New Insight into the Solution Structures of Wheat Gluten Proteins from Raman Optical Activity†

Ewan W. Blanch,*,‡,§ Donald D. Kasarda,|| Lutz Hecht,‡ Kurt Nielsen,⊥ and Laurence D. Barron*,‡

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom, United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan Street, Albany, California 94710, and Department of Chemistry, DTU 207, Technical University of Denmark, DK-2800 Lyngby, Denmark

Received October 23, 2002; Revised Manuscript Received March 7, 2003

ABSTRACT: Vibrational Raman optical activity (ROA) spectra of the wheat proteins α-gliadin (A-gliadin), *ω*-gliadin, and a 30 kDa peptide called T-A-1 from the high molecular weight glutenin subunit (HMW-GS) Dx5 were measured to obtain new information about their solution structures. The spectral data show that, under the conditions investigated, A-gliadin contains a considerable amount of hydrated α-helix, most of which probably lies within a relatively structured C-terminal domain. Smaller quantities of *β*-structure and poly(L-proline) II (PPII) helix were also identified. Addition of methanol was found to increase the α-helix content at the expense of some of the *β* and PPII structure. In comparison, *ω*-gliadin and the T-A-1 peptide were found to consist of large amounts of well-defined PPII structure with some turns but no α-helix. The results for the T-A-1 peptide are in agreement with a model in which HMW-GS are extended but not highly rigid. Application of a pattern recognition technique, based on principal component analysis (PCA), to the ROA spectra reinforces these conclusions.

The seed storage proteins, or gluten proteins, of wheat are the main determinants of the physicochemical properties of wheat flour doughs. These gluten proteins can be subdivided into two approximately equal groups based on their solubility in alcohol-water solutions, with the alcohol-soluble proteins called gliadins and the alcohol-insoluble component called glutenins. Both may be classed as prolamins because of their high proportions of proline and glutamine. Prolamins form insoluble deposits in the lumen of the endoplasmic reticulum or Golgi vesicles. They serve a storage function and are quickly broken down by proteolytic enzymes upon germination of the seed. The gliadins are essentially monomeric proteins whereas in glutenins, protein subunits are linked by intermolecular disulfide bonds to form a polymeric system (*1*, *2*).

There are four types of gliadins called α-, *β*-, *γ*-, and *ω*-gliadins on the basis of electrophoretic mobilities at acid pH in nondenaturing gels. They range in mass from 30 to 50 kDa. Each type consists of multiple components that can be grouped together on the basis of overall structure. The α-gliadins may be thought of as having two major domains, an N-terminal domain and a C-terminal domain, although other divisions have been proposed (*3*). The N-terminal domain consists mainly of degenerate repeating sequences related to the sequence Pro-Phe-Pro-$(Gln)_{3-6}$ (*4*), followed by a predominantly polyglutamine region. The C-terminal domain includes all six cysteine residues that form three intramolecular disulfide bonds, along with a second short polyglutamine sequence. Kasarda and co-workers (*5*−*10*) have extensively studied a purified fraction of α-gliadins known as A-gliadin that can form microfibrillar structures with diameters of about 0.8 nm and 300−400 nm in length depending on pH (5 or higher) and ionic strength (5 mM KCl); the microfibrils are similar in appearance to amyloid fibrils, but the proteins do not undergo major conformational changes during aggregation. These microfibrils can be disaggregated into monomers and reaggregated by varying the pH or ionic strength (*6*).

The *β*-gliadins are closely similar in structure to the α-gliadins, and the combined class has usually been referred to either as α-type or α/*β*-type. Both α- and *β*-gliadins have essentially identical N-terminal sequences and similar domain structures. The *γ*-gliadins have a similar domain structure to α- and *β*-gliadins although the C-terminal domain usually includes four intramolecular disulfide bonds instead of three. The *γ*-gliadins have a unique N-terminal sequence, and the repeating sequence has the consensus structure (Pro-Phe-Pro-Gln)$_{1-2}$ (Pro-Gln-Gln)$_{1-2}$ (*11*).

The *ω*-gliadins are distinct from the α-, *β*- and *γ*-gliadins, which are structurally related, and cluster around three average sizes: the 1A group of 34−39 kDa, the 1B group of ∼50 kDa and the 1D group of ∼40 kDa (refs *12* and *13* and personal communication from Frances Dupont). The *ω*-gliadins consist principally of proline (up to 30 mol %) and glutamine (up to 55 mol %) residues and are made up almost entirely of repeating sequences, which, in the case of the 1A and 1D *ω*-gliadins, have the consensus Pro-Phe-(Pro-Gln-Gln)$_2$. The 1B *ω*-gliadins have a slightly different

† This work was supported by a research grant from the BBSRC to L.D.B. and L.H.

* To whom correspondence should be addressed. E-mail: (E.W.B.) E.Blanch@umist.ac.uk or (L.D.B.) laurence@chem.gla.ac.uk.

§ Current address: Department of Biomolecular Sciences, UMIST, P.O. Box 88, Manchester M60 1QD, UK.

‡ University of Glasgow.

|| United States Department of Agriculture.

⊥ Technical University of Denmark.

10.1021/bi027059y CCC: $25.00
Published on Web 04/25/2003

```
147 YY-PTS-PQQ-
    PGQ-LQQ-
    PAQ-GQQ-
    PGQ-GQQ-GQQ-
    PGQ-GQP-GYY-
    PTS-SQL-Q  -
    PGQ-LQQ-
    PAQ-GQQ-GQQ-
    PGQ-AQQ-GQQ-
    PGQ-GQQ-
    PGQ-GQQ-GQQ-
    PGQ-GQQ-
    PGQ-GQQ-GQQ-
    LGQ-GQQ-GYY-
    PTS-LQQ-
    SGQ-GQP-GYY-
    PTS-LQQ-
    LGQ-GQS-GYY-
    PTS-PQQ-
    PGQ-GQQ-
    PGQ-LQQ-
    PAQ-GQQ-
    PGQ-GQQ-GQQ-
    PGQ-GQQ-GQQ-
    PGQ-GQQ-
    PGQ-GQP-GYY-
    PTS-PQQ-
    SGQ-GQP-GYY-
    PTS-SQQ-PT -QYY-
    PTS-PQQ-
    SGQ-GQP-GYY-
    LTS-PQQ-
    SGQ-GQQ-
    PGQ-LQQ-
    SAQ-GQK- 440
```

FIGURE 1: Amino acid sequence of the T-A-1 peptide, corresponding to the sequence of residues 147−440 of HMW-GS Dx5, arranged to emphasize the triplet structure.

consensus repeat, Phe-Pro-$(Gln)_{2-4}$ (*13*). As they contain no cysteine residues, the *ω*-gliadins are also known as S-poor prolamins (*14*).

