Comparison of Repetitive Sequences Derived from High Molecular Weight Subunits of Wheat Glutenin, an Elastomeric Plant Protein

Nikolaus Wellner,*† Justin T. Marsh,‡ Andrew W. J. Savage,‡ Nigel G. Halford,‡ Peter R. Shewry,‡ E. N. Clare Mills,† and Peter S. Belton§

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, Rothamsted Research, Rothamsted, Harpenden, Herts AL5 2JQ, and School of Chemical Sciences and Pharmacy, University of East Anglia, Norwich, NR4 7TJ, U.K.

Received November 23, 2005; Revised Manuscript Received February 9, 2006

A strategy has been developed to create repetitive peptides incorporating substitutions in the PGQGQQGYYPTSLQQ consensus repeat sequence of high molecular weight subunits in order to investigate natural sequence variations in elastomeric proteins of wheat gluten. After introduction of glutamic and aspartic acid residues, the peptide behaved similarly to the unmodified form at low pH, but became readily water soluble at pH > 6. Substitution of Gln for Leu at position 13 resulted in only small changes to the secondary structure of the water-insoluble peptides, as did Tyr8His and Thr11Ala. The effects of proline substitutions depended on their location: Leu13Pro substitution had little effect on solubility and structure, but Gln6Pro substitution resulted in dramatic changes. Peptides with two Gln6Pro substitutions had similar properties to the water-insoluble parental peptide, but those with 6 or 10 substitutions were readily soluble. The results indicated that specific sequences influence noncovalent intermolecular interactions in wheat gluten proteins.

The grain storage proteins of wheat are unique among plant proteins in exhibiting elastomeric properties, forming the fraction known as gluten, which underpins the wide use of wheat in food. Although these properties are important in determining processing properties, they appear to have no biological significance, which contrasts with animal systems in which elastomeric proteins occur more widely and play crucial roles in biological and physiological processes.[1] Elastomeric proteins such as elastin and fibrillin are present in mammalian muscles and connective tissues. Other examples are viscid spider silks, resilin in arthropods, and byssus in holdfasts of marine mussels. All of these proteins resemble gluten in having complex multidomain structures with at least one domain consisting of repeated amino acid sequences. Furthermore, although they differ in their precise elastomeric properties and mechanisms, reflecting their different biological roles, in all cases there appears to be an entropic component.[2] Thus, entropy appears to be the major restoring force in elastin.[3] In contrast, the situation with wheat gluten proteins appears to be more complex.[4] Although it is accepted that gluten elasticity depends on the presence of disulfide-stabilized polymers, the precise molecular basis for elasticity and the role of noncovalent interactions, notably hydrogen bonds, remains unclear.

Recent studies of gluten elasticity have focused on one group of gluten proteins called the high molecular weight (HMW) subunits of gluten. These proteins only account for about 6−12% of the total gluten fraction but are present alongside other gluten proteins in high molecular mass (approximately 1×10^6 to 10×10^6) polymers which are crucial for elasticity.[1] A number of full sequences of HMW subunits are available, showing that they comprise between about 600 and 900 amino acids with 3 structural domains. The N- and C-terminal parts of the proteins (comprising 81−104 and 42 residues, respectively) appear to be rich in α-helix and contain cysteine residues which form interchain and intrachain disulfide bonds, the former stabilizing the glutenin polymers. In contrast, the central parts of the proteins are repetitive, comprising tandem and interspersed repeats of hexapeptide (consensus PGQGQQ), nonapeptide (GYYPTS(P/L)QQ), and, in some subunits only, tripeptide (GQQ) motifs[1]. Studies of intact subunits and peptides have indicated that these sequences form β-turns and poly-L-proline II structures in solution, whereas in the solid hydrated state, a mixture of structures is observed, including intermolecular β-sheet.[5−7] The latter is considered to form during protein aggregation and appears to be stabilized by interchain hydrogen bonding (facilitated by the high contents of glutamine in the motifs) with the proportion depending on the hydration level.[6,8] The equilibrium between the presence of hydrogen-bonded β-sheet structures and flexible β-turn-rich structures has been proposed to contribute to the elastomeric mechanism.[9] However, it is not possible to determine the precise contribution of this mechanism to gluten elasticity because of the complexity of the system, with over 50 individual proteins being present. Analysis of heterologously expressed or in vitro synthesized proteins and peptides has been a powerful tool for determining the structure, biomechanical properties, and assembly of elastin.[3,10] We have therefore adopted a similar approach to determine the relationships between the sequences of recombinant peptides[11] based on HMW subunit repeat motifs and their ability to form hydrogen-bonded structures which can contribute to gluten elasticity.

Materials and Methods

Construction of Synthetic Genes Encoding Recombinant HMW Repeat Peptides. Sixteen pairs of complementary oligonucleotides were

* Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, U.K. Tel.: 0044 (0)1603 255012. E-mail: wellner@bbsrc.ac.uk.
† Institute of Food Research.
‡ Rothamsted Research.
§ University of East Anglia.

