Circular Dichroism and Nuclear Magnetic Resonance Spectroscopic Analysis of Immunogenic Gluten Peptides and their Analogs

Isabelle Parrot¹, Philip C. Huang¹ and Chaitan Khosla¹,²,³.

Departments of ¹Chemical Engineering, ²Chemistry and ³Biochemistry,
Stanford University, Stanford CA 94305-5025

Corresponding author:
Chaitan Khosla
E-mail: ck@chemeng.stanford.edu.
Phone/Fax : 650-723-6538

Running title: Structure of gluten peptides and analogs
Summary

Celiac Sprue, or gluten sensitive enteropathy, is an inheritable human disease of the small intestine that is triggered by the dietary intake of gluten. Recently, several Pro- and Gln-rich peptide sequences (most notably PQPQLPY and analogs) have been identified from gluten with potent immunogenic activity toward CD4\(^+\) T cells from small intestinal biopsies of Celiac Sprue patients. These peptides have three unusual properties. First, they are relatively stable toward further proteolysis by gastric, pancreatic and intestinal enzymes. Second, they are recognized and deamidated by human tissue transglutaminase (tTGase) with high selectivity. Third, tTGase catalyzed deamidation enhances their affinity for HLA-DQ2, the disease-specific class II MHC heterodimer. In an attempt to seek a mechanistic explanation for these properties, we undertook secondary structural studies on PQPQLPY and its analogs. Circular dichroism studies on a series of monomeric and dimeric analogs revealed a strong polyproline II helical propensity in a subset of them. Two-dimensional Nuclear Magnetic Resonance spectroscopic analysis confirmed a polyproline II conformation of PQPQLPY, and was also used to elucidate the secondary structure of the most helical variant, \((d-P)QPQLPY\). Remarkably, a strong correlation was observed between polyproline II content of naturally occurring gluten peptides and the specificity of human tTGase toward these substrates. Analogs with up to two \(d\) amino acid residues retained both polyproline II helical content and transglutaminase affinity. Since the Michaelis constant \((K_M)\) is the principal determinant of tTGase specificity for naturally occurring gluten peptides and their analogs, our results suggest that the tTGase binding site may have a preference for polyproline II helical substrates. If so, these insights could be exploited for the design of selective small-molecule inhibitors of this pharmacologically important enzyme.
Footnotes

1Gluten is a complex mixture of storage proteins found in a variety of food grains. These proteins can be fractionated into an alcohol insoluble fraction and an alcohol soluble fraction. The latter fraction from wheat, rye and barley is especially toxic to Celiac Sprue patients, and includes a family of closely related proteins called gliadins (in wheat), secalins (in rye) and hordeins (in barley). For more information on the chemistry of cereal proteins (1).
Introduction

Celiac Sprue (a.k.a. celiac disease, coeliac disease or gluten sensitive enteropathy) is a widely prevalent immune disorder of the small intestine that is triggered by the dietary intake of gluten from sources such as wheat, rye and barley (2). Like many autoimmune diseases, this inheritable disease is strongly linked to a particular allele of the class II major histocompatibility complex, HLA-DQ2 (3). Within the past few years several studies have succeeded in mapping epitopes in gluten that specifically stimulate CD4+ T cells from Celiac Sprue patients (but not control subjects) in an HLA-DQ2 dependent manner. Most of these epitopes are Pro- and Gln-rich peptides. A particular sequence, PQPQLPY, is found in gluten peptides that potently stimulate T cells from nearly all patients and are therefore considered “immunodominant” (4, 5). Gluten peptides containing the sequence PQPQLPY or analogs have three unusual properties. First, they are relatively stable toward further proteolysis by gastric, pancreatic and intestinal enzymes (6). Second, they are recognized and deamidated by human tissue transglutaminase (tTGase) with high selectivity (7, 8, 9). Third, tTGase catalyzed deamidation enhances their affinity for HLA-DQ2, the disease-specific class II MHC heterodimer (5). Therefore, understanding the structural features that underlie these properties could provide new insights into the pathogenicity of Celiac Sprue. We therefore undertook the task of characterizing PQPQLPY and selected analogs using a combination of Circular Dichroism and NMR spectroscopy.

Earlier secondary structure studies on peptides from wheat gluten have suggested the predominant existence of β-turn content (10, 11), although some polyproline II character has also been observed in longer (recombinant) polypeptides (12). Here we demonstrate that immunogenic gliadin\textsuperscript{1} peptides adopt polyproline II (PPII) helical conformations. Moreover, PPII content of these peptide correlates well with the specificity of human tTGase toward them. We
also show that incorporation of D-amino acid residues can increase their PPII character as well as the specificity of tTGase toward them.

**Experimental Procedures**

*Peptides.* The polyproline peptide used as a reference was purchased from Sigma (poly-L-proline, MW 1,000-10,000), and purified via preparative HPLC. Other peptides, synthesized at the Stanford Protein and Nucleic Acid Facility, were purified via preparative HPLC and verified to be >90% pure by analytical HPLC and mass spectrometry. The following gradient was used on a Econosphere C18 reverse phase column for all HPLC experiments: 0-25% of B (acetonitrile:H2O 80:20:0.1% TFA) over 5 min, 25-40% of B over 40 min, 40-100% of B over 1 min, 100% of B for 8 min., 100-0% of B over 1 min., and 0% of B for 7 min. In each instance, the remainder of the percentage was made up of a solution A (H2O:0.1% TFA).

