Molecular Modeling of the N-terminal Regions of High Molecular Weight Glutenin Subunits 7 and 5 in Relation to Intramolecular Disulfide Bond Formation

PETER KÖHLER,[1] BETTINA KECK-GASSENMEIER,[1,3] HERBERT WIESER,[1] and DONALD D. KASARDA[2]

Cereal Chem. 74(2):154–158

ABSTRACT

Analyses of cystine peptides derived from the high molecular weight glutenin subunits (HMW-GS) 5 and 7 indicate that, in spite of a distinct sequence homology between the two subunits in the N-terminal region, different disulfide linkages of cysteine residues are present in these regions. To investigate the structural basis for these experimental results, the conformational structures of the polypeptide chains corresponding to the N-terminal regions (first 50 amino acids) of the wheat HMW-GS 5 and 7 were modeled by computer methods. Secondary structures were predicted by the method of Rost and Sander (1993) and, to the extent appropriate, applied to the constructed polypeptide chains. The resulting structures were energy-minimized and subjected to simulated heating and dynamic equilibration. In the final structure of subunit 5, the first two cysteines were located in a region of continuous α-helix. If folding to the helical form occurs rapidly during biosynthesis as expected, the distance between the sulfhydryl groups of these two cysteines would be great enough (≈2.2 nm) to make intramolecular disulfide bond formation unlikely. Although a somewhat similar region of α-helix was predicted for the subunit 7, in some predictions the helix was interrupted between the first two cysteines, and this break was assigned either extended structure or arbitrarily modeled as an inverse γ-turn. In the final structure of subunit 7 with the assigned inverse γ-turn, after energy minimization, heating, and dynamics, the two cysteines approached one another closely (≈0.4 nm). Formation of an intramolecular disulfide bond appeared a likely possibility. This model is in accord with experimental evidence for this latter intramolecular bond (Köhler et al 1993). In agreement with the modeling, an equivalent intramolecular disulfide bond of subunit 5 has not been found and experimental evidence for a different arrangement is presented.

The N-terminal domains of x-type high molecular weight glutenin subunits (HMW-GS), e.g., subunits 5 and 7, have three cysteine residues designated as C^a, C^b, and C^d (Köhler et al 1993). The results from previous disulfide formation studies indicate that, in spite of a distinct sequence homology (Anderson and Greene 1989, Anderson et al 1989), different linkages of cysteine residues appear to be present in this domain. It was shown previously that C^a and C^b of subunit 7 form an intramolecular disulfide bond (Köhler et al 1993), whereas C^d is available for the formation of an intermolecular linkage. The work presented here indicates that C^d of subunit 5 is linked with C^b, although so far the type of linkage (intra- or intermolecular) and the origin of C^b have not been determined. We examined possible structural explanations for the different disulfide bonds formed by the homologous cysteine residues in the N-terminal regions of subunits 5 and 7 by computer molecular modeling.

MATERIALS AND METHODS

Isolation and Characterization of a Cystine Peptide from Enriched Glutenin

The thermolytic digest of enriched glutenin was fractionated by gel-permeation chromatography (GPC) on Sephadex G25. The peptide obtained in fraction 3 was further separated by reverse-phase high-performance liquid chromatography (RP-HPLC) on C_{18} silica gel (Keck et al 1995). Cystine peptides were identified by differential chromatography (RP-HPLC before and after reduction of disulfide bonds) (Köhler et al 1991). Cystine peptide fraction $P_{RTh}3/13$ was collected from several preparative runs and purified by rechromatography. The major peptide obtained ($P_{RTh}3/13a$) was reduced with dithioerythritol and alkylated with 4-vinylpyridine. The resulting constituent peptides were separated by RP-HPLC and analyzed for their amino acid sequences according to previous work (Köhler et al 1993, Keck et al 1995).