10.1021/bm050893t CCC: $33.50
Published on Web 03/10/2006

synthesized (Genosys, U.K.). See Supporting Information for sequences. The oligonucleotides were annealed (**1** with **2** (insert I); **3** with **4** (insert II); **5** with **6** (insert III); etc.) by mixing in equimolar amounts, heating briefly to 100 °C, and cooling slowly to 4 °C. The double-stranded (ds) DNA termed insert I was cloned into a NcoI and BamHI predigested plasmid pUCBM20 (Boehringer), termed pUCBM20-R0, and maintained in *E. coli* SURE cells (Stratagene). The ds DNA, insert IIa, was then cloned into a PstI predigested plasmid pUCBM20-R0 and termed pUCBM20-R1. Sequentially larger genes were made by cloning additional copies of insert IIa into the unique central PstI site in the gene. The X, N, and A nonperfect hexapeptide−nonapeptide (H−N) set of synthetic genes were constructed in a similar manner. The ds DNA, insert III, encoded the sequence for the X nonperfect H−N; the ds DNA, insert IV, encoded the sequence for the N nonperfect H−N, and the ds DNA, insert V, encoded the sequence for the A nonperfect H−N. The R5(X/N/A)1 series of genes were generated by ligating one of these inserts into a PstI predigested plasmid pUCBM20-R5. The R3(X/N/A)3 series of genes were generated by ligating three of these inserts sequentially into PstI predigested pUCBM20-R3. The R1(X/N/A)5 series of genes were generated by ligating five of these inserts sequentially into PstI predigested pUCBM20-R1. The Q and P nonperfect H−N set of synthetic genes were constructed in a slightly modified manner. The unique restriction site of PstI is positioned over the site of the Q and P modification. Thus, a ds DNA, insert IIb, was synthesized. This insert held a unique SmaI site allowing the ds DNA, insert IV, encoding the sequence for the Q nonperfect H−N and the ds DNA, insert V, encoding the sequence for the P nonperfect H−N to be ligated within the SmaI predigested DNA. The R5(Q/P)1 series of genes were generated by ligating one of these inserts into a SmaI predigested plasmid pUCBM20-R5(SmaI). The R3(Q/P)3 series of genes were generated by ligating three of these inserts sequentially into SmaI predigested pUCBM20-R3(SmaI). The R1(Q/P)5 series of genes were generated by ligating five of these inserts sequentially into SmaI predigested pUCBM20-R1(SmaI). All of the constructs were sequenced to confirm ligations had worked as expected (MWG-Biotech). An outline of this strategy is shown in Figure 1.

All thirteen synthetic genes were then excised from the pUCBM20 vectors using the restriction sites of NcoI and BamHI at either end and directionally cloned into the expression vector: pET28b+ or pET3d (Novagen). These constructs were maintained in *E. coli* JM109 (Promega) cells before transference to *E. coli* BLR (DE3) pLysS (Novagen), the strain used for protein expression.

Expression and Purification of Peptides. The recombinant peptides were produced by growing selected strains in 2YT culture (30 μg/mL kanamycin), at 30 °C, in a shaking orbital incubator (220 rpm) until an optical density (OD) of $A_{600} = 0.6$ was reached. Expression of the protein was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Incubation continued for a further 18 h. The cell cultures were harvested by centrifugation (6500 *g*). The peptides were purified on a semipreparative scale, by resuspending the cell pellets in 20 mL 90% (v/v) ethanol, 2% (w/v) 1,4-dithio-DL-threitol (DTT) in water per liter of culture, and subjecting the resulting suspension to the following treatments: (i) a heat treatment (60 °C, 90 min) of the suspension, followed by a centrifugal separation (10 000 *g*, 10 min) of the cell debris and the soluble fraction; (ii) an acetone (4 vol) precipitation of the soluble protein; (iii) a resolubilization of the precipitate in 8 M urea; 2% (w/v) DTT; (iv) a filtration through a 0.2 μm filter (Schleicher and Schuell); (v) separation using reversed-phase HPLC on a C18 column (Vydac). Elution was accomplished with a linear gradient of 0−40% (v/v) aqueous acetonitrile in 0.05% (v/v) trifluoroacetic acid over 50 min at a flow rate of 2.5 mL/min. Eluates were monitored at 280 nm. The major peaks eluting at 32−36% were collected and lyophilized. SDS-PAGE was performed essentially according to Laemmli[12] using a BioRad Mini Protean II system (BioRad) with 15% acrylamide gels. Lyophilized samples were mixed in 25 mM Tris/HCl, pH 6.8, 0.5% (w/v) SDS, 5% glycerol, 0.01% bromophenol blue, and 0.5% (v/v) 2-mercaptoethanol and boiled for 5 min prior to loading. Gels were stained with Coomassie brilliant blue R-250 dissolved in destaining solution (10% (v/v) acetic acid, 5% (v/v) methanol in water).

FT-IR Spectroscopy. FT-IR spectra of the peptides were obtained using a BioRad FTS175 (DigiLab, U.S.) spectrometer equipped with a HgCdTe detector and a single-reflection diamond attenuated total reflection (ATR) sampling accessory (Graseby Specac, Orpington, U.K.). Each spectrum was averaged from 256 scans at 2 cm^{-1} resolution and referenced against a background spectrum of the empty ATR crystal. Freeze-dried peptides were equilibrated in water (10 mg/mL). The pH value of the supernatant was checked and if necessary adjusted with 0.1 N NaOH. Only a few samples dissolved; most formed viscous, sticky pellets with little peptide detectable in the supernatant. Spectra were recorded after equilibrium was reached. Spectra of liquid water, and where necessary, water vapor, were subtracted manually. All samples were measured in duplicate. To compare the secondary structures of the peptides, the infrared spectra were baseline-corrected (absorbance at 1800 cm^{-1} set to zero) and normalized to the same height of the amide I band. The relative intensities at 1666, 1650, 1630, and 1620 cm^{-1} were calculated as the fraction of their sum. The significance of these values is discussed in the Results section. Secondary structure estimates obtained in this way were more reproducible than when band-fitting was used (less than 0.1% standard deviation compared to 0.7−1.0%), and therefore, small changes in structure could be reliably detected.[7]