Monomer 1, monomer 2 and monomer 3 correspond to D-modified analogs of the monomer PQPQLPY. Monomer 1 has one D-amino acid (D-proline1, pQPQLPY), monomer 2 has two D-amino acids (D-proline1 and D-glutamine2, pqPQLPY), monomer 3 has three D-amino acids (D-proline1, D-glutamine2, and D-proline3 ppqQLPY).

*Circular Dichroism.* Circular dichroism (CD) spectra were measured on an AVIV spectropolarimeter, model 60DS installed with a standard analysis program. The temperature was controlled using a recirculating water bath setting to 5 °C. Spectra were recorded with a quartz cell of 0.2 cm path length, using an acquisition time of 2 sec/nm, with a 1 nm spectral bandwidth, over the wavelength range from 190 nm to 260 nm. The samples were prepared by dissolving lyophilized peptide previously quantified with a 1 mM sodium citrate/1 mM sodium borate/1mM
sodium phosphate buffer, 15 mM NaCl with the pH adjusted at 7.0 (13). A number of standard buffers used for CD analysis of peptides were evaluated (13, 14, 17, 19); the citrate buffer was found to give the best resolution. Data are represented in molar ellipticities ([θ]_{max}, deg.cm^2.dmole^{-1}). The CD spectra represent an average of two accumulations at least, and were corrected by subtracting the buffer base line, and then usually fitted to a fourth-order polynomial curve, without additional filtering or smoothing.

To quantify the amount of PPII character in each peptide, it was necessary to define the CD spectra for both 100% and 0% PPII helical content. The former presumably corresponds to a rigid PPII structure whereas the latter refers to a completely disordered state. To estimate the maximum helical content (100%) of each peptide, we used the method of Creamer and coworkers (14), where [θ]_{max} at 227 nm measured in the presence of high guanidine concentrations was used as the upper bound for PPII helical content [θ]_{Gnd}. To define 0% PPII helical content, [θ]_{min} was measured at 227 nm in the presence of high concentrations of CaCl_2 ([θ]_{Ca}). From these parameters the PPII content could be estimated as follows:

\[
\%\text{PPII} = \frac{[\theta]_{\text{max}} + |[\theta]_{\text{Ca}}|}{[\theta]_{\text{Gnd}} + |[\theta]_{\text{Ca}}|} \times 100
\]

_NMR Spectroscopy._ Lyophilized peptides for NMR measurements were rehydrated in solution A or B. Solution A: DMSO-\textit{d}_6 containing TMS as an internal chemical shift reference. Solution B: a 20 mM filtered sodium phosphate buffer (pH 5.5) with 10% of D_2O. Acidic conditions were used to decrease the exchange rates of surface amide protons that are freely accessible to the solvent. Alternatively, NMR spectra were acquired in D_2O (Solution C). Final peptide concentrations of 10 mM were attained. DMSO-\textit{d}_6 and deuterium oxide (D_2O) was from...
Cambridge Isotope labs, MA. The sample is stable up to 2 months, as monitored by the 1-D $^1$H spectra.

All NMR spectra were acquired on a Varian Gemini 500 MHz spectrometer with triple resonance and three axis gradient capabilities. For most experiments a mixing time was applied, and the temperature was controlled at 25°C. Natural abundance two-dimensional experiments that were performed included: [$^1$H-$^{15}$N] Heteronuclear Single Quantum Coherence (HSQC), [$^1$H-$^{13}$C] HSQC, Total Correlation Spectroscopy (TOCSY), and Rotating frame Overhauser Enhancement Spectroscopy ROESY. The water signal was suppressed by pre-saturation. Data collection and processing were performed on Silicon Graphics workstations using VNMR programs. Coupling constants ($^3J_{HN\alpha}$) were extracted from the one-dimensional spectrum. The NOE cross-peak intensities (mixed time, 25°C) were determined using the SPARKY software.

**PPII Model Building.** The ideal PPII conformation model was built using the molecular modelling program SYBYL (Tripos Associates, St Louis, MO). The coupling constant ($^3J_{HN\alpha}$) is related to the dihedral angle $\phi$ as expressed by the relationship (15): $^3J_{HN\alpha} = 6.7\cos^2\phi - 1.3\cos\phi + 1.5$. The coupling constants were converted into $\phi$ values, which were introduced as a dihedral constraints, with $\psi$ values of 146.

**Expression, Purification and Kinetic analysis of tTGase.** Expression, purification and kinetic analysis of tTGase were performed as described previously (7). Glutamate dehydrogenase (GDH) was purchased from Biozyme (San Diego, CA), benzyloxycarbonyl-glutamylglycine (Z-Gln-Gly) from Sigma.
Results

CD Spectroscopic Analysis of PQPQLPY and Analogs. As a standard for polyproline II left-handed helical peptides, we measured the spectrum of a poly-L-proline at 5°C, in a 1 mM sodium citrate/1 mM sodium borate/1mM sodium phosphate buffer, 15 mM NaCl with the pH adjusted at 7.0 (13). As seen in Figure 1, this spectrum reproduced well the characteristic features of a stable polyproline II helix in polar environments: a strong negative band around 210 nm (π-π* transition) and a weak positive band around 229 nm (n-π* transition) (16-18).

