terminal turn segment (Fig. 2).

Hydrogen-bonding Patterns—A consensus sequence for the 18 Z19 and Z22 homologous repeat units (Fig. 2) was determined by the following conditions: (i) an amino acid type must be utilized in a given position at least four times and (ii) at least nine of the repeat units must contribute an amino acid for a consensus position to be defined. The resulting consensus

residues with numbered positions are

$$\underline{\underline{L}}\ \underline{\underline{Q}}\ \underline{\underline{Q}}\ \frac{\underline{\underline{F}}}{\underline{\underline{L}}}\ \underline{\underline{L}}\ \underline{\underline{P}}\ \frac{\underline{\underline{A}}}{\underline{\underline{F}}}\ \underline{\underline{N}}\ \underline{\underline{Q}}\ \underline{\underline{L}}\ \frac{\underline{\underline{A}}}{\underline{\underline{L}}}\ \frac{\underline{\underline{A}}}{\underline{\underline{V}}}\ \underline{\underline{A}}\ \underline{\underline{N}}\ \underline{\underline{S}}\ \underline{\underline{P}}\ \underline{\underline{A}}\ \underline{\underline{Y}}\ \underline{\underline{L}}\ \underline{\underline{Q}}\ \underline{\underline{Q}}$$

$$-2\ -1\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15\ 16\ 17\ 18\ 19$$

The polar amino acids have been underlined twice. It is proposed that positions −2, −1, and 19 constitute the turn regions between helices composed of amino acids in positions 1 through 18. An interesting mirror symmetry exists in the helical segment about positions 9 and 10 with respect to the polar and hydrophobic residues. In proceeding from position 1 to 18 or from 18 to 1, the polar and hydrophobic residues are encountered in the same sequential manner.

A helical wheel of this 18-residue consensus span shows three polar segments along the helical surface (Fig. 4). Side groups of residues 1 and 12 will be along a line nearly parallel to the helical axis and form a polar segment on the helical surface; similarly, positions 7 and 18 and positions 6 and 13 will yield two other lines approximately parallel to the helical axis. The three lines have a 120° angular relationship. Since the turn regions between helices are only about 4 residues in length, it is reasonable to propose that sequential helices will be spatially adjacent and topologically antiparallel. Two polar side groups in each of the neighboring helices will be properly positioned to form hydrogen bonds. For example, side chains at positions 1 and 12 in one helix can make respective hydrogen bonds with side groups at positions 18 and 7 in the neighboring antiparallel helix. Fig. 5 illustrates one possible pattern of bonds among three adjacent helices; the scheme can be repeated for the next three helices and so on. Fig. 6 displays in projection how nine such antiparallel and sequential helices with three polar lines or segments on the helical surface can be arranged within a distorted cylinder of oval cross-section. The distortion is necessitated by the ability of side groups to pack in the cylindrical interior. The helical axes would parallel that of the cylinder and two of the three polar segments would be utilized in intramolecular hydrogen-bonding contacts. The third polar segment would be available for intermolecular contacts such that a molecular plane of zein

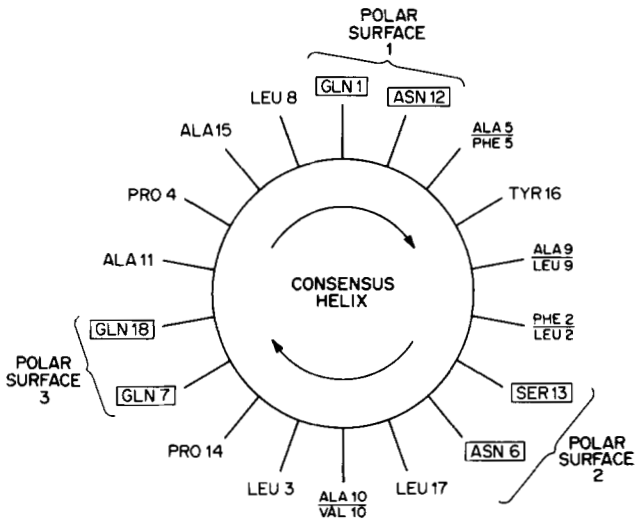

FIG. 4. **Helical wheel (Schiffer and Edmundson, 1967) for the 18-residue consensus repeat sequence in Z19 and Z22.** The *spokes* refer to the amino acid side chain directions in the projected helix. Pairs of residues separated by a line indicate the ambivalent choice of amino acids in the consensus helix as discussed in the text. The figure shows three polar regions each composed of two amino acids whose C_α atoms can be joined by a line nearly parallel to the helix axis.