Amino Acid Sequences for Molecular Modeling

The N-terminal amino acid sequences of subunits 5 and 7 through 50 residues were taken from Anderson et al (1989) and Anderson and Greene (1989), who determined the gene DNA sequences. Both genes were obtained from a genomic DNA library from the cultivar Cheyenne. The sequences in single letter code are as follows (numbers correspond to the positions of cysteine residues in the sequences):

HMW-GS 5:

```
         10              25              40
EGEASEQLQCERELQELQERELKACQQVMDQQLRDISPECHPVVVSPVAG-
```

HMW-GS 7:

```
         10     17             32
EGEASGQLQCEHELEACQQVVDQQLRDVSPGCRPITVSPGTRQYEQQPVV-
```

Structure Prediction

Secondary structure prediction based on the sequences was performed by the PHD method of Rost and Sander (1993, 1994), which uses sequence profiles and neural networks. The program was utilized by way of Internet access to the Predict Protein server at Heidelberg, Germany (Rost et al 1994). Two differently trained architectures (versions) of the secondary structure prediction software were used: 11.93 and 5.94_317.

Molecular Modeling

A Silicon Graphics, Inc. (Mountain View, CA), Personal Iris W-4D/35workstation running Quanta 4.0 and CHARMm (version

[1]Deutsche Forschungsanstalt für Lebensmittelchemie and Kurt-Hess-Institut für Mehl- und Eiweissforschung, Lichtenbergerstrasse 4, D-85748 Garching, Germany.

[2]USDA, Agricultural Research Service, Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710. Mention of a product is for informational purposes only and is not meant to imply recommendation by the U. S. Department of Agriculture over others that may be suitable.

[3]Present address: Nestlé Ltd. Research Centre, PO Box 44, Lausanne 26, CH 1000, Switzerland.

Publication no. C-1997-0205-05R.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. American Association of Cereal Chemists, Inc., 1997.

23.1) software from Molecular Simulations, Inc. (San Diego, CA) was used for molecular modeling. The Quanta software was used to construct peptides equivalent to the first 50 residues of the N-terminal sequences of subunits 5 and 7 from a library of geometric parameters included in the software and to apply the secondary structure predictions. Because the predictive method does not specify the peptide backbone ϕ, ψ dihedral angles for amino acid residues involved in loops or turns, these angles were not applied to the resulting peptide where predicted, with the exception of one inverse γ turn. The ϕ, ψ dihedral angles for the central residue of the three-residue inverse γ-turn were arbitrarily set as –75 and 59º, which are within the range found for γ-turns in proteins (Milner-White 1990, Kumosinski et al 1993). After construction of the peptide in the computer and application of secondary structure, the resulting structure was energy-minimized in CHARMm by the method of Steepest Descents, followed by the Adopted Basis Newton-Raphson (ABNR) method, both of which are part of the software package. Minimization generally consisted of 100 steps with the method of Steepest Descents followed by 500 steps with the ABNR method. Minimization parameters were: Coordinate Update Frequence, 5; Energy Gradient Tolerance, 0.01; Energy Value Tolerance, 0; Initial Step Size, 0.02; and Step Value Tolerance, 0.

Heating from the equivalent of 0–300ºK was then simulated, followed by equilibration of the energy of the molecular system for the equivalent of 30 psec. Heating parameters used were: Output Frequency, 10; Time Step, 0.001. Equilibration parameters used were: Output Frequency, 10; Equilibration Frequency, 200; Time Step, 0.001. The force field used in CHARMm to define the energy of the molecule during these calculations is described by Brooks et al (1983). Finally, the molecule was energy-minimized once again by the ABNR method until no significant change was noted in the resulting energy of the system. No attempt was made to include solvent in the modeling.