Initial studies were performed on PQPQLYPQPQLPY, an immunodominant peptide from α-2 gliadin (5). This peptide is selectively deamidated by tTGase at Gln-4 and Gln-11 (7). The CD spectra of PQPQLYPQPQLPY and its mono-deamidated analog PQPELPYPQPQLPY displayed both characteristics of a polyproline II helix (Figure 1). The 229 nm maxima of these gliadin peptides were slightly shifted to lower wavelengths in comparison with the poly-L-proline reference. Presumably this results from a loss of tertiary amide chromophores due to fewer proline residues (18).

Two additional experiments provided evidence for PPII helical content in these gliadin peptides. First, unlike most secondary structural motifs, hydrogen bonding does not contribute significantly to the conformation of PPII peptides. Consistent with this, the Circular Dichroism spectra of both peptides show an increase in PPII content in the presence of 4-6 M chaotropic agents such as guanidine or urea (Figure 2A). Presumably, this increase results from interactions between the denaturing agent and the peptide backbone, which force the peptide to assume a more extended conformation in the sterically limited PPII region of (φ, ψ) space. Second, also consistent with earlier studies on PPII peptides, the molar ellipticities of both peptides showed a
decrease at $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ in the presence of 0.1-1.5 M CaCl$_2$ (14, 19) (Figure 2B), until the positive band disappeared altogether due to disruption of the extended helix (20, 21).

To test whether the PPII character of the above peptides could be ascribed to the PQPQLPY motif, CD spectroscopy was performed on this monomeric peptide, as well as four naturally occurring analogs (PQPQLSY, SQPQLPY, PQQASP, PQQQPP) found in gluten proteins. The first two analogs are from naturally occurring gliadin sequences, whereas the latter two are the closest homologs of PQPQLPY from the sequences of rice gluten proteins. (Rice is non-toxic in the Celiac diet.) As shown in Figure 3, both gliadin peptides had strong PPII helical character, whereas both rice peptides had substantially reduced helicity.

_Effect of Incorporating $\alpha$-Amino Acids on the PPII Content of PQPQLPY and Analogs._ Recently it was shown that substitution of $\alpha$-amino acids into a biologically active heptapeptide resulted in an increase of its PPII content (22). To assess the effect of substituting Pro$^1$, Gln$^2$ and Pro$^3$ in PQPQLPY for their $\alpha$-counterparts, the CD spectra of pQPQLPY, pqPQLPY and pqpQLPY were analyzed. (Lowercase letters correspond to D-amino acids.) As shown in Table 3, whereas monomers 1 and 2 were highly helical, helicity was substantially reduced in monomer 3. Indeed, monomer 1 (pQPQLPY) had the highest polyproline II helical content amongst all PQPQLPY analogs evaluated in this study, and was therefore the principal target for structural analysis via NMR spectroscopy (see below).

_NMR Assignments for pQPQLPY._ To verify that PQPQLPY and selected analogs did indeed adopt a polyproline II helical conformation, the structure of pQPQLPY (the analog with the highest helical content) was determined by two-dimensional NMR spectroscopy. As a prerequisite, the chemical shifts of individual hydrogen atoms needed to be assigned. A standard
method (23) was used to sequentially assign all the proton resonances of this peptide at 25°C. TOCSY spectra were used to identify spin systems, and ROESY spectra were used to obtain inter-residue connectivities and to distinguish equivalent spin systems. Unambiguous assignments were made in the following three steps.

First, the $^{15}$N-$^1$H HSQC of pQPQLPY in solution A (DMSO) or B (phosphate buffer) was ascertained to contain the correct number of amide proton peaks before further peak assignment. Since proline residues do not contain amide protons and the N-terminal amino group proton is not observed in aqueous buffers, only four residues in the seven-residue peptide were expected to show cross peaks corresponding to the amide protons. In addition, two peaks for amide protons in each of the glutamine side chains were expected. In aggregate, ten strong cross peaks were observed in Solution A, whereas eight peaks were observed in Solution B. (An extra peak observed in each case was weaker than the others, and was therefore assumed to correspond to the protonated state of the N-terminal residue.) Second, we examined the $\alpha$-δH/NH fingerprint regions of the TOCSY spectra. The spin systems of the individual amino acids residues were readily identified in the TOCSY spectrum. For example, the two glutamine spin systems showed characteristic $\beta$H and γH, whereas the leucine spin system had a characteristic unique delta methyl proton chemical shift at <1 ppm, and the tyrosine residue had characteristic signals in the aromatic region of the spectrum. Further examination of the full TOCSY and $^1$H-$^{13}$C HSQC spectrum also revealed three sets of peaks in the carbon (F1) dimension, corresponding to the proline Cδ-Hδ correlation Pro$^1$, Pro$^3$, Pro$^6$. Third, using these assigned residues together with the sequential NOE connectivities from ROESY, sequential specific resonance assignments could be completed.
The complete $^1$H NMR assignments for pQPQLPY in Solution B and D$_2$O are summarized in Table 1, whereas $\alpha$H and NH chemical shifts in DMSO-$d_6$ are summarized in supporting information. The NH shift dispersion was approximately equivalent (ca. 0.3 ppm) in either solvent, implying that solvent polarity and dielectric had little influence on the conformation of the peptide.