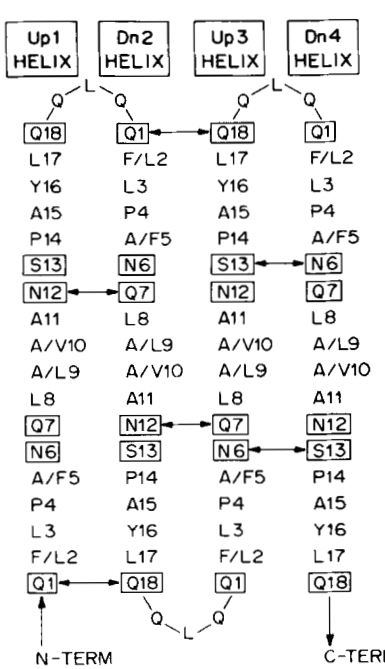

FIG. 5 (*left*). **Possible hydrogen bonding interactions (↔) between polar groups in spatially adjacent, topologically antiparallel consensus helices.** The pair of amino acids within a helix are associated with one of the polar regions indicated on the wheel of Fig. 4.

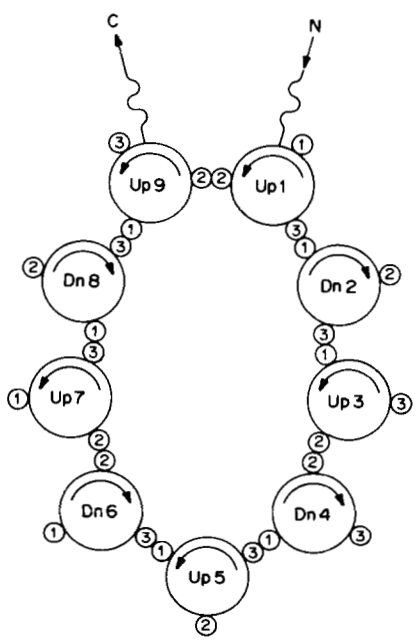

FIG. 6 (*center*). **A possible nine-helical zein protein structural model shown in projection with the helical axes orthogonal to the figure plane.** The hydrogen-bonding polar residue segments (shown as *small circles*) are numbered as in the helical wheel of Fig. 4. The helices are numbered as in Fig. 5. *Up* refers to a helical propagation direction toward the viewer from the NH_2- to COOH terminus while *Dn* indicates the opposite direction. The noninteracting third polar region would be available for hydrogen bonding to helical polar groups in neighboring zein molecules.

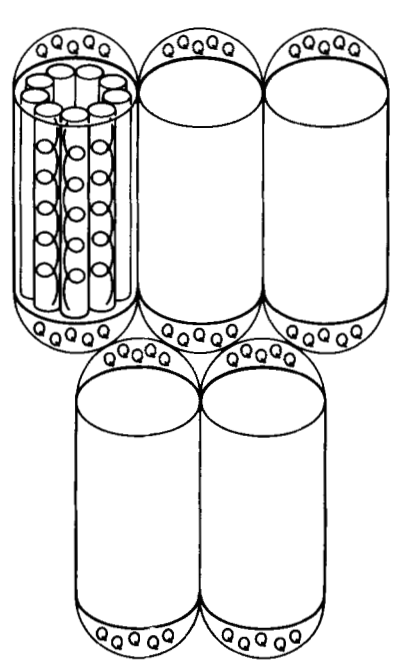

FIG. 7 (*right*). **A possible model for the arrangement of zein proteins within a plane as well as for the stacking of molecular planes.** The glutamine residues (*Q*) would allow hydrogen bond interactions among molecules in neighboring planes. An alternative packing model would align the cylindrical axes.

proteins could be formed. The glutamine residues which compose the turn regions between the antiparallel helices would exist at the caps or ends of the cylindrical clusters. Their interaction would allow stacking of the zein molecular planes. One possible arrangement is illustrated in Fig. 7.

Supporting Evidence for the Model—Residues that are within the repeating units shown in Fig. 2 and included in the consensus helix positions 1–18 comprise about 60% of the Z19 or Z22 primary structures. This helical composition is in good agreement with that indicated by a Greenfield-Fasman analysis of the CD data.

Sedimentation and diffusion constant studies (Foster and Edsall, 1945) and viscosity measurements in urea solution (Ermolenko and Ginzberg, 1956) suggest a rod-like or prolate ellipsoidal shape for the zein molecules. Films of zein proteins cast from ethanol solutions can be formed and yet retain their α-helical structure (Kretschmer, 1957). These observations are consistent with the cylindrical structural model proposed here. A rough calculation based on the standard geometry of α-helices shows a length to width ratio of about 2:1 for the zein model. The film-forming ability of the molecules can be explained through glutamine interactions resulting in stacked zein molecular planes.