High molecular weight glutenin subunits (HMW-GS)[1] constitute ∼10 wt % of wheat gluten and have a number of different allelic forms, usually four or five of them expressed in a given wheat cultivar (*15*), which play a significant role in determining the quality of wheat flour (*16*, *17*). The HMW-GS can be classified, depending on their molecular mass and primary sequence, as either *x*-type, with a mass of 83−88 kDa, or *y*-type, with a mass of 67−74 kDa. Both types of HMW-GS have N- and C-terminal domains of about 100 and 50 residues, respectively, and a large central domain of 480−690 residues made up of repeating sequences similar to the following (albeit with considerable variation): in *x*-types, Pro-Gly-Gln-$(Gly\text{-}Gln\text{-}Gln)_{1-2}$ and Pro-Gly-Gln-Gly-Gln-Pro-Gly-Tyr-Tyr-Pro-Thr-Ser-Pro-Gln-Gln; and in *y*-types, Pro-Gly-Gln-Gly-Gln-Gln and Pro-Gly-Gln-Gly-Gln-Gln-Gly-Tyr-Tyr-Pro-Thr-Ser-Pro-Gln-Gln. These repeating sequences can also be arranged to display sextet, and to some degree, triplet elements (Figure 1). In common with the other wheat proteins studied here, HMW-GS are rich in glutamine (37−39 mol %) and proline (12−14 mol %), but in addition they are rich in glycine (14−19 mol %). They also contain several cysteine residues in or near the N- and C-terminal domains that can form intramolecular or intermolecular disulfide bonds (*2*, *18*, *19*), the intermolecular disulfide bonds being the basis for the polymeric system that provides elasticity to a wheat flour dough. Strong hydrogen-bonding interactions are likely to occur between the amide side chains of glutamine residues in the central domain and other gluten proteins or intramolecularly with main chain amides and other glutamine side chains (*20*). These interactions contribute to the viscoelasticity of doughs and may have structural implications (*21*−*23*).

Despite the glutens being among the first proteins to be studied, the structures of these cereal grain proteins are not yet satisfactorily defined because they have so far not been crystallized for X-ray diffraction analysis. Also, they do not lend themselves to NMR structure determination for a number of reasons: they are fairly large, have redundancies of sequence because of the extensive repeats, and have large numbers of prolines, which tend to flip back and forth between cis and trans conformations and also have no protons on the imido nitrogen. A number of spectroscopic and other techniques have provided useful information, but many of the details of the secondary, tertiary, and quarternary structures of gluten proteins are still unknown. In particular, the degree to which certain domains of the gluten proteins, especially those containing intramolecular disulfide bonds, are highly structured is not known; IR studies have provided some indication that these domains may be loosely folded (*10*).

A novel form of Raman spectroscopy called Raman optical activity (ROA), which measures small differences in the vibrational Raman spectra of chiral molecules in right- and left-circularly polarized incident light (*24*, *25*), has recently been applied to studies of the structure of biomolecules in solution (*26*). ROA bears the same relation to conventional Raman spectroscopy as ultraviolet circular dichroism (UVCD) does to conventional ultraviolet absorption spectroscopy. ROA is sensitive to chirality and thereby is more incisive than conventional vibrational spectroscopy in the study of biomolecules such as proteins because only the few vibrational coordinates that most directly sample the skeletal chirality, such as those in the peptide backbone, make the largest contributions to the ROA spectrum. This leads to ROA band patterns being simpler and more sensitive to conformation than those in the parent Raman spectrum in which the bands from the peptide backbone are often obscured by those from the amino acid side chains. As well as bands arising from secondary structure and side chains, protein ROA spectra contain bands from loops and turns that can provide information about the tertiary fold of the peptide backbone. Since ROA has proven especially useful for the study of partially or fully unfolded proteins (*27*−*29*), it was used here to investigate the solution structures of wheat α-(A)-gliadin, wheat *ω*-gliadin, and a peptide called T-A-1 derived by tryptic digestion from the HMW-GS Dx5.

EXPERIMENTAL PROCEDURES

Sample Preparation. The sample of A-gliadin was prepared according to the method of Bernardin et al. (*5*). The *ω*-gliadin sample, prepared according to the method of Kasarda et al. (*12*), was a mixture of *ω*-1 and *ω*-2 components of the 1D group. These two *ω*-gliadin 1D proteins are likely to

[1] Abbreviations: CCD, charge coupled device; Da, Daltons; FTIR, Fourier transform infrared spectroscopy; HMW-GS, high molecular weight glutenin subunit; HPLC, high performance liquid chromatography; MALDI, matrix assisted laser desorption/ionization time-of-flight; PCA, principal component analysis; PHD, profile-based neural network prediction of secondary structure and solvent accessibility; PPII, poly(L-proline) II; ROA, Raman optical activity; SDS−PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; UVCD, ultraviolet circular dichroism.

differ only by the larger having an additional eight residues at the N-terminus (ref *12* and unpublished data, D. D. Kasarda). The sample of the T-A-1 peptide fragment corresponds to the sequence from residues 147−440 of HMW-GS Dx5 and was prepared as follows. Flour from the wheat cultivar Cheyenne was extracted first with 0.1 M sodium chloride and then 50% *n*-propanol/50% water (v/v). The remaining residue, as a slurry in 0.1 M ammonium carbonate, titrated to pH 7.5, was then subjected to proteolytic digestion with trypsin using an enzyme/substrate ratio of ∼100:1 for 2 h at room temperature with gentle stirring provided by a magnetic stirrer. After centrifugation, digestion was stopped by the addition of acetic acid to the clarified supernatant solution until a pH of 5.5 was achieved. The supernatant was then freeze-dried in small batches, which were used for gel filtration fractionation on a Biogel P-30 column (BioRad) with 0.1 M acetic acid as the eluant. The T-A-1 peptide was contained in the peak eluting at the void volume of the column (designated Peak A), which was then freeze-dried directly from 0.1 M acetic acid. This fraction was subjected to reverse-phase HPLC on a C-18 Vydac semipreparative column with the first peak (designated Peak 1) being found to contain the desired peptide in essentially pure form. Identification of the peptide was by way of N-terminal protein sequencing and mass determination of 30 727 Da by MALDI mass spectrometry for comparison with the complete sequence of Dx5 (*30*). The sequence of the T-A-1 peptide derived from these analyses and the DNA-based sequence is shown in Figure 1. Sample solutions were prepared using either dilute HCl or NaOH. A digital pH meter was used to determine the final pH of each solution.