Secondary Structure Elucidation of pQPQLPY. Since the one-dimensional and two-dimensional spectra of pQPQLPY showed the same characteristics in phosphate buffer, D$_2$O and DMSO-$d_6$, the solution conformation of this peptide was explored in either solvent by determination of C$\alpha$ proton chemical shift values, three-bond spin-spin ($^3J$) coupling constants, medium/long range NOE coupling and inter-proton distances between the $\alpha$H of the preceding residues and the proline $\alpha$H and $\delta$H.

The deviation of $\alpha$H chemical shifts from random coil values provides some information about peptide secondary structure, since well-defined chemical shift changes are associated with $\alpha$-helices and extended $\beta$-strands (24, 25). The Chemical Shift Index (CSI) developed by Wishart and Sykes compares each $\alpha$H chemical shift in a peptide/protein to a range of random coil values, as determined from a large database of peptides and proteins (26). A chemical shift value lower than the range indicates that the residue lies in a region of $\alpha$-helix, while a value greater than the range suggests participation in a $\beta$-strand extended structure. A grouping of four or more upfield shifts (average negative shift -0.38 ppm) is designated as an $\alpha$-helix, while three or more consecutive downfield shifts (average positive shift 0.38 ppm) imply $\beta$-strand (24, 25). Changes in chemical shifts of $\alpha$H resonances from random coil for pQPQLPY in aqueous buffer are summarized in Table 1. While several $\alpha$H chemical shifts fell outside the random coil range, our data did not suggest local regions of classical $\alpha$-helix or $\beta$-strand secondary structure in aqueous...
solution, with CSI$_{\alpha H}$ values < 1.2 ppm. Therefore, it was likely that pQPQLPY has an alternative, possibly more rigid, conformation than other peptides.

The torsion angles $\phi$ and $\psi$ are key parameters for defining the backbone conformation of a polypeptide chain (27). In solution these angles can be detected experimentally by NMR via spin-spin coupling constants, which in case of $^3J$ coupling constants can be related to specific torsion angles by Karplus relationships (15, 28). Three bond $^3J_{HN\alpha}$ coupling constants for extended chains are 6-9 Hz, whereas smaller coupling constants (2-5 Hz) are observed for $\alpha$-helix and for the second and/or the third residues of $\beta$-strand secondary structure (29). All $^3J_{HN\alpha}$ coupling constants measured for pQPQLPY in aqueous buffer are summarized in Table 1. Those coupling constants were found to be in the 6.5 to 7.5 Hz range, in accordance with an extended structure.

The intensity of the ratio between sequential $\alpha H$-$NH$ (i, i+1) and $NH$-$NH$ (i, i+1) NOEs is useful as a measure for secondary structure, since it depends on the dihedral angles ($\phi$ and $\psi$). Using population-weighted random coil, Dobson and co-workers predicted that the intensity ratio of $\alpha H$-$NH$ (i, i+1)/$NH$-$NH$ (i, i+1) NOEs was about 1.4 for random structure, notably distinct from the ratio for extended $\beta$-strand conformation (about 55) (30). An analysis of NOE intensities for pQPQLPY revealed that in aqueous solution or in DMSO-$d_6$, $NH$-$NH$ (i, i+1) were not detected for any residues of the peptide, whereas $\alpha H$-$NH$ (i, i+1) were medium/strong (figure 4 and 5). This suggested a high $\alpha H$-$NH$ (i, i+1)/$NH$-$NH$ (i, i+1) intensity ratio, also consistent with a predominantly extended conformation rather than random coil. The NOE intensity ratio $\alpha H$-$NH$ (i, i+1)/$\alpha H$-$NH$ (i, i) provided further evidence for extended structure in solution. A value about 2.3 is predicted for the population-weight random coil model, while value $>$ 4 are predicted for a $\beta$-strand (31). Our analysis of NOE intensities in aqueous buffer revealed that just the $\alpha H$-$NH$ (i, i) were not detected with an exception for the residue Gln$^4$, which had an intensity ratio $\alpha H$-$NH$ (i, i)
(i, i+1)/αH-NH (i, i) of 3. However in DMSO-\textit{d}_6, all residues had an intensity ratio αH-NH (i, i+1)/αH-NH (i, i) larger than 3.5, consistent with a large population of extended structure through the entire sequence (Table 1, figure 4 and supporting information).

A key requirement for a PPII helix is that the peptide bonds preceding Pro residues adopt \textit{trans} configurations. To distinguish between the \textit{cis}/\textit{trans} configuration of the imide bonds in pQPQLPY, the difference in inter-proton distances between the αH of the preceding residues and the proline αH and δH were measured (32). A short distance (strong NOE signal) between αH of a preceding residue and δH of a proline is diagnostic for the \textit{trans} configuration, whereas a short distance (strong NOE signal) between αH of a preceding residue and αH of a proline is diagnostic for the \textit{cis} configuration. In DMSO-\textit{d}_6 (supporting information) or in D\textsubscript{2}O (figure 5), the αδ(i, i+1) NOE cross peaks in the ROESY spectrum strongly indicated that the proline residues of pQPQLPY adopted a \textit{trans} configuration at 25°C, confirming PPII conformation.