The amino acids most frequently appearing in the hydrophobic surfaces of the suggested helices are alanine, leucine, phenylalanine, and valine. It is noteworthy that these five amino acids rank in the top 7 residues with the highest membrane-bound preference parameters (Table I). The ready absorption of zein proteins to membrane fractions and their self-association properties are consistent with this observation. The bacteriorhodopsin structure (Engelman *et al.*, 1980; Engelman and Zaccai, 1980) which is largely buried within a hydrophobic lipid bilayer has a structural motif similar to that proposed for the zein proteins. The bacteriorhodopsin architecture consists of seven spatially adjacent, sequential, and topologically antiparallel helices contained within a distorted cylinder that traverses the lipid bilayer. The residues facing the membrane are similar to those found in the proposed hydrophobic surfaces of the zein helices.

The zein consensus helix contains two prolines near the NH_2 and COOH termini. It is generally agreed that proline is a helix-breaking amino acid in soluble protein structures. However, proline has been observed to occur frequently at helical NH_2-terminal positions (Chou and Fasman, 1978; Argos and Palau, 1982). This is not the case for COOH-terminal sites although there are examples of proline within helices of known protein structures (Argos and Palau, 1982). Proline is also a frequent constituent of the membrane-buried helices in signal sequences (Austen, 1979) and the bacteriorhodopsin structural model (Engelman *et al.* (1980)) which has a strong basis in experimental fact; namely, high resolution electron diffraction studies (Henderson and Unwin, 1975) and neutron scattering experiments (Engelman and Zaccai, 1980). The proline residues tend to "kink" some of the helices in bacteriorhodopsin while generally preserving the overall parallel nature of their axes. A similar distortion may also exist in the zein molecules.

Within the homologous repeat sequence region of Z22 (Fig. 2) and Z22.3 (Marks and Larkins, 1982), there are 18 amino acid exchanges. It would be expected that by chance seven of these substitutions would occur within the helical polar regions and flanking turn spans (consensus helix positions −2, −1, 1, 6, 7, 12, 13, 18, and 19); however, only two exchanges are observed in the strongly polar positions, suggesting their structural importance in the preservation of interhelical hydrogen bonding. Furthermore, zein protein 19.1 displays a 32-residue insertion at position 115 of Z19 and a deletion of 19 amino acids corresponding to Z19 residues 122–140 (Pedersen *et al.*, 1982). It is noteworthy that the insertions and deletions maintain the repeat sequence pattern allowing both proteins an integral number of helices as required by the structural model.

It must be emphasized that the zein structural model proposed here is speculative; however, the "numerology" is compelling. The model represents only an idealized structure as some of the repeats are shorter than the consensus unit requiring some structural alterations. Nonetheless, the deletions are generally 3 or 4 residues in length which would correspond to about one turn of an α-helix. This integral turn deletion would still preserve the spatial relationship among the remaining polar residues and thus not disturb the basic hydrogen-bonding pattern.

TABLE I

Scaled and normalized physical parameters used in this study for the 20 amino acids

Amino acid	Hydration potential	Chou-Fasman helical potential	Chou-Fasman turn potential	Membrane-buried helical potential
Met	1.47	1.43	0.41	2.96
Leu	2.06	1.26	0.62	2.93
Phe	1.58	1.04	0.61	2.03
Ile	2.03	0.94	0.54	1.67
Ala	2.00	1.25	0.82	1.56
Cys	1.51	1.08	0.84	1.23
Val	2.00	0.88	0.49	1.14
Trp	0.79	0.96	0.79	1.08
Thr	0.94	0.80	1.08	0.91
Ser	0.92	0.80	1.40	0.81
Pro	0.93	0.50	2.01	0.76
Tyr	0.75	0.70	1.11	0.68
Gly	2.07	0.54	1.73	0.62
Gln	0.25	1.23	1.02	0.51
Arg	0.00	0.93	0.93	0.45
His	0.12	1.18	0.73	0.29
Asn	0.20	0.87	1.35	0.27
Glu	0.13	1.40	1.05	0.23
Lys	0.23	1.19	1.01	0.15
Asp	0.01	1.01	1.48	0.14

CONCLUSIONS

Empirical observations show the zein proteins to contain repeat sequences, to be hydrophobic and possess a rod-like structure with high helical content, and to self-aggregate, form fibers, and associate with membranes. The zein structural model presented here accounts for these phenomena. A cluster of α-helices corresponding to the zein homologous repeat units and contained within a distorted cylindrical surface capped by polar glutamine residues would allow hydrogen-bonding patterns and hydrophobic van der Waals interactions to maintain zein structural integrity as well as intermolecular packing.