Results

Peptide Design. The HMW subunits of wheat are classified into two subgroups called x-type and y-type, which differ in the detailed structure of their repetitive domains.[13] Both comprise nona- and hexapeptide repeat motifs which always occur in alternate fashion to give a 15-amino-acid block and also tandem arrays of hexapeptide repeats alone. In addition, the x-type subunits contain tripeptides which are also always interspersed with hexapeptides to give a second nonapeptide motif. Although the motifs are conserved in length, there is variation in the frequency of amino acid substitutions, both at different positions within the motifs and at the same position within the corresponding motifs of x-type and y-type subunits (see Supporting Information). We decided to use the consensus 15-amino-acid ^{1}PGQGQQGYYPTSLQQ^{15} (i.e., hexapeptide + nonapeptide (H−N)) motif of the y-type subunits, as a basis for creating a range of mutants to explore the effects of single or combinations of amino acid substitutions on the structures and interactions of the peptides, using a standard peptide length of 203 amino acids (Figure 2a). The synthetic gene, encoding the standard peptide, was created by the successive addition of ds DNA, encoding the H−N block, into either a similar ds DNA or into a "starter" ds DNA, encoding the N- and C-terminal ends (16 and 7 amino acids, respectively). The N-terminal included a cysteine residue as a potential cross-linking site and an N-terminal methionine encoded by the ATG initiation codon. The C-terminal included a cysteine residue and a stop codon.[11] This strategy is summarized in Figure 1 and described in detail in Materials and Methods. Incorporation of blocks containing mutations allowed the production of forms containing 1 (30 residues), 3 (90 residues), or 5 (150 residues) blocks of mutant repeats, within the standard peptide containing 6 blocks (180 residues). These are called R5−1, R3−3, R1−5, and R6, respectively (Figure 2b,c).

The following mutations were designed for analysis:

(i) The first mutant peptide was not based on naturally occurring substitutions but designed to introduce a high propor-

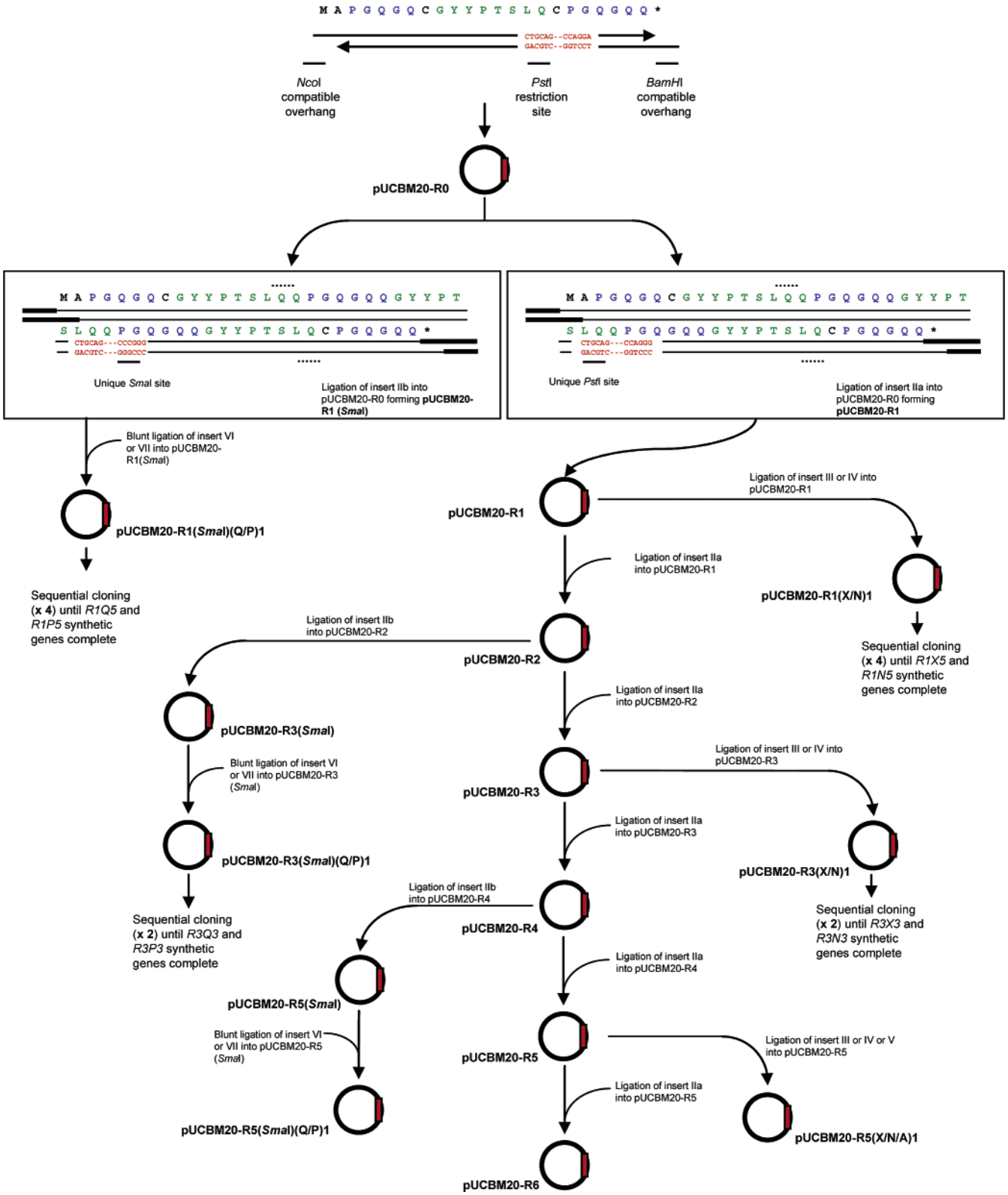

Figure 1. Overview of the construction of the standard and variant synthetic genes.

tion of charged residues without disrupting the predicted β-turn-rich structure (based on the use of standard algorithms as discussed previously). This mutant sequence contained Glu instead of Gln/Gly at positions 5, 6, 7, and 15 and Asp instead of Tyr at position 9, and is called A for "acidic" for convenience.

(ii)/(iii) Leu occurs at position 13 of the 15-amino-acid motif in approximately 54% of y-type repeats but in only 11% of x-type repeats. Two mutations were therefore constructed in which this residue was replaced by the two most frequently occurring substitutions: Pro (70% x-type, 21% y-type) and Gln (19% of y-type, 0% x-type). The sequence with the Leu13Pro mutation is called P. The sequence with the corresponding Leu13Gln mutation is called Q.