\textit{NMR Spectroscopy on PQPQLPY and pqpQLPY in phosphate buffer.} To confirm the generality of the CD spectroscopic results on PQPQLPY and its analogs, NMR spectroscopy was also performed on PQPQLPY itself (predicted to have a relatively high PPII content) and its analog pqpQLPY (predicted to have a lower PPII content). The spin system of individual amino acids residues in both peptides were easily identified in the TOCSY and ROESY spectra, and compared to assignments that were made for pQPQLPY. In buffer B, two-dimensional and one-dimensional spectra of PQPQLPY revealed the presence of isomers for Gln\textsuperscript{4} and Leu\textsuperscript{5}, whereas isomers of Gln\textsuperscript{2}, Gln\textsuperscript{4} and Leu\textsuperscript{5} were observed for pqpQLPY (supporting information). This suggested the existence of minor conformations in both cases in dynamic equilibrium with a major conformation. The percentage of these isomers, estimated by peak integration, were
approximately 22 % for PQPQLPY and 38 % for the pqpQLPY, suggesting that the former peptide is more structured than the latter. Moreover, as shown in Table 2, the $\alpha$H chemical shifts were about the same (ca. 0.04 ppm dispersion) for the major isomers of all three peptides. The same was also the case for NH chemical shifts (dispersion ca.0.3 ppm), suggesting that the major conformation was strongly preserved in the different peptides. Finally, the $^{3}J_{HN}$ coupling constants range between 6.5 to 8 Hz for all peptides, indicating an extended conformation for the backbone, rather than a $\beta$-turn or $\alpha$-helix. Together these results suggest that the predominant conformations of both PQPQLPY and (to a lesser extent) pqpQLPY are PPII helices.

**Correlation between PPII Content and tTGase Specificity.** Table 3 summarizes the estimated PPII helical content of selected gliadin peptides showing different affinities for human tTGase. Remarkably, a strong correlation is observed between PPII helical content and tTGase specificity. In particular, PQPQQPP, which has more proline residues than PQPQLPY, has negligible PPII helical content, and is a very poor substrate of human tTGase. Similarly, PQPQAPS, which has substantially reduced PPII content, is a much weaker substrate of tTGase than PQPQLPY. From the data summarized in Table 3 it also appears that variability in tTGase substrate specificity can primarily be explained by differences in the Michaelis constant ($K_{M}$). Since this parameter presumably reflects differences in ground-state binding abilities of the peptides for the tTGase active site, it is likely that tTGase has an intrinsic preference for PPII helical peptide substrates. However, our data also indicates that, although PPII conformation increases enzyme specificity for the corresponding peptide, it alone cannot explain the observed specificity. For example, the highly helical peptide PQPQP PPP is not a substrate. Thus it appears that human tTGase prefers PPII helical peptides with an aromatic residue (Y or F) at the C-terminus.
Discussion

An investigation into the conformation of immunogenic peptides from gliadin is important for the understanding of the pathogenicity of these peptides in Celiac Sprue, a hereditary disease of the small intestine. Several decades ago, Frazer and coworkers showed that pretreatment of gliadins with pepsin, trypsin and/or chymotrypsin does not result in loss of gliadin toxicity (33), demonstrating that the intact tertiary structure of gliadin was not relevant for pathogenesis. However, since all known toxic peptides from gluten are at least 8 residues in length, the secondary structures of the proteolytic products of gluten could be extremely important to their toxicity. Using circular dichroism (CD) and NMR spectroscopy, we have shown here that immunogenic gluten peptides have a strong PPII character. More remarkably, the observed specificity of human tTGase toward these peptides correlates well with their helical content.

Proline-rich regions are believed to adopt left-handed polyproline II (PPII) conformations, similar to poly-L-proline and collagen (34). This helical conformation, the hallmark of SH3 domains (35), plays an important role in biological signal transduction. Notably, the preferred peptide ligands of class II MHC molecules are also bound in the PPII conformation (36, 37). PPII helices can also be found in sequences not rich in proline, such as poly(lysine), poly(glutamate), and poly(aspartate) (38). In particular, glutamine, an amino acid that makes up ~30% of gliadin proteins, is an excellent PPII helix former (14, 34). Finally, PPII helices are short, the majority being just four residues in length (34). Indeed, only three proline residues are required to form and detect a PPII helix. For example the PQP and the PPP motifs have 65 % and 66% PPII content, respectively.

The secondary structure of a PPII helix shows a relatively extended conformation in which all peptide bonds are trans and the helix is left-handed with 3 residues per turn (17, 18). Since the
circular dichroism (CD) spectra of peptides are very sensitive to secondary structure, CD has proven to be a powerful and discriminating tool in the identification of PPII helices. In the first part of our study, results obtained by CD spectroscopy suggested that PQPQLPYPQPQLPY (the immunodominant peptide from α-2gliadin), its monodeaminated analog PQPELPYPQPQLPY, as well as PQPQLPY and a number of its analogs adopt a common PPII conformation (figures 1-4). In each case a positive shoulder around 229 nm is clearly indicative of the presence of a stable PPII helix in polar environments. This is further supported by calcium antagonism and guanidine (or urea) dependent enhancement of helical content (figure 2A and 2B). Our CD spectroscopic investigations also highlighted the potential correlation between the immunogenicity of naturally occurring gluten peptides and their PPII helical content. Whereas PQPQLPY and two other analogs (PQPQLSY, PSQPQLPY) from gluten sources that are toxic in the Celiac diet adopted PPII conformations (figure 3, table 3), the closest analogs from rice gluten (a non-toxic food grain) were not substantially helical in nature. Finally, CD spectroscopy also revealed that substitution of Pro¹ and/or Gln² with their D-counterparts in PQPQLPY led to an increase in PPII content.