ROA Spectroscopy. The Raman and ROA measurements were carried out using an instrument described previously (*31*). It utilizes a backscattering geometry, which is essential for measurements on biomolecules, and employs a single-grating spectrograph fitted with a backthinned charge coupled device (CCD) camera as detector. An edge filter is used to block the Rayleigh line. The ROA spectra are accumulated over time by synchronizing the acquisition of the Raman spectra with an electrooptic modulator that switches the polarization of the incident argon-ion laser beam between right- and left-circular forms at a suitable rate. The spectra are displayed in analog-to-digital counter units as a function of the Stokes Raman wavenumber shift with respect to the excitation wavelength. The ROA spectra are presented as the circular intensity differences, $I^R - I^L$, while the parent Raman spectra are presented as the circular intensity sums, $I^R + I^L$, where I^R and I^L are the Raman backscattered intensities in right- and left-circularly polarized incident light, respectively. The experimental conditions were as follows: laser wavelength 514.5 nm; laser power at the sample ∼700 mW; spectral resolution ∼10 cm^{-1}; protein and peptide concentrations ∼20−50 mg/mL; and acquisition times ∼30−60 h (depending on the sample). Scale bars are shown in each Raman and ROA spectrum to represent the scattering intensities. These scale bars have not been normalized to any standard set of conditions, and the scattering intensities shown reflect differences between experimental conditions. Therefore, the absolute values of the scattering intensities shown are not a reliable means of comparing different spectra.

Dynamic Light Scattering. A DynaPro-801 Dynamic Light Scattering instrument was used to measure the hydrodynamic radii of the T-A-1 peptide and A-gliadin in solution. For the dynamic light scattering measurements, the protein concentrations were 3−4 mg/mL, and solutions were at the same pH and temperature as used for the ROA measurements. Under these conditions, the T-A-1 peptide and A-gliadin solutions were found to be monodispersed. It was not possible to determine the hydrodynamic radius of *ω*-gliadin as this protein formed polydispersed solutions with even the species with lowest molecular weight present appearing to be an oligomer.

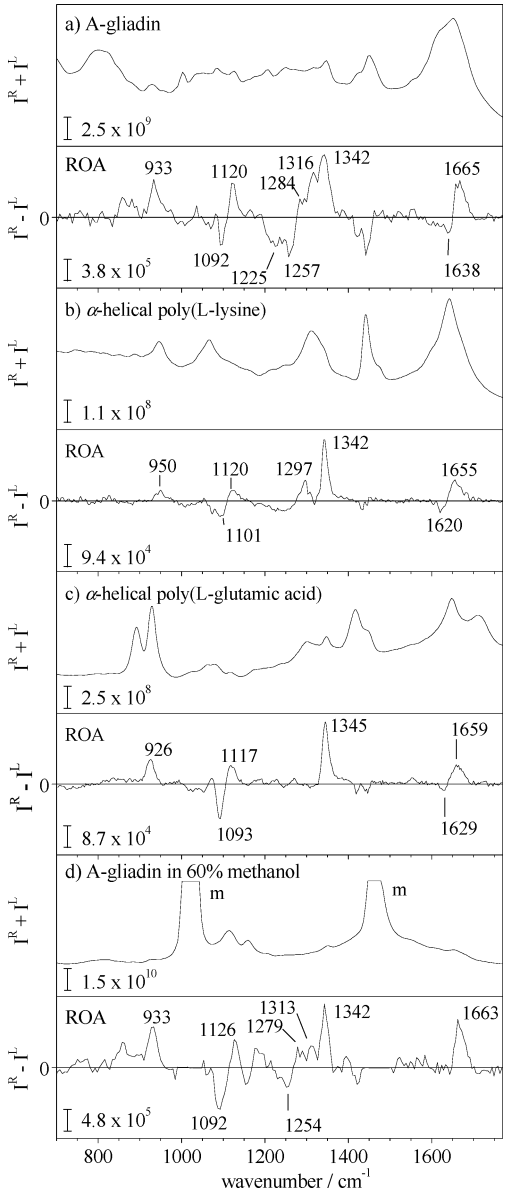

FIGURE 2: Backscattered Raman ($I^R + I^L$) and ROA ($I^R - I^L$) spectra of (a) A-gliadin at pH 3.5, (b) the α-helical form of poy(L-lysine) at pH 11.0, (c) the α-helical form of poy(L-glutamic acid) at pH 4.8, and (d) A-gliadin in 60% methanol at pH 3.6. The strong bands marked m in the latter Raman spectrum are due to the methanol solvent, and no reliable signals were recorded in these regions in the corresponding ROA spectrum.

RESULTS AND DISCUSSION

A-Gliadin. The backscattered Raman and ROA spectra of A-gliadin (an α-type gliadin) are shown in Figure 2a. In the ROA spectrum of A-gliadin, the largest band is the positively signed feature at ∼1342 cm^{-1} in the extended amide III

region (which originates mainly from N−H and C_α−H deformations and the C_α−N stretch). This band at ∼1342 cm^{-1} is very similar to the dominant bands found in the ROA spectra of the α-helical forms of poly(L-lysine) and poly(L-glutamic acid) (*26*) as shown in Figure 2b,c, which can be obtained when the side chains of these homopolypeptides are neutralized. Positive ROA bands in the region of ∼1340−1345 cm^{-1} are characteristic of α-helix in a hydrated state. There is no clear positive ROA band in the region of ∼1295−1310 cm^{-1}, which we assign to α-helix in a more hydrophobic environment. Electron spin resonance studies of double spin-labeled alanine-rich peptides (*32*, *33*) have provided new insight into the natures of these hydrated and hydrophobic variants of α-helix. A new and more open conformation of the α-helix was identified from these studies, and computer modeling suggests that in this conformation the hydrogen-bonding network is intact but that the C−O···N angle is changed. This results in the backbone amide carbonyls being splayed away from the helix axis and into solution. Therefore, this more open conformation may be the preferred one in aqueous solution since it allows main-chain hydrogen bonding with water molecules. The ROA couplet in the amide I region (originating mainly in the peptide C═O stretch), negative at ∼1638 cm^{-1} and positive at ∼1665 cm^{-1}, and the positive band in the backbone skeletal stretch region (originating mainly from C_α−C, C_α−C_β, and C_α−N stretch modes) at ∼933 cm^{-1} are also characteristic of α-helix. Although both FTIR and UVCD have indicated significant amounts of α-helix in A-gliadin (*7*, *10*), these techniques do not differentiate between hydrated and hydrophobic forms. Further discussion of the principal ROA band assignments used here can be found elsewhere (*26*).

In the extended amide III region there is a small negative band at ∼1225 cm^{-1} that we assign to a small content of β-structure, possibly in the form of hydrated single β-strands (*26*). Large numbers of prolines makes the formation of long runs of residues in the β-sheet or β-strand conformation unlikely (*34*). There is also a couplet, negative at ∼1257 cm^{-1} and positive at ∼1284 cm^{-1}, which may originate in turn structures, but we are unable at this time to discriminate between different turn types. A small but clear positive band at ∼1316 cm^{-1} indicates the presence of a small amount of poly(L-proline) II, or PPII, structure. Similar bands are observed in the ROA spectra of polypeptide models known to contain significant amounts of PPII structure, as shown in Figure 3b,c, and are discussed in more detail later.