(iv) Tyr and Thr occur at positions 8 and 11 in 98% and 96% of x-type motifs, respectively, but only 54% and 60%, respectively, in y-type motifs. The most frequent substitutions in y-type motifs are Tyr8His (42%) and Thr11Ala (37%), and these usually occur together within the same motif. These

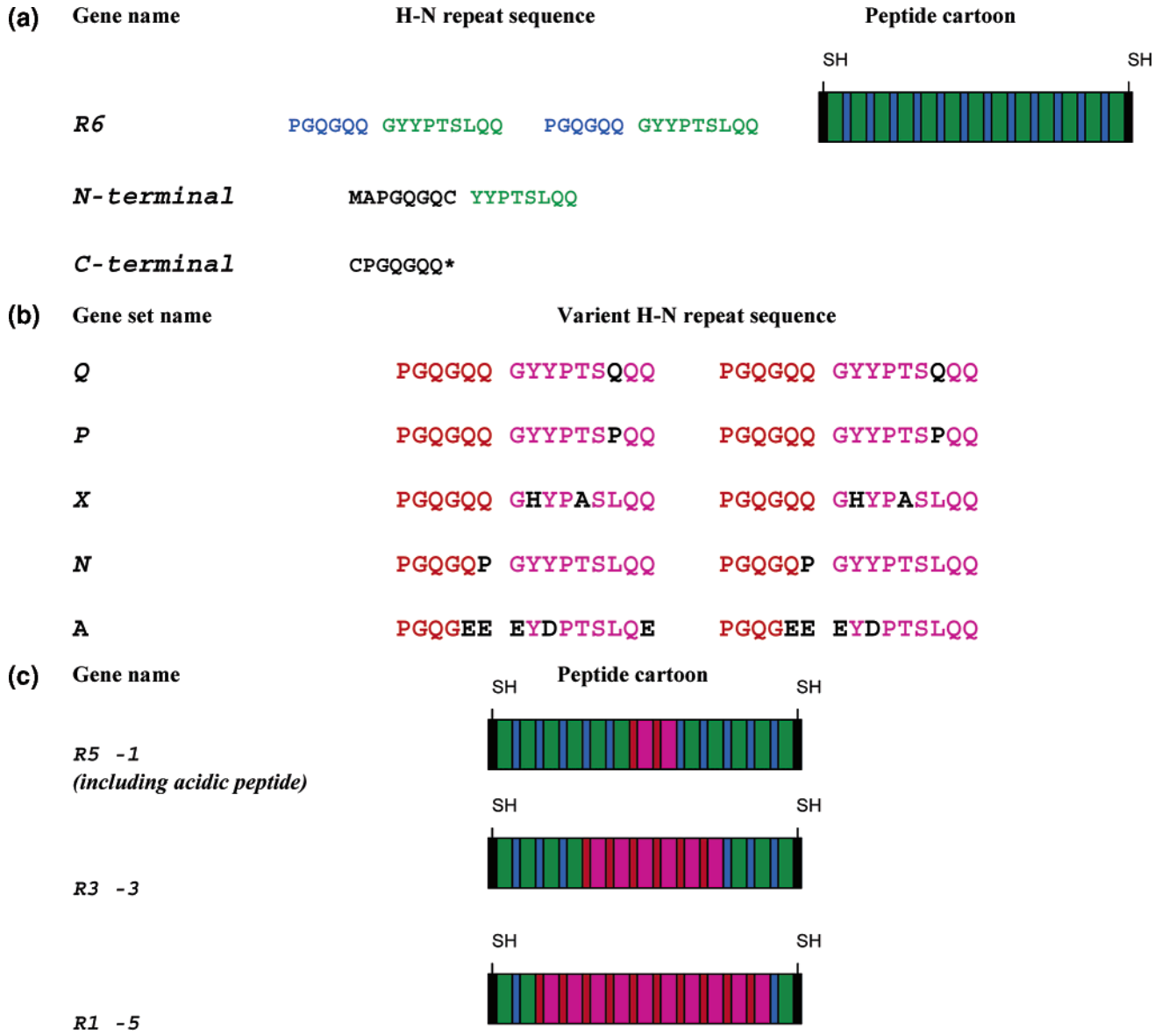

Figure 2. Repeat sequences: (a) standard peptide, (b) the four variant H−N sequence blocks and the acidic sequence block, (c) position and number of substituted sequence blocks.

substitutions were therefore combined to give Tyr8His/Thr11Ala and the corresponding repeat sequence denoted as X.

(v) Gln occurs at position 6 in 94% of y-type motifs but only 80% of x-type motifs. The most frequent substitution in the latter is Pro (15%), which was used to design Gln6Pro substituted sequences, denoted as N.

It was expected that the deprotonated form of the acidic A sequence in the middle of the peptide would prevent aggregation. In contrast, the QQQ motif in the Q sequence was expected to confer greater aggregation. Tyrosine stacking may also help to stabilize the aggregates, and the substitution of one Tyr in the X sequence could therefore make the peptide more soluble. The N and P sequences both introduced prolines, which should prevent β-sheet formation and thus aggregation. While one copy of the A mutation already had a strong effect on the peptide, the influence of the other mutations was weaker. Therefore, either one, three, or five copies of the N, P, Q, and X sequences were incorporated to determine the relationship between the frequency of these mutations and the effects on properties.