To confirm our findings from CD spectroscopy, we conducted NMR studies on selected peptides. NMR spectra of pQPQLPY revealed narrow deviations of αH chemical shifts from random coil values, narrow αH chemical shifts dispersions, and coupling constants ranging between 6.5-7.5 Hz. Together, these observations allowed us to rule out α-helical or β-stranded secondary structure. Moreover, since sequential αH-NH and NH-NH NOEs could not be observed in the spectra, an extended structure with inter-proton distances larger than 5.5 Å (e.g. PPII), rather than a random coil structure, was preferred. Finally, the observation of medium range NOEs, such as αN(i, i+2), together with the existence of a large NOE intensity ratio αN(i,
i+1)/ αN(i, i), indicated that pQPQLPY was not a random coil. Conclusive evidence for a PPII helix came from strong NOE signals between the δH of proline residues and αH of the preceding residues, a feature that is diagnostic of a trans-Pro configuration. Similar NMR studies on PQPQLPY and pqpQLPY revealed that the former peptide was more helical than the latter (but less helical than pQPQLPY), a finding that was entirely consistent with our results from CD spectroscopy.

A model for the secondary structure of pQPQLPY, derived from NMR-derived dihedral angle constraints and subsequent energy minimization, is shown in Figure 6. Upon adopting a PPII conformation with 3.0 amino acids per turn, the two glutamine side chains are aligned on the same side of the helix, whereas the hydrophobic residue leucine and the amphipathic residue tyrosine is on the opposite side. This backbone and side-chain conformation is presumably important for the pathogenic properties of the PQPQLPY motif to Celiac Sprue patients in two ways. First, the strong correlation between PPII content and tTGase specificity (Table 3) suggests that the binding pocket of human tTGase has a preference for peptide substrates that intrinsically adopt this conformation. Although the crystal structure of human tTGase has recently been solved (39), the substrate-binding pocket of this enzyme is insufficiently defined. Future co-crystallographic studies with PPII peptides could not only help in understanding the structural basis for this correlation, but would also be useful in the design of selective inhibitors for this therapeutically important enzyme. (In addition to Celiac Sprue, tTGase is also believed to play an important role in Huntington’s disease and certain skin diseases (40).) Second, the intrinsic conformation of PQPQLPY is also well-matched to the presumed active site of Celiac Sprue associated HLA-DQ2, which has been modeled based on the recent X-ray structure of the closely related HLA-DQ8 (41, 42). Again, this insight may be translatable into pharmacologically useful
HLA-DQ2 antagonists. In summary, our structural studies on the conformation of immunogenic gluten peptides in solution provide fundamentally new insights into the structural biology of gluten, and represent a first step in understanding its remarkable ability for interacting with two key human proteins, tTGase and HLA-DQ2.
Acknowledgements: We thank Dr. S. Lynch for advice in connection with the acquisition and analysis of NMR data. This research was supported by the Alan T. Waterman Award from the National Science Foundation to C.K. I.P. is a recipient of a postdoctoral fellowship from the Fondation pour la Recherche Médicale (France).
References:

Table 1. $^1$H NMR sequence-specific resonance assignments (ppm), $^3J_{HN_a}$ (Hz), and CSI$_{\alpha H}$ (ppm) for pQPQLPY in solution C, D$_2$O, and in solution B, a 20 mM sodium phosphate buffer (pH 5.5) with 10 % of D$_2$O.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$\alpha H$</th>
<th>$\beta H$</th>
<th>$\delta H$</th>
<th>$\gamma H$</th>
<th>$\delta NH^a$</th>
<th>$^3J_{HN_a}$</th>
<th>$\phi$</th>
<th>others</th>
<th>CSI$_{\alpha H}$ $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro$^1$</td>
<td>4.25</td>
<td>2.34</td>
<td>3.25, 3.28</td>
<td>1.85</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.19</td>
</tr>
<tr>
<td>Gln$^2$</td>
<td>4.52</td>
<td>1.81</td>
<td>-</td>
<td>2.32</td>
<td>8.70</td>
<td>6.5 Hz</td>
<td>-77.9°</td>
<td>$\delta NH_2^a$ 7.43, 7.47</td>
<td>0.15</td>
</tr>
<tr>
<td>Pro$^3$</td>
<td>4.29</td>
<td>1.74, 2.13</td>
<td>3.54, 3.67</td>
<td>1.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.15</td>
</tr>
<tr>
<td>Gln$^4$</td>
<td>4.18</td>
<td>1.83, 1.92</td>
<td>-</td>
<td>2.26</td>
<td>8.37</td>
<td>7.5 Hz</td>
<td>-78°</td>
<td>$\delta NH_2^a$ 6.75, 6.78</td>
<td>-0.19</td>
</tr>
<tr>
<td>Leu$^5$</td>
<td>4.45</td>
<td>1.28, 1.44</td>
<td>0.78</td>
<td>1.50</td>
<td>8.19</td>
<td>7.5 Hz</td>
<td>-78°</td>
<td>-</td>
<td>0.07</td>
</tr>
<tr>
<td>Pro$^6$</td>
<td>4.27</td>
<td>1.75, 2.06</td>
<td>3.45, 3.63</td>
<td>1.86</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.17</td>
</tr>
<tr>
<td>Tyr$^7$</td>
<td>4.43</td>
<td>2.92, 2.96</td>
<td>-</td>
<td>-</td>
<td>7.67</td>
<td>7.5 Hz</td>
<td>-78°</td>
<td>$2.6H$ 7.01 3.5H 6.72</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