RESULTS AND DISCUSSION

Experimental Evidence for Disulfide Bonds of Subunits 5 and 7

Evidence for an intramolecular disulfide bond between the cysteine residues C^a and C^b of subunit 7 was previously provided by the isolation and analysis of the cystine peptide $P_{Tr}2/1$ (Table I) from a tryptic digest of whole glutenin (Köhler et al 1993). In a continuation of recent studies on the thermolytic digest of enriched glutenin (Keck et al 1995), peptide fraction 3 obtained by GPC on Sephadex G25 was further separated by RP-HPLC on C_{18} silica gel (Fig. 1). Cystine peptides were identified by differential chromatography. The cystine peptide fraction $P_{RTh}3/13$ (peak 13) was isolated by preparative RP-HPLC and purified by rechromatography (chromatogram not shown). The major peptide obtained ($P_{RTh}3/13a$) was reduced and alkylated with 4-vinylpyridine. A further separation then yielded the two S-alkylated peptides (13a-1, 13a-2), which were analyzed for amino acid sequences. The results (Table I) show that fragment 13a-2 was derived from subunit 5 (Anderson et al 1989), whereas fragment 13a-1 occurs in all the HMW-GS present in the cultivar Rektor investigated to-date (subunits 5, 7, 9, 10). Thus, peptide $P_{RTh}3/13a$ represents a disulfide bond between C^d of subunit 5 with C^b of subunit 5 or some other subunit. Even though the type of linkage (intra- or intermolecular) and the origin of C^b could not be determined, the results indicate that the disulfide formation within the homologous N-terminal regions of subunits 5 and 7 differs. Although we cannot say with certainty that no other possible connections involving the cysteine side chains in question exist in glutenin, we have no significant findings that point in that direction at present. These considerations served as the basis for the computer modeling studies.

Molecular Modeling

HMW-GS 5. Secondary structure prediction for subunit 5 by the PHD method indicated α-helix for residues 5–32 without interruption. This region includes the first two cysteines C^a and C^b (residues 10 and 25) of the N-terminal sequence. Other predictions were for extended structure and loops, although in this program loop means neither α-helix nor extended structure and so may include turns of various types as well as unordered structure. For some residues, no prediction was significant. Because the default condition is extended structure and the predictive method provides no indication of what types of loops or turns are present (i.e., what torsional angles might be applied), only the predicted α-helical structure was applied to the molecule. During energy minimization, the energy of the peptide rapidly became negative

TABLE I
Amino Acid Sequence and Origin of Cystine Peptides Derived From High Molcular Weight Glutenin Subunits (HMW-GS)

Peptide	Sequence	HMW-GS	Cysteine Residue[a]
$P_{Tr}2/1$	EGEASGQLQCEHELEACQQVVDQQLR	7	C^a, C^b
$P_{RTh}3/13a-1$	AC	5, 7, 9, 1	C^b
$P_{RTh}3/13a-2$	ISPECHP	5	C^d

[a] Designation of cysteine residues according to Köhler et al (1993).

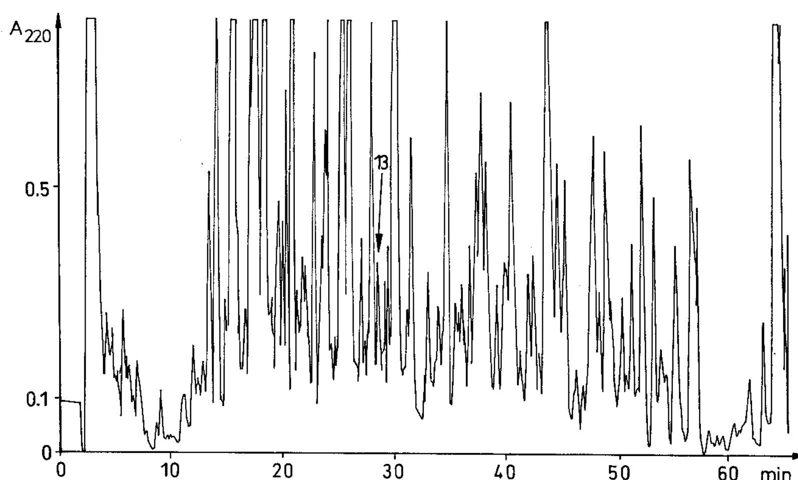

Fig. 1. Reversed-phase high-performance liquid chromatography of fraction 3 obtained from a thermolytic digest of enriched glutenin by gel-permeation chromatography (Keck et al 1995).

and was satisfactorily negative after heating, equilibration, and reminimization, indicating that the resulting structure was apparently without any important unfavorable interactions. Absolute values of the energy are not given because they are in part composition-dependent and we are comparing peptides with slightly different sequences. The RMS force is the more appropriate measure of a satisfactory state for the equilibrated molecular structure. The RMS force was <0.3 for all structures considered here. Energy minimization and molecular dynamics produced some changes in torsional angles and displacement of atoms, but the α-helix survived intact. In the resulting molecule, the cysteine residues at positions 10 and 25 (C^a and C^b) were well separated in the α-helical part of the molecule with a distance of 2.2 nm between the sulfur atoms (Figs. 2a and 3a).