Expression and Purification. Constructs encoding 14 peptides (R6, R5A1, and R5−1, R3−3, and R1−5 forms of the other 4 mutants) were expressed in *E. coli* using pET28b+ or pET3d vectors (Novagen) and purified on the basis of their solubility in 90% (v/v) ethanol and 2% (w/v) DTT, followed by an acetone precipitation and RP-HPLC. SDS-PAGE (Figure 3) showed that they were essentially pure, although traces of dimeric forms were observed for some peptides. There were however some differences in mobility, with the acidic peptide and those containing 3 and 5 copies of the Q and P mutations

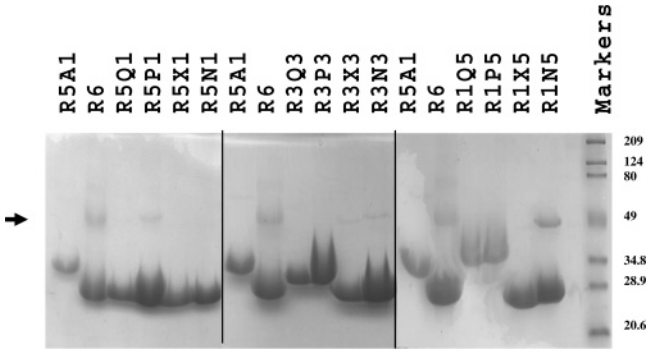

Figure 3. SDS-PAGE gel showing different mobilities of the RP-HPLC purified peptides. The arrow marks dimers.

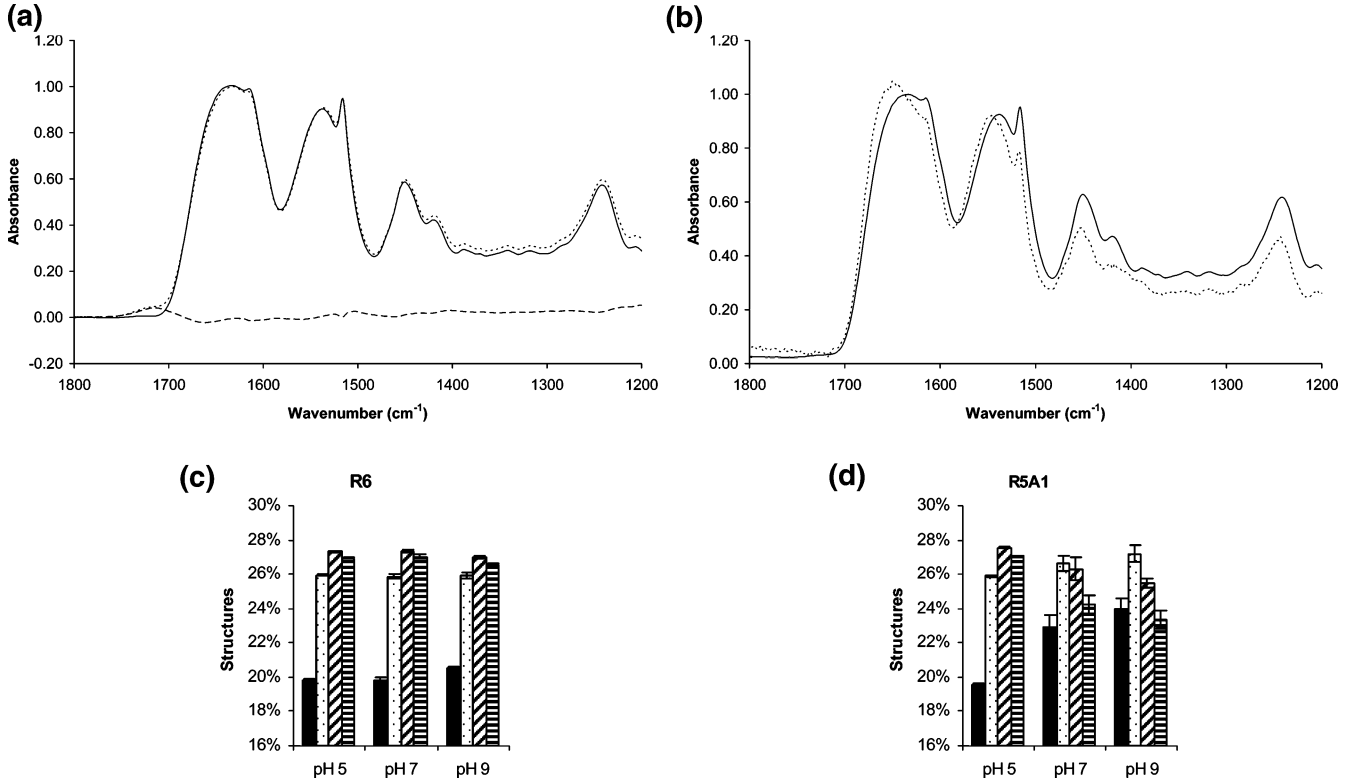

Figure 4. Comparison of R6 and R5A1 peptides: (a) FT-IR spectra of (—) R6 and (···) R5A1 at pH 5 (water spectra subtracted), (- - -) difference spectrum R5A1 − R6; (b) FT-IR spectra of (—) R6, (···) R5A1, at pH 9 (water spectra subtracted); (c) estimated secondary structures in R6 at different pH values; and (d) estimated secondary structures in R5A1 at different pH values. Key to (c) and (d): solid black bar, β-turn; dotted bar, random; diagonal hashed bar, intramolecular β-sheet; horizontal hashed bar, intermolecular β-sheet.

being particularly retarded. This presumably resulted from effects of the mutations on the proportion of SDS-resistant secondary structure.