$^a$ in solution B. $^b$ CSI$_{\alpha H}$ = ppm$_{obs}$-ppm$_{rc}$. 
**Table 2.** Chemical shifts (ppm) and $^3J_{HN}$ coupling constants for three PQPQLPY analogs.

<table>
<thead>
<tr>
<th>Residue</th>
<th>pQPQLPY</th>
<th>PQPQLPY</th>
<th>pqpQLPY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$H</td>
<td>NH$^a$</td>
<td>$^3J_{HN}$</td>
</tr>
<tr>
<td>Pro$^1$</td>
<td>4.25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gln$^2$</td>
<td>4.52</td>
<td>8.70</td>
<td>6.5 Hz</td>
</tr>
<tr>
<td>Pro$^3$</td>
<td>4.29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gln$^4$</td>
<td>4.18</td>
<td>8.37</td>
<td>7.5 Hz</td>
</tr>
<tr>
<td>Leu$^5$</td>
<td>4.45</td>
<td>8.19</td>
<td>7.5 Hz</td>
</tr>
<tr>
<td>Pro$^6$</td>
<td>4.27</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyr$^7$</td>
<td>4.43</td>
<td>7.67</td>
<td>7.5 Hz</td>
</tr>
</tbody>
</table>
**Table 3.** Summary of PPI content and kinetic data for PQPQLPY, PQPQLPYPQPQLPY, gliadin peptides and derivatives. Kinetic parameters for peptides marked with an asterisk are from ref. (7).

<table>
<thead>
<tr>
<th>tTGase substrates</th>
<th>% PPII</th>
<th>K_M (mM)</th>
<th>k_cat/K_M (min⁻¹.mM⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PQPQLSY*</td>
<td>93.2</td>
<td>0.3</td>
<td>87</td>
</tr>
<tr>
<td>PQPQLPY*</td>
<td>87.8</td>
<td>0.3</td>
<td>82</td>
</tr>
<tr>
<td>PQPQLPF*</td>
<td>84.4</td>
<td>0.4</td>
<td>72</td>
</tr>
<tr>
<td>SQPQLPY*</td>
<td>83.2</td>
<td>0.3</td>
<td>96</td>
</tr>
<tr>
<td>PQPQAPS*</td>
<td>20.7</td>
<td>&gt;10</td>
<td>7</td>
</tr>
<tr>
<td>PQPQQPP*</td>
<td>&lt;1</td>
<td>&gt;10</td>
<td>1</td>
</tr>
<tr>
<td>pqPQLPY</td>
<td>96</td>
<td>0.2</td>
<td>95</td>
</tr>
<tr>
<td>pqpQLPY</td>
<td>90.4</td>
<td>0.1</td>
<td>98</td>
</tr>
<tr>
<td>pqpQLPY</td>
<td>48.6</td>
<td>0.7</td>
<td>41</td>
</tr>
<tr>
<td>PQPQPPP</td>
<td>89 %</td>
<td>&gt;10</td>
<td>&lt;10</td>
</tr>
</tbody>
</table>
CAPTIONS TO FIGURES

**Figure 1.** CD spectra of poly-l-proline, dimer PQPQLPYPQPQLPY and its deaminated derivative PQPELPYPQPQLPY, at 5°C, in a 1 mM sodium citrate/1 mM sodium borate/1mM sodium phosphate buffer, 15 mM NaCl with the pH adjusted at 7.0.

**Figure 2.** CD spectra of dimer PQPQLPYPQPQLPY at 5°C, in a 1 mM sodium citrate/1 mM sodium borate/1mM sodium phosphate buffer, 15 mM NaCl with the pH adjusted at 7.0. (A). in the presence of 0-1.5 M of CaCl$_2$. (B). in the presence of 0-6 M of guanidine.

**Figure 3.** CD spectra of selected gliadin peptides, at 5°C, in a 1 mM sodium citrate/1 mM sodium borate/1mM sodium phosphate buffer, 15 mM NaCl with the pH adjusted at 7.0.

**Figure 4.** Expanded ROESY region for pQPQLPY in solution B (a 20 mM filtered sodium phosphate buffer (pH 5.5) with 10 % of D$_2$O), showing αH-NH trough space connectivity, of which resolved cross peaks are labelled.

**Figure 5.** (A). Expanded TOCSY region for pQPQLPY, in solution C (D$_2$O). (B). Expanded ROESY region for pQPQLPY, in solution C (D$_2$O).