Gluten proteins are synthesized by ribosomes attached to the endoplasmic reticulum with simultaneous insertion of the lengthening polypeptide chain through the membrane into the lumen (Galili et al 1993, Shewry 1995). It seems likely that secondary structures, such as α-helix, form rapidly as the polypeptide chain emerges from the membrane, making it unlikely that cysteines C^a and C^b (positions 10 and 25) approach one another sufficiently closely during biosynthesis to form an intramolecular disulfide bond. The relative position of the third cysteine C^d in the molecule (residue 40) was not well defined because we were unable to specify the torsional angles of some predicted loop structures surrounding this cysteine. Our modeling does not rule out the possibility that either the cysteine at position 40 (C^d) or the cysteine at position 96 (C^e; not included in this study) might fold back to form an intramolecular disulfide bond with any of the preceding cysteines in the polypeptide chain. The possibility of the four cysteines in the N-terminal domain of subunit 5 combining to form two intramolecular disulfide bonds seems unlikely. Subunit 5 must form two or four intermolecular disulfide bonds by way of its N-terminal domain. Otherwise, this subunit would not be conducive to formation of extensive polymers with only one cysteine at the C-terminal end of the protein molecule.

HMW-GS 2 is identical to subunit 5 in sequence through the first 124 residues, with the exception of a serine residue at position 96 instead of the cysteine C^e found in the sequence of subunit 5 at this position, thus having only three cysteine residues (C^a, C^b, and C^d at positions 10, 25, and 40, respectively) near the N-terminus of the protein. Accordingly, it may be speculated on the basis of our results for HMW-GS 5, that subunit 2 forms either one or three intermolecular disulfide bonds. In either case, it would be expected to be less effective in forming intermolecular disulfide crosslinks than subunit 5. Both proteins have only one cysteine near the C terminus (C^z) and would not be expected to differ in reactivity in the C-terminal domain. We note, however, that when a genetically engineered mutant gene of subunit 2 in which the codon for the cysteine of the C-terminal domain had been changed to that of serine, thus eliminating the C-terminal cysteine of the equivalent protein, was expressed in tobacco (Shani et al 1994), the resulting protein formed mainly polymers larger than dimers. This indicates, insofar as the expression in tobacco endosperm may be extrapolated to expression in wheat endosperm, that all three cysteines of the N-terminal domain form intermolecular disulfide bonds to a large extent (there was a small amount of the homodimer of the subunit 2 mutant present even in the absence of reducing agent).

HMW-GS 7. Version 11.93 of the PHD method predicted α-helical secondary structure for residues 6–13, residues 16–20, and residues 24–26 of HMW-GS 7 (50 residues considered). For a few residues, no prediction was made, and others were predicted to have extended structure (default structure in the modeling) or loop or turn structures. The α-helical conformation was applied to the appropriate residues in the computer-constructed model of subunit 7. In our first approach, loops or turns were not applied because of a lack of information about torsional angles for the amino acid residues involved. The resulting structure was subjected to simulated energy minimization, heating, and dynamic equilibration as was the model of the HMW-GS 5. In the resulting model, the sulfur atoms of the cysteine residues C^a and C^b (positions 10 and 17) had a distance of 0.95 nm after energy minimization. This was not as close as would be expected to favor the formation of an intramolecular disulfide bond between the two cysteines.

Structure prediction by a differently trained architecture of the PHD neural network method (Version 5.94_317) resulted in a prediction of α-helical structure for residues 8–27 in contrast to