The ∼127 residue N-terminal domain of A-gliadin contains various repeat sequences, and there are a total of 25 proline residues within the N-terminal domain sequence. Prolines tend to disrupt α-helices as the imide nitrogen breaks at least two adjacent hydrogen bonds, and the constrained pyrrolidine ring causes a kink in the polypeptide backbone (*35*). It is therefore likely that the N-terminal domain contains little of the identified α-helix, although this domain probably contains small but significant amounts of PPII structure and turns—proline often being found in turns as well as in PPII helix. However, within the ∼127-residue N-terminal domain there is a sequence from residue 96 to 122 that contains no prolines but that consists mainly of glutamine residues, including an 18-residue polyglutamine sequence. Glutamine residues have been found to favor the α-helical region of the Ramachandran surface for a set of refined nonredundant protein structures from the Protein Data Bank (*36*), but recent biophysical studies of polyglutamine peptides indicate that random coil and β-strand are more likely for such peptides (*37*−*40*), and glutamine residues have also been reported to have a high PPII helix-forming propensity (*23*). Consequently, no firm conclusion can be drawn regarding the structure of the polyglutamine stretch in A-gliadin. Furthermore, the ultraviolet circular dichroism (UVCD) spectra of polyglutamine peptides attributed to random coil (*41*) might well correspond to some degree of PPII structure, which has a similar UVCD spectrum to that of random coil (*39*).

Secondary structure prediction methods can provide useful information about the local conformation propensities of residues within proteins. Analysis of the sequence of A-gliadin by the PHD program of Rost and Sander (*42*) predicts that most regions within the N-terminal domain, which contain a large number of prolines, have only a small or negligible propensity to form α-helix or β-strand. Proline-rich sequences are much rarer in the C-terminal domain of A-gliadin, which contains the six cysteine residues of the protein that form three specific intramolecular disulfide bonds (*43*). The C-terminal domain is predicted by the PHD program to contain more stretches of α-helix and β-strand than the N-terminal domain. FTIR spectroscopy supports the likelihood that most α-helical regions are found in the C-terminal domain (*10*). The PHD program also predicts that the residues within the C-terminal domain have, on average, a 3-fold lower accessibility to the solvent than those in the N-terminal domain.

We attempted to obtain an ROA spectrum of a cyanogen bromide-cleaved peptide of A-gliadin (CN-1) (*3*) corresponding to the N-terminal domain, but this was not possible, as the peptide formed a gel, inaccessible to ROA spectroscopy, under conditions where the intact A-gliadin had been stable in solution. This suggests that the more structured C-terminal domain helps to counteract the tendency for the less compactly structured N-terminal domain to undergo aggregation in solution. In a UVCD study of α-gliadin (*7*), a loss in mean residue ellipticity near 280 nm upon changing the pH from 5 to 3 indicated a conformational change that increased the mobility of the tyrosine or tryptophan side chains. This was accompanied by a small decrease in α-helical content. Measurements of the intrinsic viscosity of A-gliadin in the range of pH 3−5 suggested that the monomer has a relatively compact conformation at pH 4 or higher. At pH 4, although the protein was in an equilibrium between the monomer and a slightly aggregated form incorporating on the average about 20 protein subunits, the intrinsic viscosity was indicative of a compact conformation, possibly for both forms, whereas the somewhat higher viscosity at pH 3 may be a consequence of the N-terminal domain becoming extended or flexible for the essentially monomeric protein at this pH (*9*).

As noted above, the α-helical structure of A-gliadin, which is likely to be found mainly in the C-terminal domain, appears to be in the hydrated form with no α-helix in a hydrophobic environment. We have performed dynamic light scattering measurements under the same conditions of pH and temperature at which the ROA spectrum was recorded, although at lower concentration, and found the hydrodynamic radius of A-gliadin to be 2.5 nm. This corresponds to the

hydrodynamic radius expected for a folded globular protein of the same molecular weight (*44*). However, the low mean hydrophobicity of the A-gliadin sequence is characteristic of a natively unfolded protein (*44*, *45*). These data imply that under the conditions of this study (pH 3.5 and ~20 °C), the C-terminal domain lacks the hydrophobic core typical of many natively folded proteins and that solvent water molecules are able to readily penetrate the interior of this domain.

Methanol-Induced α-Helix Formation in A-Gliadin. The ROA spectrum of A-gliadin was also recorded in a 60% methanol/40% aqueous HCl solution at pH 3.6 (v/v) and is shown in Figure 2d. Although no ROA signal is observed in the ~990−1050 and ~1430−1500 cm^{-1} regions because of the saturation of the detector by large Raman bands from the solvent methanol at these wavenumbers, it is evident that the sharp positive peak at ~1342 cm^{-1} in the amide III region of the ROA spectrum now dominates the ROA spectrum indicating an increase in the amount of solvated α-helix. From the decrease in the intensities of the bands at ~1316 and ~1225 cm^{-1} originating from PPII helix and hydrated *β*-structure, respectively, it appears that addition of methanol converts approximately half of the PPII structure and most of the *β*-strand of A-gliadin into α-helix. The intensity of the band at ~933 cm^{-1} also appears to increase upon the addition of methanol, which further supports the assertion that there is an increase in α-helical content. Alcohols are known to destabilize the native structure of proteins and to promote the formation of α-helix (*41*, *46*, *47*). The mechanism by which this occurs is not well-understood but may be related to the decrease in the solvent polarity with decrease in water concentration, leading to a decrease in the strength of hydrophobic interactions between hydrophobic regions of the protein. The decrease in solvent dielectric constant also probably leads to an enhancement of intrapeptide hydrogen bonding. This is the first ROA data reported for a polypeptide dissolved in a methanol solution. Therefore, it is not clear whether elements of α-helix induced by methanol solvation give rise to vibrational modes, and consequently, ROA band patterns in the extended amide III region significantly different to those characteristic of the hydrated variant of α-helix. However, the absence of any new ROA bands assigned to α-helix in this region while all known α-helical marker bands throughout the spectrum appear to increase in proportion to each other upon the addition of methanol indicates that this is probably not the case.

ω-Gliadin. The parent Raman spectrum of *ω*-gliadin exhibits a relatively high level of background fluorescence that reduces the signal-to-noise ratio in the associated measured ROA spectrum, both of which are shown in Figure 3a. However, it is still apparent that the ROA spectrum is dominated by a positive band in the extended amide III region at ~1321 cm^{-1}. A very similar band dominates the ROA spectra of the disordered states of poly(L-lysine) at acidic pH and poly(L-glutamic acid) at alkaline pH, shown in Figure 3b,c for comparison (*26*). As these disordered homopolypeptides are thought to contain significant amounts of left-handed polyproline II (PPII) helix (*48*−*51*), this band at ~1321 cm^{-1} has been assigned to a vibrational mode of the PPII helix, or at least residues that have the appropriate *φ*,*ψ* angles (*26*), indicating that *ω*-gliadin contains large amounts of PPII helix, probably as the predominant type of