**Figure 6.** Energy minimized model of the PPII helical conformation of pQPQLPY. The initial model was derived from dihedral angles calculated from NMR spectroscopic data.
Figure 1.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6.
Supporting information

**Table S1.** Key $^1$H NMR sequence-specific resonance assignments (ppm) for pQPQLPY in DMSO-$d_6$.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$\alpha$H</th>
<th>NH</th>
<th>others</th>
<th>$\alpha$H-NH (i, i+1)/ $\alpha$H-NH (i, i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro$^1$</td>
<td>4.15</td>
<td>8.53, 9.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln$^2$</td>
<td>4.55</td>
<td>8.85</td>
<td>$\delta$NH$_2$ 7.20, 7.30</td>
<td>3.8</td>
</tr>
<tr>
<td>Pro$^3$</td>
<td>4.34</td>
<td>-</td>
<td>$\delta$NH$_2$ 6.75, 6.82</td>
<td>6.3   5.0</td>
</tr>
<tr>
<td>Gln$^4$</td>
<td>4.21</td>
<td>8.06</td>
<td>$\delta$NH$_2$ 6.75, 6.82</td>
<td></td>
</tr>
<tr>
<td>Leu$^5$</td>
<td>4.51</td>
<td>7.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro$^6$</td>
<td>4.38</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr$^7$</td>
<td>4.28</td>
<td>7.90</td>
<td>2.6H 7.02, 3.5H 6.66</td>
<td>3.6</td>
</tr>
</tbody>
</table>
Table S2. Distinguishing features for random coil, and extended structures

<table>
<thead>
<tr>
<th></th>
<th>random coil</th>
<th>β structure</th>
<th>PII helix</th>
<th>pQPQLPY data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CD</strong></td>
<td>( \lambda_{\text{max}} = 215 \text{ nm} )</td>
<td>( \lambda_{\text{max}} = 196 \text{ nm} )</td>
<td>( \lambda_{\text{max}} = 215 \text{ nm} )</td>
<td>( \lambda_{\text{max}} = 210 \text{ nm} )</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{\text{min}} = 197 \text{ nm} )</td>
<td>( \lambda_{\text{min}} = 218 \text{ nm} )</td>
<td>( \lambda_{\text{min}} = 230 \text{ nm} )</td>
<td>( \lambda_{\text{min}} = 226 \text{ nm} )</td>
</tr>
<tr>
<td><strong>CD in presence of urea or guanidine</strong></td>
<td>( \uparrow \lambda_{\text{max}} )</td>
<td>( \uparrow \lambda_{\text{max}} )</td>
<td>( \uparrow \lambda_{\text{max}} )</td>
<td>( \uparrow \lambda_{\text{max}} )</td>
</tr>
<tr>
<td><strong>CD in presence of CaCl(_2)</strong></td>
<td>( \downarrow \lambda_{\text{max}}, \downarrow \lambda_{\text{min}} )</td>
<td>( \downarrow \lambda_{\text{max}}, \downarrow \lambda_{\text{min}} )</td>
<td>( \downarrow \lambda_{\text{max}}, \downarrow \lambda_{\text{min}} )</td>
<td>( \downarrow \lambda_{\text{max}}, \downarrow \lambda_{\text{min}} )</td>
</tr>
<tr>
<td>( \alpha )H chemical shift dispersion</td>
<td>narrow, within 1 ppm</td>
<td>wide</td>
<td>narrow</td>
<td>narrow</td>
</tr>
<tr>
<td>( \alpha )H chemical shift difference</td>
<td>&lt; 0.1 ppm</td>
<td>&gt; 0.1 ppm</td>
<td>&lt; 0.2 ppm</td>
<td>&lt; 0.2 ppm</td>
</tr>
<tr>
<td>( \beta )J(_{\text{HN}}) coupling constant</td>
<td>6.5-7.0 Hz, 5.8 Hz for Gly</td>
<td>&gt; 8.5 Hz</td>
<td>= 7.5 Hz (( \phi = -78^\circ ))</td>
<td>6.5-7.5 Hz</td>
</tr>
<tr>
<td>NOE NN(i, i+1)</td>
<td>medium</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>NOE ( \alpha )N(i, i+1)/ NN(i, i)</td>
<td>1-2.7 avg 1.4</td>
<td>&gt; 55</td>
<td>very large</td>
<td>very large</td>
</tr>
<tr>
<td>NOE ( \alpha )N(i, i+1)/ ( \alpha )N(i, i)</td>
<td>2.3</td>
<td>&gt; 4</td>
<td>&gt; 2.3</td>
<td>&gt; 3.5</td>
</tr>
<tr>
<td>Medium range NOE ( \alpha )N (i, i+2)</td>
<td>observed</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>( \alpha \delta )(i, i+1) NOE</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
**Figure S1.** Expanded ROESY region for pQPQLPY in solution A (DMSO-d$_6$), showing $\alpha$H-NH trough space connectivity, of which resolved cross peaks are labelled.

**Figure S2.** (A). Expanded TOCSY region for pQPQLPY, in solution A (DMSO-d$_6$). (B). Expanded ROESY region for pQPQLPY, in solution A (DMSO-d$_6$).

**Figure S3.** Expanded TOCSY region for pqpQLPY, in solution B (a 20 mM filtered sodium phosphate buffer (pH 5.5) with 10 % of D$_2$O).

**Figure S4.** Expanded TOCSY region for PQPQLPY, in solution B (a 20 mM filtered sodium phosphate buffer (pH 5.5) with 10 % of D$_2$O).
Figure S1
Figure S2.
Figure S3.
Figure S4.