Effect of Mutations on the Structure and Properties of Repetitive Peptides. *(i) Acidic Mutation.* Our previous studies of the perfect repeat peptide R6, used as the parent peptide in the current study, have shown that in the hydrated state an equilibrium was present between helical structure (β-reverse turns and poly-L-proline II) and β-sheet, in particular, intermolecular β-sheet, which rendered the peptide water-insoluble.[6] HMW subunits and other gluten proteins generally have few charged residues at neutral pH. The acidic peptide R5A1, containing one block of the A mutant repeat, was therefore designed to determine the effect of electrostatic charge repulsion on this structure equilibrium. At pH 5, the R5A1 peptide was as insoluble as the R6 peptide. Its infrared spectrum was also similar to that of R6, the most marked difference being the presence of a νC═O band at 1720 cm^{-1} from carboxylic acid side groups of the introduced glutamic acid and aspartic acid residues in R5A1 (Figure 4a). The difference spectrum of the two peptides under these conditions also revealed a loss of intensity around 1662 cm^{-1} due to the loss of glutamine side chains, which absorb at 1658 cm^{-1},[14] and there was also a small negative band at 1517 cm^{-1} in the amide II band from the replaced Tyr residue. The overall similarity of the amide bands of these two peptides indicated that their secondary structures were comparable, in particular, their proportions of intermolecular β-sheet. This indicated that both glutamic and aspartic acid residues could be incorporated into the secondary structure without major effects on the conformation. The spectrum, and by implication, the structure of R6, was largely unchanged between pH 5 and pH 9. However, the R5A1 peptide was readily water-soluble at pH 7 and above, whereas R6 remained virtually insoluble. This indicated that the negative charges on only one block of the mutant A sequence were enough to make the peptide water-soluble. At pH 9, the νC═O band at 1720 cm^{-1} was absent from the spectrum of the R5A1 peptide, confirming that the Asp and Glu residues were deprotonated. There was also a marked reduction of the absorption in the 1630−1610 cm^{-1} region of the amide I band and in the 1530−1510 cm^{-1} region of the amide II band, which are characteristic of β-sheet structure.[15,16] Instead, the main peaks were at 1655 and 1550 cm^{-1}. The secondary structures were estimated by the relative intensities at four characteristic positions in the amide I band: 1666 cm^{-1}, indicating β-reverse turn structures; 1650 cm^{-1}, unordered structure and glutamine side chain contributions; 1630 cm^{-1}, intermolecular β-sheet; and 1616 cm^{-1}, intramolecular β-sheet.[15−19] Thus, R5A1 in solution had an increased amount of unordered structures and β-turns and a loss of β-sheet structures, in particular, intermolecular β-sheet (Figure 4d). The conversion of intermolecular β-sheet into β-turn structures has been observed as a typical feature whenever aggregated gluten proteins are dissolved.[8] The behavior is consistent with the general observation that aggregation of gluten proteins, in particular, in the hydrated solid state, is mediated by intermolecular sheets;[20] whereas in solution, the molecules adopt a helical structure or β-spiral or other extended structure, which is characterized by repetitive β-turns interspersed with poly-L-proline II structure.[6,8,21]

(ii) Leu13Pro Substitution. The gluten protein repeat sequences are rich in proline, a residue that terminates regular structures and severely disrupts the formation of β-sheet. The repeat sequence of the peptides prepared in this study has two stretches which can adopt β-strand structure and thus are possible candidates for intermolecular β-sheet formation—a short five-residue stretch from Thr11 to Gln15 and a longer eight-residue stretch from Gly2 to Tyr9, both flanked by prolines. The replacement of Leu with Pro at position 13 introduces an

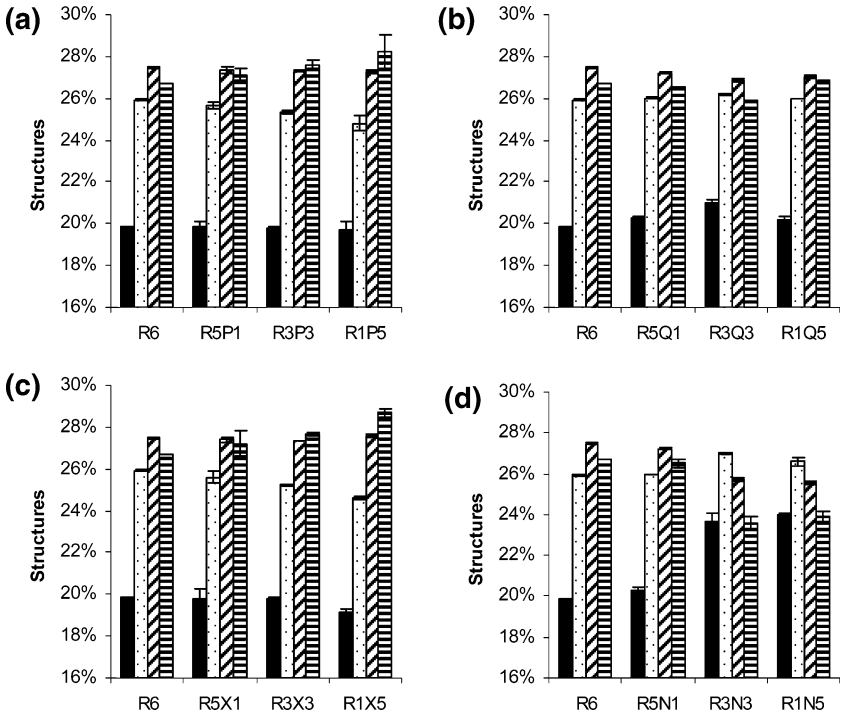

Figure 5. Effect of increasing numbers of substitutions on the secondary structures of peptides in water: (a) Leu13Pro; (b) Leu13Gln; (c) Tyr8His/Thr11Ala; (d) Gln6Pro. Key: solid black bar, β-turn; dotted bar, random; diagonal hashed bar, intramolecular β-sheet; horizontal hashed bar, intermolecular β-sheet.

additional "kink" in the middle of the shorter stretch, preventing it from participating in the formation of β-sheets. Consequently, the amount of intermolecular β-sheet might be expected to decrease as increasing numbers of these modified repeats are introduced. However, the opposite result was observed, with all three substituted peptides being practically insoluble in water. In fact, the infrared spectra showed that the content of intermolecular β-sheet actually increased from R6 to R1P5 (Figure 5a). This was accompanied by a decrease in unordered structure. The proportions of other structures remained fairly constant. This suggests that the additional proline residue, rather than impeding β-sheet formation, may have actually promoted it. A possible explanation would be that the section of sequence containing the substitution does not normally participate in β-sheet formation but acts as a randomly structured spacer between stretches of β-sheet. The substituted proline could therefore assist in the formation of a more tightly folded structure which may be sterically more favorable for the aggregation of the whole molecule.