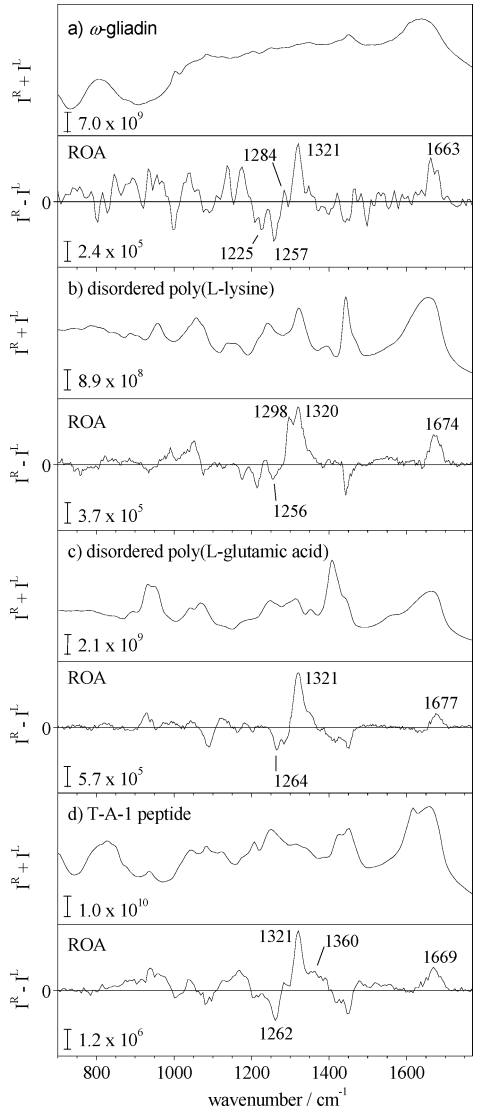

FIGURE 3: Backscattered Raman ($I^R + I^L$) and ROA ($I^R - I^L$) spectra of (a) *ω*-gliadin at pH 2.6, (b) the disordered form of poly(L-lysine) at pH 3.0, (c) the disordered form of poly(L-glutamic acid) at pH 12.6, and (d) the T-A-1 peptide at pH 3.5.

secondary structure. The presence of significant amounts of PPII structure in *ω*-gliadin is in agreement with the results of a UVCD study by Dupont et al. (*13*) who reported that an *ω*-gliadin 1D fraction consisted mainly of flexible coil and PPII structure but possibly with small amounts of *β*-structure and some turns, either *β*- or *γ*-type. The ROA couplet in the amide III region, negative at ~1257 cm^{-1} and positive at ~1284 cm^{-1}, and the negative band at ~1225 cm^{-1} also indicate the presence of turns and hydrated *β*-structure, respectively. However, the weak positive amide I band at 1663 cm^{-1} indicates less extended structure than found for A-gliadin. The spectra of Dupont et al. (*13*) lacked a positive band at approximately 220 nm, and the negative band at ~200 nm was weaker than would be expected if the protein was primarily in the PPII form, but a possible explanation is that a large proportion of the residues in the protein have *φ*,*ψ* angles close to those characteristic of PPII, but the segment length of the helix is short (*52*).

T-A-1 Peptide. Figure 3d shows the backscattered Raman and ROA spectra of the T-A-1 peptide. This peptide corresponds to a 294 amino acid sequence from the large

central domain of HMW-GS Dx5 and consists almost entirely of variations on several different glutamine-, proline-, and glycine-rich repeating sequences. We have arbitrarily arranged the sequence shown in Figure 1 in a nontraditional way to illustrate the likelihood of there being a three-residue basis to the overall repeating sequence, although the elements of the traditional 6-, 9-, and 15-residue repeats are unquestionable. The amide III region of the ROA spectrum of the T-A-1 peptide is dominated by a sharp positive band at $\sim$1321 cm^{-1}. We therefore conclude that the T-A-1 peptide also contains large amounts of PPII structure. An elemental 3-residue repeat as illustrated in Figure 1 might promote formation of PPII helix, which is based on 3-residue turns. Similar ROA bands have been reported for a number of proteins with irregular, natively unfolded, and non-native folds (*26*−*28*, *53*) and have also been assigned to residues with the PPII-helical conformation in loop regions of a viral coat protein (*54*). The ROA spectrum of disordered poly(L-lysine), but not those of disordered poly(L-glutamic acid) or the T-A-1 peptide, shows a positive band at $\sim$1298 cm^{-1}, which we assign to a propensity for individual residues to assume α-helical ϕ,ψ torsion angles. All three ROA spectra exhibit a broad, weak positive amide I signal at $\sim$1669−1677 cm^{-1} that is consistent with a lack of any extended secondary structure. The ROA spectrum of T-A-1 also bears strong similarities to that of a 17 residue peptide that lies within the disordered N-terminal tail of the sheep prion protein in solution (Blanch, E. W., Gill, A. C., Rhie, A. G. O., Hope, J., Hecht, L., and Barron, L. D., unpublished data) and that is thought to contain significant amounts of PPII helix (*55*).

The negative band at $\sim$1262 cm^{-1} in T-A-1 may originate from residues in β- or γ-turns. However, there may also be a contribution to this band from the heterocyclic ring of proline residues, which constitute $\sim$14 mol % of the peptide since the ROA spectrum of poly(L-proline) in aqueous solution displays a similar feature (Blanch, E. W., Hecht, L., and Barron, L. D., unpublished data). For the same reason, the small positive feature at $\sim$1360 cm^{-1} may originate in the proline residues. The ROA spectrum of the T-A-1 peptide indicates the presence of a large amount of PPII helix, which, although it might be expected to promote an extended structure, is apparently imperfect to some degree, as indicated by a hydrodynamic radius of 4.0 nm as measured by dynamic light scattering under the same conditions of pH and temperature as for the ROA spectra. This corresponds to the hydrodynamic radius expected for a natively unfolded premolten-globular protein of the same molecular weight (*44*). The ROA band at $\sim$1262 cm^{-1} suggests the possibility of some turns being present although the type of turn cannot be specified. Similar bands assigned to elements of PPII structure and β-turns were observed in the ROA spectrum of the Bowman-Birk protease inhibitor (*27*) from soybean (not shown here). This is a small protein 61 residues long, and analysis of the X-ray crystal structure 1pi2 in the Protein Data Bank shows that $\sim$22 of these residues have ϕ,ψ torsion angles characteristic of PPII helix occurring with short segment lengths of two to five residues. Thus, we suggest that T-A-1, like the ω-gliadin, has a very short segment length for the PPII helix.