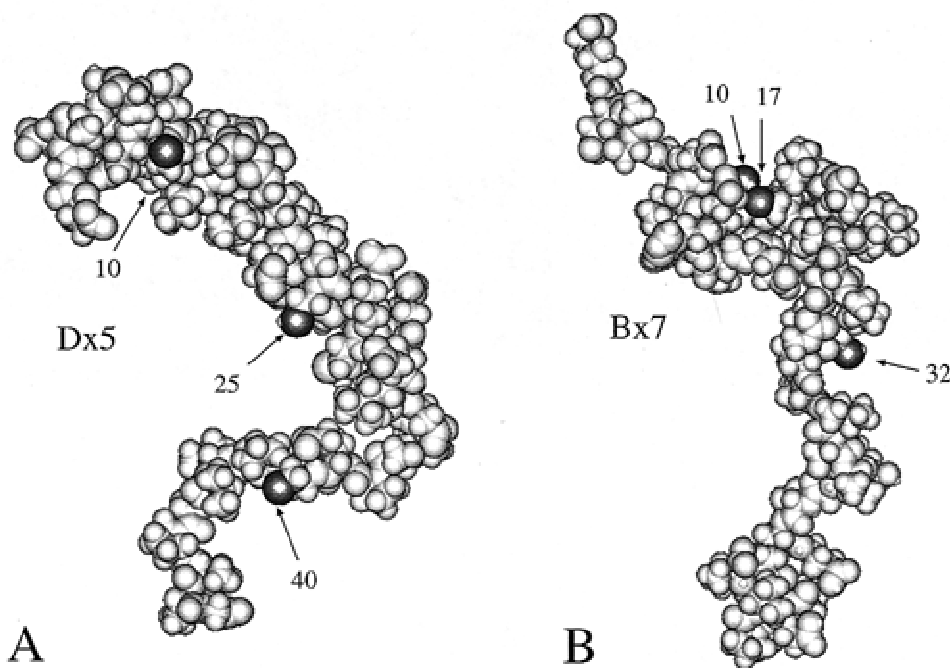

Fig. 2. Final models in space filling form (Van der Waals radii for atoms) of the first 50 amino acid residues of high molecular weight glutenin subunits 5 (**A**) and 7 (**B**). Sulfur atoms in darker shade. Numbers indicate cysteine positions.

the previous prediction where residues 6–13 and 16–20 inclusive, were predicted to be in the α-helical form. The differences apparently result mainly because secondary structure prediction is based on comparisons with homologous or similar sequences in proteins of known three-dimensional structure. A scarcity of homologous sequences in the structure databases makes predictions for gluten proteins much weaker than for many globular proteins. This weakness of homology combines with the nature of neural networks to give different results for the different versions or architectures (Rost and Sander 1991, 1993). However, no difference was noted in the prediction for subunit 5 in the range between cysteines 10 and 25 when predictions by the two architectures were compared.

Because our molecular modeling indicated that it was not possible to form the disulfide bond in question without disrupting a continuous α-helical conformation including the two cysteine residues in question and there is good evidence that these two residues are linked by an intramolecular disulfide bond (Köhler et al 1993), we used the structure prediction that indicated a break in the α-helix between cysteines 10 and 17 as the basis for further molecular modeling. A change in the distance between the sulfur atoms of the cysteines C^a and C^b could easily be achieved by changing the dihedral angles ϕ and ψ for the α-carbon atoms of the amino acid residues 14–16, where no prediction for helix has been made. For example, simple rotation about ϕ for the residue 14 α-carbon of the first structure (again followed by energy minimization, which indicated that the rotations had produced no unfavorable interactions) brought the sulfur atoms to within 0.52 nm of one another, a distance that makes the formation of an intramolecular disulfide bond between C^a and C^b likely. Our third approach involved amino acid residues 14, 15, and 16. No conformation was predicted for these residues in the PHD prediction. In a subset prediction, however, the method predicted that residue 15 was in a loop conformation, whereas no prediction was made for residues 14 and 16. Because only three residues were involved, we chose to apply an inverse γ-turn, which is common in proteins (Milner-White 1990), to residues 14–16 as a likely possibility by assigning dihedral angles $\phi = -75°$ and $\psi = 59°$ to residue 15. The γ-turn is a three-residue turn found in proteins in which residues i and i+2 are hydrogen bonded (Milner-White1990, Kasarda et al 1994) and would not disturb the predicted adjacent helical regions as would a four-residue turn. The resulting structure was processed as for the previous models. Again, the energy rapidly became negative during minimization and remained satisfactorily negative throughout the dynamics calculations and the RMS force was small. Although some changes in torsional angles and some atom displacement occurred during minimization and dynamics calculations, the overall structure was similar to that of the starting structure to which only α-helical structure and a single γ-turn had been applied.