(iii) Leu13Gln Substitution. All three substituted peptides R5Q1, R3Q3, and R1Q5 were as insoluble in water as the R6 peptide. Substitution of the hydrophobic leucine residues with hydrophilic glutamine residues did not increase the solubility, although increases in the degree of substitution resulted in small changes in the infrared spectra (Figure 5b). Glutamine accounts for 32%, 33%, 35.5%, and 37.5% of the amino acid content in R6, R5Q1, R3Q3, and R1Q5, respectively, and has an absorption at 1658 cm^{-1} due to its amide side chain.[13] Consequently, the amide I band component at around 1660 cm^{-1} increased gradually as the degree of substitution increased. In general, the structures of the four peptides were dominated by inter- and intramolecular β-sheet and unordered structures, each accounting for around 20% of the total structure. Glutamine residues are known to promote intermolecular aggregation of proteins, and the stability of these aggregates increases with the number of consecutive glutamine residues.[22] Therefore, it was expected that extending existing diglutamine stretches in the repeat sequence to triglutamine should favor aggregation. However, R3Q3 had the highest amounts of intramolecular β-sheet and unordered structures and the lowest amount of intermolecular β-sheet and a slightly higher content of β-turns. Thus, additional glutamine residues did not increase the amount of intermolecular β-sheet as expected. It is possible that the effects of the glutamine substitution were tempered by a steric effect, as it was placed in a short five-residue stretch flanked by proline residues on either side, which would restrict the ability to form an extended β-strand structure. In this case, the hydrophilic glutamine residues would increase the hydration and hence the tendency of the peptide chain to form turns or unordered structures. This effect appeared to dominate in R5Q1 and R3Q3 with the additional glutamine side chain only contributing to the content of intermolecular β-sheet in R1Q5.

(iv) Tyr8His/Thr11Ala Substitutions. The peptides R6, R5X1, R3X3, and R1X5 were all virtually insoluble in water at neutral pH. FT-IR spectra of these peptides showed clear increases in the 1616 cm^{-1} band with increasing number of Tyr8His/Thr11Ala substitutions. This indicated the formation of increasing amounts of intermolecular β-sheet, accompanied by decreasing contents of unordered (1650 cm^{-1}) and β-turn structures (1666 cm^{-1}) (Figure 5c). It had been postulated that tyrosine stacking may help stabilize gluten protein aggregates, and a substitution replacing Tyr8 would therefore be expected to increase solubility. However, the infrared data presented here suggest that the substitution of tyrosine by histidine actually increased the ability of the peptides to aggregate. A possible explanation for this observation comes from models of the HMW subunit repetitive domain. These showed that the side chains of two adjacent tyrosine residues interact in the center of the β-spiral and thus help stabilize the structure.[23,24] Consequently, the substitution of a tyrosine residue would destabilize the soluble β-spiral structure and shift the equilibrium toward the formation of more insoluble intermolecular β-sheet

structures. Whereas the role of the tyrosine residue may be explained, the effect of the accompanying Thr8Ala substitution is not clear.

(v) Gln6Pro Substitution. Substitution of glutamine by proline at position 6 resulted in a dramatic change in the properties of the peptides. The peptide R5N1 was as insoluble as R6, but the R3N3 and R1N5 peptides were readily soluble in water at neutral pH. The replacement of glutamine with a proline residue would be expected to introduce a "kink" into the sequence which would prevent the formation of β-sheet. This assertion was supported by the FT-IR data. The dissolved peptides R3N3 and R1N5 had much lower contents of intermolecular β-sheet and higher contents of β-turns than the insoluble R6 and R5N1 peptides (Figure 5d). The effect of the substitution was dose-dependent. One modified sequence block (R5N1) could be accommodated without affecting solubility, but the presence of three or five adjacent modified sequence blocks out of six clearly shifted the balance between protein−protein and protein−water interactions toward the latter. The results suggest that the eight-residue stretch from Gly2 to Tyr9 is indeed crucial for the formation of intermolecular β-sheet in gluten protein aggregates. Few structural differences were observed between R3N3 and R1N5, which had three and five modified sequence blocks, respectively. Both dissolved peptides adopted a typical β-turn-rich structure in solution, and this structure did not appear to be greatly affected by the substitutions. Modeling suggested that this solution structure is open and flexible;[24] therefore, it could easily accommodate proline substitutions.

Discussion

Elastomeric properties are exhibited by a number of proteins, which have different evolutionary origins, but all exist as large aggregates or arrays, which are stabilized by covalent cross-links and/or strong noncovalent interactions. The molecular mechanisms of these elastomeric protein systems, which include spider silks, gluten, and elastin, may vary in detail, but all generally involve both entropic and enthalpic contributions.[2] The entropic component requires flexible sequences which can be easily deformed, while the enthalpic contribution arises from intermolecular interactions. Influencing the relative contributions of these mechanisms could therefore result in fine-tuning of the physicochemical properties of the elastomers. In wheat gluten, the noncovalent intermolecular interactions are driven by intermolecular β-sheet formation between the repetitive domains of gluten proteins. We have therefore studied the effects of amino acid substitutions, based on those found in nature, on the structures and the aggregation properties of a series of synthetic perfect repeat peptides based on the elastomeric HMW subunits of wheat gluten. Incorporation of aspartic acid and glutamic acid residues had no effect at low pH, but once deprotonated, their electrostatic charges rendered the molecule water-soluble. This effect may not be desirable in a storage protein such as gluten, as it could affect the ability of the proteins to be packaged in dense deposits within the cells of the developing grain. This may provide an explanation for the low charge density of wheat gluten and the related storage proteins of other cereals. Tyrosine residues did not appear to stabilize the aggregation of the peptides but may stabilize the coiled structure of the dissolved molecules and thus move the equilibrium away from aggregation. Although the peptides contained single cysteine residues at each end, these did not form significant proportions of disulfide bonds, with aggregation of the peptides occurring mainly via the formation of intermolecular β-sheet structures.[6,8,10] The peptide sequence had two stretches of amino acid residues (GQGQQGYY and TSLQQ) which have high probabilities of β-sheet formation, which probably accounted for the insolubility. The TSQQQ mutation should have strengthened the propensity of the shorter stretch to form β-sheet structures, because additional glutamine residues would be expected to promote aggregation, but the results indicate that if aggregation is hindered by sterical reasons the additional polar side chain may instead promote hydration. Equally, although this stretch was disrupted by a proline in the TSPQQ sequence, those peptides remained insoluble. Thus, the TSLQQ sequence may not contribute to the intermolecular β-sheet structure, but it could influence aggregation in other ways. Notably, the conformational change resulting from the Leu13Pro substitution may improve the steric interactions between the GQGQQGYY sequences and so make aggregation more favorable overall. In contrast, the GQGQPGYY sequence resulting from the Gln6Pro mutation had little propensity to form β-sheet, and the R3N3 and R1N5 peptides differed from the others in being water-soluble at neutral pH. There was also a clear "dosage" relationship between the number of modified peptides inserted and the effects on the properties.