Studies of a Dy10/Dx5 HMW-GS recombinant hybrid protein overexpressed in wheat (*56*) demonstrated that the expressed protein occurred to a significant extent in a monomeric form in which a single disulfide bond connected cysteine residues in the N- and C-terminal regions. The gene coding for the hybrid protein had been constructed from the DNA coding for the N-terminal domain of HMW-GS Dy10 combined with that coding for the central repeat sequence and C-terminal domains of HMW-GS Dx5. The linkage between the N- and the C-terminal domains in the hybrid was somewhat surprising because the most popular model for the structure of the large intervening repeating sequence domain of HMW-GS corresponds to a rodlike structure in which the polypeptide chain assumes a β-spiral conformation, or at least some form of spiral (*57*−*61*). This unusual disulfide linkage in the Dy10/Dx5 hybrid indicated that the central repeating sequence domain of subunit Dx5 is at least sufficiently flexible to allow close approach between the widely separated domains, which would not occur if the central domain had a continuous rigid-rodlike structure. While this is suggestive of chain flexibility, even a relatively small hinge region might provide the necessary flexibility. Gilbert et al. (*62*), on the basis of UVCD and FTIR spectroscopies, concluded that a somewhat similar peptide corresponding to the central repeating sequence domain of Dx5 contained both PPII helix and β-turns in equilibrium with one another when in solution, such that the former predominated at low temperatures and the latter at higher temperatures. The turn structure appeared to be favored at 22 °C. In contrast, our results indicate a definite predominance of the PPII structure at room temperature. β- or γ-turns might well be present in the T-A-1 structure to some degree without their necessarily participating in a well-defined spiral structure, although we cannot exclude the presence to some degree of the proposed spiral structure. Although we favor a flexible chain resulting from a short segment length for PPII helix, we cannot rule out the possibility of PPII helix interspersed with other flexible regions; this would be in contrast to a rigid-rodlike structure in which the repeats exist in something like a β-spiral structure (*63*−*65*). Most recently, Parchment et al. (*60*) constructed computer molecular models for the repeating sequence regions of HMW-GS, which involved spiral structures, including one in which the polypeptide chain assumed a loose spiral composed of β-turns and intervening β-strands. The structure was described as having little intramolecular hydrogen bonding. Such a model might have inherent flexibility, but our results do not support the likelihood of this being a major conformational structure for the repeating sequence domain of HMW-GS because we see little evidence for β-strand. Belton et al. (*65*) found that the elasticity of the HMW-GS is not in accord with hydrophobic interactions as might be expected from comparison with the β-spiral of elastin (*66*) but rather exhibited a hydrophilic character for the repeating sequence domain in which hydrogen bonding determines elasticity. Finally, McIntire et al. (McIntire, T. M., Lew, E. J.-L., Adalsteins, A. E., Anderson, O., Brant, D. A., and Kasarda, D. D., unpublished data) concluded from their noncontact atomic force microscopy of the hybrid monomer (*56*) that the results were mostly in accord with an extended form for

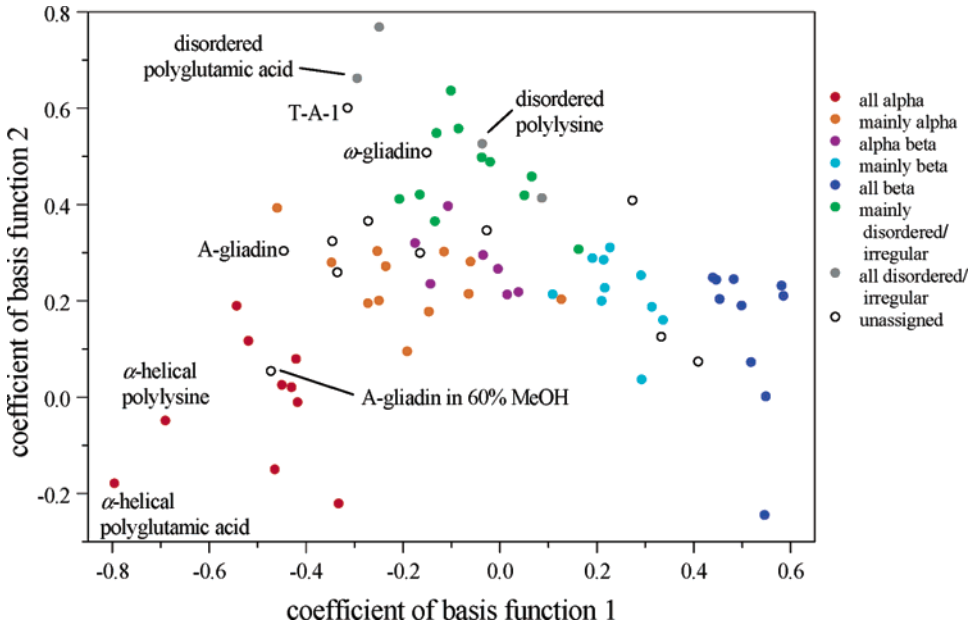

FIGURE 4: Plot of the PCA coefficients for the two most important basis for a set of 75 polypeptide, protein, and virus ROA spectra. The structural types are defined as follows: all α, > ∼60% α-helix with the rest mainly disordered and little or no β-sheet; mainly α, > ∼35% α-helix and a small amount (∼5−15%) of β-sheet; α β, similar significant amounts of α-helix and β-sheet; mainly β, > ∼35% β-sheet and a small amount (∼5−15%) of α-helix; all β, > ∼45% β-sheet with the rest mainly disordered and little or no α-helix; mainly disordered, little secondary structure; all disordered, no secondary structure.

the repeating sequence domain, either PPII or flexible random-coil, rather than any of the proposed spiral structures (*60*, *67*). Thus, more recent results, in combination with those presented here, seem to favor a flexible character for the repeating sequence domains of HMW-GS, in contrast to a uniform overall spiral structure.

Principal Component Analysis of the ROA Spectra. We are developing a pattern recognition program, based on principal component analysis (PCA), to identify protein folds from ROA spectral band patterns. Using a preliminary library of ROA spectral data, PCA calculates a set of subspectra that serve as basis functions that may be used to reconstruct any member of the original set of experimental ROA spectra. Figure 4 shows a plot of the coefficients for the whole set of 75 polypeptide, protein, and virus ROA spectra for the two most important basis functions. The protein and polypeptide positions are color-coded with respect to the seven different structural types listed on the figure, which provide a useful initial classification that will be refined in later work. The spectra separate into clusters corresponding to different types of protein structure, with increasing α-helix content to the left, increasing β-sheet content to the right, and increasing disorder from bottom to top. The disordered states of poly(L-glutamic acid) and poly(L-lysine) are found near the top of the ordinate axis and in the middle of the abscissa axis of the graph in accordance with their classification as disordered polypeptides. There is a considerable separation between the locations of these two homopolypeptides within this plot, suggesting that the disordered state of poly(L-glutamic acid) is more disordered than that of poly(L-lysine). Close to the coordinates representing disordered poly(L-glutamic acid) are those for the T-A-1 peptide that further support the earlier interpretation of the ROA spectrum that this peptide has little well-defined tertiary structure. The position of ω-gliadin lies between those of the T-A-1 peptide and disordered poly(L-lysine) indicating that it is largely disordered, consistent with the analysis of specific features of the ROA spectrum showing ω-gliadin to contain mostly PPII helix and a small amount of β-structure. The α-helical forms of poly(L-lysine) and poly(L-glutamic acid) are positioned at the bottom left-hand side of the plot, reflecting their high α-helix contents. A-gliadin is located at the edge of the mainly α region, which is consistent with the identification of a large amount of α-helix with some PPII helix and β-strand. When dissolved in 60% methanol, A-gliadin moves further to the left and downward into the all α region, which is in agreement with our earlier suggestion that methanol induces a conformational change in A-gliadin that results from some of the PPII helix and also possibly the β-structure being converted into solvated α-helix. This movement down the ordinate-axis also suggests that A-gliadin in 60% methanol is less disordered and more structured than in the absence of methanol.