The final structure for the HMW-GS 7 is shown in Figures 2B and 3B, where cysteines C^a and C^b (positions 10 and 17) are located in short lengths of α-helix that are bent back on one another by the short loop centered at residue 15. Although in the final structure the applied γ-turn no longer had a hydrogen bond linking residues 14 and 16, the backbone dihedral angles of residue 15 differed only slightly from those applied. The two sections of helix interacted through one backbone-to-backbone hydrogen bond that linked residues 10 and 17. The sulfur atoms of cysteines 10 and 17 approached one another relatively closely (0.43 nm) in the final model of subunit 7, in contrast to those of cysteines 10 and 25 of subunit 5 (2.21 nm). It would be expected that this close approach would favor an intramolecular disulfide bond between these two residues in subunit 7. This is in accordance with the experimental results of Köhler et al (1993). The position of the third cysteine in the sequence, C^d (residue 32), relative to the first two cysteines, was not sufficiently well defined to indicate whether or not it would be likely to interact with either of the first two cysteines, but the experimental evidence does not support such an interaction. Either way, our modeling suggests that the N-terminal region of the HMW-GS 7 is most likely to form a single intramolecular disulfide bond and one intermolecular disulfide bond. Thus, subunit 7 (one intermolecular disulfide bond) is likely to be equivalent to or less effective than subunit 2 (one or three intermolecular disulfide bonds) and definitely less effective than subunit 5 (two or four intermolecular disulfide bonds) in formation of glutenin polymers.

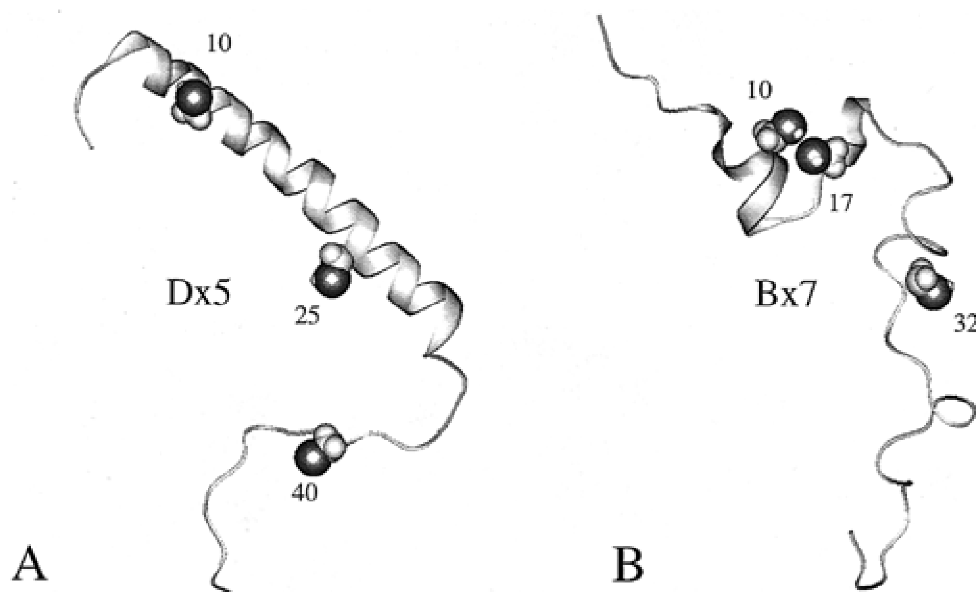

Fig. 3. Final models as in Fig. 2 in protein cartoon form (polypeptide backbone only). Cysteine side chains added with atoms of appropriate Van der Waals radii for high molecular weight glutenin subunits 5 (**A**) and 7 (**B**). Sulfur atoms in darker shade.

CONCLUSION

Molecular modeling based on secondary structure predictions in the absence of any known constraints on bond angles and atom distances that might be obtained from X-ray diffraction of crystals or from nuclear magnetic resonance spectroscopy of the peptides in solution must be considered highly speculative. There is heavy dependence in our modeling on the reliability of predictions for α-helical structure based on amino acid sequences, which have an accuracy >70% for the method of Rost and Sander (1993) for most globular proteins, but much less (to an unknown extent) for gluten proteins, which have few similarities in sequence to proteins of known three-dimensional structure.