We have shown, therefore, that the introduction of naturally occurring mutations into a perfect repeat peptide based on the elastomeric HMW subunits of wheat results in changes in their structures and properties, including their solubility in water. These proteins are immensely important in wheat utilization, as their elastomeric properties are largely responsible for the ability to make bread, pasta, and many other food products.[1] Differences in the amount and properties of these proteins relate to differences in processing quality between flours from different wheat types. The results described here therefore provide information on the molecular basis for these differences. In the wider context of elastomeric proteins, this study has shown that mutations in the sequence provide another way to control the physicochemical properties besides molecular weight and cross-linking. The spectrum of elastomeric proteins ranges from the very hydrophobic elastin to the hydrophilic gluten. Generally, their repetitive domains have conserved sequences and secondary structures, which could be regarded as optimized for the purpose of each of these proteins. Mutations which destabilize these structures would diminish the functionality, but the effect of several mutations appears to be cumulative; therefore, some variation may be tolerated. On the other hand, specific variations of repeat sequences may be utilized intentionally, as in the case of different spider silks.[2]

Acknowledgment. Rothamsted Research and the Institute of Food Research receive grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. The work described was supported by BBSRC competitive research grant no. D14544.

Supporting Information Available. Tables showing the sequences of the oligonucleotides used in gene construction and the positional occurrences of amino acids in the x- and y-type HMW glutenin repeats. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Shewry, P. R.; Halford, N. G.; Tatham, A. S.; Popineau, Y.; Lafiandra, D.; Belton, P. S. *Adv. Food Nutr. Res.* **2003**, *45*, 221−302.
(2) Tatham, A. S.; Shewry, P. R. *Philos. Trans. R. Soc. London* **2002**, *357*, 229−234.
(3) Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T. *Philos. Trans. R. Soc. London* **2002**, *357*, 169−184.

(4) Tatham, A. S.; Hayes, L.; Shewry, P. R.; Urry, D. W. *Biochim. Biophys. Acta* **2001**, *1548*, 187−193.
(5) Tatham, A. S.; Drake, A. F.; Shewry, P. R. *J. Cereal Sci.* **1990**, *11*, 189−200.
(6) Feeney, K. A.; Wellner, N.; Gilbert, S. M.; Halford, N. G.; Tatham, A. S.; Shewry, P. R.; Belton, P. S. *Biopolymers* **2003**, *72*, 123−131.
(7) Gilbert, S. M.; Wellner, N.; Belton, P. S.; Greenfield, J. A.; Siligardi, G.; Shewry, P. R.; Tatham, A. S. *Biochim. Biophys. Acta* **2000**, *1479*, 135−146.
(8) Wellner, N.; Belton, P. S.; Tatham, A. S. *Biochem. J.* **1996**, *319*, 741−747.
(9) Belton, P. S. *J. Cereal Sci.* **1999**, *29*, 103−107.
(10) Keeley, F. W.; Bellingham, C. M.; Woodhouse, K. A. *Philos. Trans. R. Soc. London* **2002**, *375*, 185−189.
(11) Feeney, K. A.; Tatham, A. S.; Gilbert, S. M.; Fido, R. J.; Halford, N. G.; Shewry, P. R. *Biochim. Biophys. Acta* **2001**, *1456*, 346−355.
(12) Laemmli, U. K. *Nature (London)* **1970**, *227*, 359−365.
(13) Payne, P. I. *Annu. Rev. Plant Phys. Plant Mol. Biol.* **1987**, *38*, 141−153.
(14) Venyaminov, S. Y.; Kalnin, N. N. *Biopolymers* **1990**, *30*, 1243−1257.
(15) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469−487.
(16) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *30*, 95−120.
(17) Mantsch, H. H.; Perczel, A.; Hollósi, M.; Fasman, G. D. *Biopolymers* **1993**, *33*, 201−207.
(18) Surewicz, W. K.; Mantsch, H. H.; Chapman, D. *Biochemistry* **1993**, *33*, 389−394.
(19) Bandekar, J. *Biochim. Biophys. Acta* **1992**, *1120*, 123−143.
(20) Pézolet, M.; Bonenfant, S.; Dousseau, F.; Popineau, Y. *FEBS Lett.* **1992**, *299*, 247−250.
(21) Blanch, E. W.; Kasarda, D. D.; Hecht, L.; Nielsen, K.; Barron, L. D. *Biochemistry* **2003**, *42*, 5665−5673
(22) Perutz, M. F.; Johnson, T.; Suzuki, M.; Finch, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5355−5358.
(23) Purcell, J. M.; Kasarda, D. D.; Wu, C. S. C. *J. Cereal Sci.* **1988**, *7*, 21−32.
(24) Parchment, O.; Shewry, P. R.; Tatham, A. S.; Osguthorpe, D. J. *Cereal Chem.* **2001**, *78*, 658−662.

BM050893T