Acknowledgment—Marilyn Anthony was essential in the preparation of this manuscript.

REFERENCES

Argos, P., and Palau, J. (1982) *Int. J. Peptide Protein Res.* **19,** 380–393

Austen, B. M. (1979) *FEBS Lett.* **103,** 308–313

Brahms, S., and Brahms, J. (1980) *J. Mol. Biol.* **138,** 149–178

Chang, C. T., Wu, C.-S. C., and Yang, J. T. (1978) *Anal. Biochem.* **91,** 13–31

Chen, Y., Yang, J. T., and Chau, K. H. (1974) *Biochemistry* **13,** 3350–3359

Chothia, C. (1976) *J. Mol. Biol.* **105,** 1–14

Chou, P. Y., and Fasman, G. D. (1974a) *Biochemistry* **13,** 211–222

Chou, P. Y., and Fasman, G. D. (1974b) *Biochemistry* **13,** 222–245

Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47,** 45–148

Dailey, H. A., and Strittmatter, P. (1981) *J. Biol. Chem.* **256,** 3951–3955
Danzer, L. A., Ades, H., and Rees, E. D. (1975) *Biochim. Biophys. Acta* **386,** 26–31
Engelman, D. M., and Zaccai, G. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 5894–5898
Engelman, D. M., Henderson, R., McLachlan, A. D., and Wallace, B. A. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2023–2027
Ermolenko, N. F., and Ginzberg, D. A. (1956) *Vestn. Akad. Nauk Beloruss. SSSR Ser. Fiz. Tekh. Nauk* **2,** 91
Foster, J. F., and Edsall, J. T. (1945) *J. Am. Chem. Soc.* **67,** 617–625
Geraghty, D., Peifer, M. A., Rubenstein, I., and Messing, J. (1981) *Nucleic Acids Res.* **9,** 5163–5174
Gianazza, E., Viglienghi, V., Righetti, P. C., Salamini, F., and Soave, C. (1977) *Phytochemistry* **16,** 315–317
Greenfield, N., and Fasman, G. D. (1969) *Biochemistry* **8,** 4108–4116
Henderson, R., and Unwin, P. N. T. (1975) *Nature (Lond.)* **257,** 28–32
Hurkman, W. J., Smith, L. D., Richter, J., and Larkins, B. A. (1981) *J. Cell Biol.* **89,** 292–299
Kretschmer, C. B. (1957) *J. Phys. Chem.* **61,** 1627–1631
Larkins, B. A., and Hurkman, W. J. (1978) *Plant Physiol.* **62,** 256–263
Lee, K. H., Jones, R. A., Dalby, A., and Tsai, C. Y. (1976) *Biochem. Genet.* **14,** 641–650
Levitt, M. (1978) *Biochemistry* **17,** 4277–4285
Marks, M. D., and Larkins, B. A. (1982) *J. Biol. Chem.* **257,** 9976–9983
Moews, P. C., Knox, J. R., Waxman, D. J., and Strominger, J. L. (1981) *Int. J. Peptide Protein Res.* **17,** 211–218
Pedersen, K., Devereux, J., Wilson, D. R., Sheldon, E., and Larkins, B. A. (1982) *Cell* **29,** 1015–1026
Provencher, S. W., and Glockner, J. (1981) *Biochemistry* **20,** 33–37
Rees, E. D., and Singer, S. J. (1955) *Nature* **176,** 1072–1073
Rees, E. D., and Singer, S. J. (1956) *Arch. Biochem. Biophys.* **63,** 144–159
Righetti, P. G., Gianazza, E., Viotti, A., and Soave, C. (1977) *Planta* **136,** 115–123
Rose, G. E., and Siddhartha, R. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 4643–4647
Schiffer, M., and Edmundson, A. B. (1967) *Biophys. J.* **7,** 121–135
Siegel, J. B., Steinmetz, W. E., and Long, G. L. (1980) *Anal. Biochem.* **104,** 160–167
Tomita, M., Furthmayr, H., and Marchesi, V. T. (1978) *Biochemistry* **17,** 4756–4770
von Heijne, G. (1981) *Eur. J. Biochem.* **116,** 419–422
Wall, J. S., and Paulis, J. W. (1978) in *Advances in Cereal Science and Technology* (Pomeranz, Y., ed) Vol. 2, pp. 135–219, American Association of Cereal Chemists,
Wolfenden, R. V., Cullis, P. M., and Southgate, C. C. F. (1979) *Science* **206,** 575–577