Natively Unfolded Proteins. There is increasing interest in proteins that are wholly or partially unfolded under physiological conditions (*44*, *68*). Many proteins have been identified as being completely unfolded in their biologically active states, which has led to a reassessment of the structure−function paradigm (*69*). While proteins displaying a compact tertiary fold containing large amounts of extended secondary structure have been well-characterized in terms of widely recognized classes of fold, there is no similar clear consensus when describing different types of unfolded proteins. As these proteins are often difficult to characterize by experimental techniques, they have often been called unordered or random coil. The random coil was originally conceived as a collection of a large number of possible conformations of a polypeptide in which chain flexibility arises from internal rotation, with some degree of hindrance, around the covalent backbone bonds (*70*). The random coil is the limiting case of a dynamic type of disorder in which there is a distribution of Ramachandran ϕ,ψ angles for each

amino acid residue, giving rise to an ensemble of rapidly interconverting conformers. ROA scattering intensity is sensitive to absolute chirality, and the cancellation of contributions to the ROA spectrum from the enantiomeric structures characterizing many of these interconverting conformers results in a weak ROA spectrum with poorly resolved band structure for polypeptides displaying the dynamic type of disorder. The second limiting case is of a static type of disorder in which a sequence of residues has fixed but nonrepetitive ϕ,ψ angles (*26*, *29*). This is the situation of many of the residues, exclusive of those found in elements of secondary structure, within the longer loops of proteins that have well-defined tertiary folds. Polypeptide sequences containing residues with fixed but nonrepetitive ϕ,ψ angles appear to give rise to strong and well-defined ROA bands. Elements of secondary structure are defined by the characteristic and repetitive ϕ,ψ angles of their constituent residues and include α-helix, β-sheet, β-strand, and PPII-helix, although β- and γ-turns are also sometimes included in secondary structure despite their not having repetitive ϕ,ψ angles. ROA spectroscopy appears to be able to discriminate between the static type of disorder found in the loop and turn regions of compactly folded domains and the flexible coil often found in unfolded domains. The results presented here for the T-A-1 peptide indicate that many of the residues in the polypeptide chain have the ϕ,ψ angles characteristic of PPII but with short segment lengths for multiple residues having such torsion angles. In this regard, the PPII structure of the T-A-1 peptide, and indeed most polypeptides, may differ from polyproline itself, which, in the PPII form, has a significant segment length (*71*).

CONCLUSIONS

This ROA study has shown that A-gliadin contains a significant amount of a hydrated form of α-helix with smaller amounts of PPII helical- and β-structure. Some of each of the latter two structural types are converted into α-helix by methanol. On the basis of the amino acid sequence, it is likely that most of the α-helix lies within a more structured C-terminal domain. Interactions between the more structured C-terminal domain and the less structured N-terminal domain could help to stabilize the conformation of the latter in solution. In contrast, both ω-gliadin and the T-A-1 peptide of HMW-GS Dx5 appear to contain large proportions of residues with torsion angles characteristic of PPII helix, although without these residues enforcing a rigid rodlike structure on the peptide, possibly because of a short segment length for multiple residues in the PPII helical form. The ROA spectra also indicate the presence of some turns but indicate little or no α-helix, β-sheet, or β-strand in solution for ω-gliadin or the T-A-1 peptide and suggest that both are natively unfolded.

ACKNOWLEDGMENT

We thank Holly Jones for preparing the A-gliadin, Elva Adalsteins and Ellen Lew for contributing to the preparation of the T-A-1 peptide, Earl Cole for preparing the ω-gliadin fraction, and Robert W. Woody for valuable discussions.

REFERENCES

1. Shewry, P. R., and Tatham, A. S. (1990) *Biochem. J. 267*, 1−12.
2. Shewry, P. R., Tatham, A. S., Barro, F., Barcelo, P., and Lazzeri, P. (1995) *Bio/Technology 13*, 1185−1190.
3. Kasarda, D. D., Okita, T. W., Bernardin, J. E., Baecker, P. A., Nimmo, C. C., Lew, E. J.-L., Dietler, M. D., and Greene, F. C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4712−4716.
4. Anderson, O. D., and Greene, F. C. (1997) *Theor. Appl. Genet. 95*, 59−65.
5. Bernardin, J. E., Kasarda, D. D., and Mecham, D. K. (1967) *J. Biol. Chem. 242*, 445−450.
6. Kasarda, D. D., Bernardin, J. E., and Thomas, R. S. (1967) *Science 155*, 203−205.
7. Kasarda, D. D., Bernardin, J. E., and Gaffield, W. (1968) *Biochemistry 7*, 3950−3957.
8. Kasarda, D. D. (1980) *Ann. Technol. Agric. 29*, 151−173.
9. Cole, E. W., Kasarda, D. D., and LaFiandra, D. (1984) *Biochim. Biophys. Acta 787*, 244−251.
10. Purcell, J. M., Kasarda, D. D., and Wu, C.-S. C. (1988) *J. Cereal Sci. 7*, 21−32.
11. Anderson, O. D., Hsia, C., and Torres, V. (2001) *Theor. Appl. Genet. 103*, 323−330.
12. Kasarda, D. D., Autran, J.-C., Lew, E. J.-L., Nimmo, C. C., and Shewry, P. R. (1983) *Biochim. Biophys. Acta 747*, 138−150.
13. DuPont, F. M., Vensel, W. H., Chan, R., and Kasarda, D. D. (2000) *Cereal Chem. 77*, 607−614.
14. Shewry, P. R., Miflin, B. J., and Kasarda, D. D. (1984) *Philos. Trans. R. Soc. London Ser. B 304*, 297−308.
15. Lawrence, G. J., and Shepherd, K. W. (1980) *Aust. J. Biol. 33*, 221−233.
16. Rousset, M., Carrillo, J. M., Qualset, C. O., and Kasarda, D. D. (1992) *Theor. Appl. Genet. 83*, 403−412.
17. Payne, P. I., Nightingale, M. A., Krattiger, A. F., and Holt, L. M. (1987) *J. Sci. Food Agric. 40*, 51−65.
18. Kasarda, D. D. (1999) *Cereal Foods World 44*, 566−571.
19. Shewry, P. R., Field, J. M., Faulks, A. J., Parmar, S., Miflin, B. J., Dietler, M. D., Lew, E. J.-L., and Kasarda, D. D. (1984) *Biochim. Biophys. Acta 788*, 23−34.
20. Jeffrey, G. A., and Saenger, W. (1994) *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin.
21. Ewart, J. A. D. (1989) *Food Chem. 32*, 135−150.
22. Belton, P. (1999) *J. Cereal Sci. 29*, 103−107.
23. Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F., and Creamer, T. P. (2001) *Biochemistry 40*, 14376−14383.
24. Barron, L. D., and Hecht, L. (2000) in *Circular Dichroism, Principles and Applications* (Berova, N., Nakanishi, K., and Woody, R. W., Eds.) pp 667−701, John Wiley and Sons, New York.
25. Nafie, L. A. (1997) *Annu. Rev. Phys. Chem. 48*, 357−386.
26. Barron, L. D., Hecht, L., Blanch, E. W., and Bell, A. F. (2000) *Prog. Biophys. Mol. Biol. 73*, 1−49.
27. Smyth, E., Syme, C. D., Blanch, E. W., Hecht, L., Vasak, M., and Barron, L. D. (2001) *Biopolymers 58*, 138−151.
28. Syme, C. D., Blanch, E. W., Holt, C., Jakes, R., Goedert, M., Hecht, L., and Barron, L. D. (2002) *Eur. J. Biochem. 269*, 1−10.
29. Barron, L. D., Hecht, L., and Blanch, E. W. (2002) *Adv. Protein Chem. 62*, 51−90.
30. Anderson, O. D., and Greene, F. C. (1989) *Nucleic Acids Res. 17*, 461−462.
31. Hecht, L., Barron, L. D., Blanch, E. W., Bell, A. F., and Day, L. A. (1999) *J. Raman Spectrosc. 30*, 815−825.
32. Hanson, P., Anderson, D. J., Martinez, G., Millhauser, G., Formaggio, F., Crisma, M., Toniolo, C., and Vita, C. (1998) *Mol. Phys. 95*, 957−966.
33. Bolin, K. A., and Millhauser, G. L. (1999) *Accs. Chem. Res. 32*, 1027−1033.
34. Smith, C. K., and Regan, L. (1997) *Acc. Chem. Res. 30*, 153−161.
35. Kim, M. K., and Kang, Y. K. (1999) *Protein Sci. 8*, 1492−1499.
36. Pal, D., and Chakrabarti, P. (2000) *Acta Cryst. D56*, 589−594.
37. Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5355−5358.
38. Altschuler, E. L., Hud, N. V., Mazrimas, J. A., and Rupp, B. (1997) *J. Pept. Res. 50*, 73−75.
39. Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001) *J. Mol. Biol. 311*, 173−182.