We think, however, that our results are in agreement with the experimental evidence available so far and are intuitively satisfying. Further experimental evidence concerning the arrangement of disulfide bonds in glutenin polymers is needed, but our models should provide some guidance in the design of new experiments aimed at elucidating the relationship between conformational structure and disulfide bond formation. One experiment might be the chemical synthesis of a peptide corresponding to residues 5–25 of HMW-GS 7. Nuclear magnetic resonance measurements in solution or X-ray diffraction studies after crystallization of the peptide could be used to determine whether there is an inherent tendency to assume a conformation that favors close approach of the sulfur atoms of the cysteine side chains corresponding to the first two cysteines of the protein. Future work is required also to determine whether the linkage of the C^d (of subunit 5) and C^b cysteines is intramolecular or intermolecular.

ACKNOWLEDGMENTS

Travel support for P. Köhler provided by an exchange grant from the German government (No. 115-0870/USA) is gratefully acknowledged.

LITERATURE CITED

ANDERSON, O. D., GREENE, F. C., YIP, R. E., HALFORD, N., SHEWRY, P. R., and MALPICA-ROMERO, J.-M. 1989. Nucleotide sequences of the two high-molecular-weight glutenin genes from the D-genome of a hexaploid bread wheat, *Triticum aestivum* L. cv. Cheyenne. Nucleic Acids Res. 17:461-462.

ANDERSON, O. D., and GREENE, F. C. 1989. The characterization and comparative analysis of high molecular weight glutenin genes from genomes A and B of a hexaploid wheat. Theor. Appl. Genet. 77:689-700.

BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., and KARPLUS, M. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations. J. Computational Chem. 4:187-217

GALILI, G., ALTSCHULER, Y., and LEVANONY, H. 1993. Assembly and transport of seed storage proteins. Trends Cell. Biol. 3:437-443.

KASARDA, D. D., KING, G., and KUMOSINSKI, T. F. 1994. Comparison of spiral structures in wheat high molecular weight glutenin subunits and elastin by molecular modeling. Pages 209-220 in: Molecular Modeling: From Virtual Tools to Reality. T. F. Kumosinski and M. N. Liebman, eds. ACS Symp. Ser. No. 576, Am. Chem. Soc.: Washington, DC.

KECK, B., KÖHLER, P., and WIESER, H. 1995. Disulphide bonds in wheat gluten: Cystine peptides derived from gluten proteins following peptic and thermolytic digestion. Z. Lebensm. Forsch. 200:432-439.

KÖHLER, P., BELITZ, H.-D., and WIESER, H. 1991. Disulphide bonds in wheat gluten: Isolation of a cystine peptide from glutenin. Z. Lebensm. Forsch. 192:234-239.

KÖHLER, P., BELITZ, H.-D., and WIESER, H. 1993. Disulphide bonds in wheat gluten: further cystine peptides from high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin and from γ–gliadins. Z. Lebensm. Forsch. 196:239-247.

KUMOSINSKI, T. F., BROWN, E. M., and FARRELL, H. M., JR. 1993. Three-dimensional molecular modeling of bovine caseins: An energy-minimized β-casein structure. J. Dairy Sci. 76:931-945.

MILNER-WHITE, E. J. 1990. Situations of gamma-turns in proteins: their relation to alpha-helices, beta-sheets, and ligand binding sites. J. Mol. Biol. 216:385-397.

ROST, B., and SANDER, C. 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 232:584-599.

ROST, B., and SANDER, C. 1994. Combining evolutionary information and neural networks to predict secondary structure. CABIOS 10:53-60.

ROST, B., SANDER, C., and SCHNEIDER, R. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19:55-72.

SHANI, N., ROSENBERG, N., KASARDA, D. D., and GALILI, G. 1994. Mechanisms of assembly of wheat high molecular weight glutenins inferred from the expression of wild-type and mutant subunits in transgenic tobacco. J. Biol. Chem. 269:8924-8930.

SHEWRY, P. R. 1995. Plant storage proteins. Biol. Rev. 70:375-426.

[Received March 18, 1996. Accepted December 13, 1996.]