40. Chen, S., Berthelier, V., Hamilton, J., O'Nuallain, B., and Wetzel, R. (2002) *Biochemistry 41*, 7391−7399.
41. Hirota-Nakaoka, N., and Goto, Y. (1999) *Bioorg. Med. Chem. 7*, 67−73.
42. Rost, B., and Sander, C. (1993) *J. Mol. Biol. 232*, 584−599.
43. Müller, S., and Wieser, H. (1995) *J. Cereal Sci. 22*, 21−27.
44. Uversky, V. N. (2002) *Eur. J. Biochem. 269*, 2−12.
45. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) *Proteins 41*, 415−427.
46. Hamada, D., and Goto, Y. (1997) *J. Mol. Biol. 269*, 479−487.
47. Hirota, N., Mizuno, K., and Goto, Y. (1998) *J. Mol. Biol. 275*, 365−378.
48. Tiffany, M. L., and Krimm, S. (1968) *Biopolymers 6*, 1379−1382.
49. Dukor, R. K., and Keiderling, T. A. (1991) *Biopolymers 31*, 1747−1761.
50. Keiderling, T. A., Silva, R. A. G. D., Yoder, G., and Dukor, R. K. (1999) *Biorg. Med. Chem. 7*, 133−141.
51. Woody, R. W. (1992) *Adv. Biophys. Chem. 2*, 37−79.
52. Shi, Z., Woody, R. W., and Kallenbach, N. R. (2002) *Adv. Prot. Chem. 62*, 163−240.
53. Blanch, E. W., Morozova-Roche, L. A., Cochran, D. A. E., Doig, A. J., Hecht, L., and Barron, L. D. (2000) *J. Mol. Biol. 301*, 553−563.
54. Blanch, E. W., Robinson, D. J., Hecht, L., and Barron, L. D. (2001) *J. Gen. Virol. 82*, 1499−1502.
55. Gill, A. C., Ritchie, M. A., Hunt, L. G., Steane, S. E., Davies, K. G., Bocking, S. P., Rhie, A. G. O., Bennett, A. D., and Hope, J. (2000) *EMBO J. 19*, 5324−5331.
56. Shimoni, Y., Blechl, A. E., Anderson, O. D., and Galili, G. (1997) *J. Biol. Chem. 272*, 15488−15495.
57. Tatham, A. S., Miflin, B. J., and Shewry, P. R. (1985) *FEBS Lett. 177*, 205−208.
58. Field, J. M., Tatham, A. S., and Shewry, P. R. (1987) *Biochem. J. 247*, 215−221.
59. Miles, M. J., Carr, H. J., McMaster, T. J., I'Anson, K. J., Belton, P. S., Morris, V. J., Field, J. M., Shewry, P. R., and Tatham, A. S. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 68−71.
60. Parchment, O., Shewry, P. R., Tatham, A. S., and Osguthorpe, D. J. (2001) *Cereal Chem. 78*, 658−662.
61. Tatham, A. S., Marsh, M. N., Wieser, H., and Shewry, P. R. (1990) *Biochem. J. 270*, 313−318.
62. Gilbert, S. M., Wellner, N., Belton, P. S., Greenfield, J. A., Siligardi, G., Shewry, P. R., and Tatham, A. S. (2000) *Biochim. Biophys. Acta 1479*, 135−146.
63. Tatham, A. S., Shewry, P. R., and Miflin, B. J. (1984) *FEBS Lett. 177*, 205−208.
64. Tatham, A. S., Drake, A. F., and Shewry, P. R. (1990) *J. Cereal Sci. 11*, 189−200.
65. Belton, P. S., Colquhoun, I. J., Field, J. M., Grant, A., Shewry, P. R., Tatham, A. S., and Wellner, N. (1995) *Int. J. Biol. Macromol. 17*, 74−80.
66. Tatham, A. S., Miflin, B. J., and Shewry, P. R. (1985) *Cereal Chem. 62*, 405−412.
67. Kasarda, D. D., King, G., and Kumosinski, T. F. (1994) in *Proceedings of the Symposium for Molecular Modeling* (Kumosinski, T. F., and Liebman, M., Eds.) pp 209−220, ACS, Washington, DC.
68. Tompa, P. (2002) *TIBS 27*, 527−533.
69. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
70. Poland, D., and Scheraga, H. A. (1970) *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press, New York.
71. Mandelkern, L. (1967) in *Poly-α-Amino Acids* (Fasman, G., Ed.) pp 675−722, Marcel Dekker, New York.

BI027059Y