Homology modeling and molecular dynamics simulations of the N-terminal domain of wheat high molecular weight glutenin subunit 10

ROLAND CAZALIS,¹ THIERRY AUSSENAC,² LARBI RHAZI,¹ ANTOINE MARIN,³
AND JEAN-FRANÇOIS GIBRAT³

¹Laboratoire d'Agrophysiologie, UMR 1054 INRA, ESA Purpan, 31076 Toulouse cedex 3, France

²ISA Beauvais, 60026 Beauvais cedex, France

³Mathématique, Informatique et Génome, Centre de Recherche de Versailles, INRA, 78026 Versailles cedex, France

(RECEIVED August 28, 2002; ACCEPTED October 17, 2002)

Abstract

High molecular weight glutenin subunits (HMW-GS) are of a particular interest because of their biomechanical properties, which are important in many food systems such as breadmaking. Using fold-recognition techniques, we identified a fold compatible with the N-terminal domain of HMW-GS Dy10. This fold corresponds to the one adopted by proteins belonging to the cereal inhibitor family. Starting from three known protein structures of this family as templates, we built three models for the N-terminal domain of HMW-GS Dy10. We analyzed these models, and we propose a number of hypotheses regarding the N-terminal domain properties that can be tested experimentally. In particular, we discuss two possible ways of interaction between the N-terminal domains of the γ -type HMW glutenin subunits. The first way consists in the creation of interchain disulfide bridges. According to our models, we propose two plausible scenarios: (1) the existence of an intrachain disulfide bridge between cysteines 22 and 44, leaving the three other cysteines free of engaging in intermolecular bonds; and (2) the creation of two intrachain disulfide bridges (involving cysteines 22–44 and cysteines 10–55), leaving a single cysteine (45) for creating an intermolecular disulfide bridge. We discuss these scenarios in relation to contradictory experimental results. The second way, although less likely, is nevertheless worth considering. There might exist a possibility for the N-terminal domain of Dy10, Nt-Dy10, to create oligomers, because homologous cereal inhibitor proteins are known to exist as monomers, homodimers, and heterooligomers. We also discuss, in relation to the function of the cereal inhibitor proteins, the possibility that this N-terminal domain has retained similar inhibitory functions.

Keywords: Fold recognition; cereal inhibitor family; glutenin polymer; disulfide bridge pattern

Wheat gluten proteins are critical for bread-making and other food uses as they confer unique rheological properties on doughs, a combination of elasticity and viscous flow. Wheat gluten is a complex mixture of proteins broadly divided into two groups, the monomeric gliadins and the glu-

tenins, which form polymers linked by interchain disulfide bonds. Both groups contribute to the rheological properties of gluten and dough. However, a glutenin subgroup, the high molecular weight glutenin subunits (HMW-GS), is considered the most important determinants of these polymers (Payne 1987).

Early studies showed that the differences in the HMW-GS composition of bread wheat cultivars were associated with variation in breadmaking quality (Payne et al. 1979), whereas more recent studies have shown direct effects of HMW-GS composition on the proportion of high molecular

Reprint requests to: Jean-François Gibrat, Mathématique, Informatique et Génome, Centre de Recherche de Versailles, INRA, Route de St Cyr, 78026 Versailles cedex, France; e-mail: gibrat@versailles.inra.fr; fax: +33 (1) 30 83 33 59.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0229803>.

weight glutenin polymers (Popineau et al. 1994; Gupta et al. 1995) and the level of gluten viscosity (Popineau et al. 1994). The structures and interactions of the HMW-GS are, therefore, of considerable interest in relation to understanding and manipulating gluten functionality.

HMW-GS proteins consist of two closely related subunit types, called x and y on the basis of their mobility in SDS-PAGE (Payne 1987). The x and y types share a similar primary structure, which consists of three domains: (1) a small N-terminal domain ranging from 80 to 120 amino acids with three to five cysteine residues (Shewry and Tatham 1997); (2) a large central domain made up of repeating sequences rich in glutamines, prolines, and glycines; (3) the C-terminus, a nonrepetitive domain of ~50 residues with one cysteine. The nonrepetitive domains, then, contain most or all of the cysteine residues, some of which form interchain disulfide bonds (Gupta et al. 1995) to stabilize the high molecular glutenin polymers.

However, x- and y-type subunits differ in some features. Sequence alignments show that the y-type subunit exhibits an insertion of 18 residues in the middle of the N-terminal domain, including two cysteine residues, which is not found in the x-type subunit. Therefore, y types have five cysteines in this domain, whereas x types usually have three. y types have an additional cysteine in the repetitive domain that is also characterized by differences in composition and order of the repeat motifs between the two types (Anderson and Greene 1989; Miles et al. 1991; Matsushima et al. 1992).

Physical studies (Miles et al. 1991; Matsushima et al. 1992; Thomson et al. 1999; Humphris et al. 2000) and molecular modeling (Matsushima et al. 1990; Kasarda et al. 1994; Parchment et al. 2001) of the central repeat domain indicate a β -turn conformation organized in a β -spiral structure that might be responsible for the elastic properties of the molecule.

It is known that cross-linking of the various glutenin subunits by intermolecular disulfide bridges plays a critical role in the formation of high molecular glutenin polymers. However, the experimental results, as to which cysteines of the different glutenin subunits are involved in these interchain bridges, are, so far, contradictory (Köhler et al. 1991, 1993; Keck et al. 1995; Shimoni et al. 1997; Kasarda 1999).

Recently the first 50 residues of both HMW-GS Dx5 and Bx7 were modeled by computer methods (Köhler et al. 1997). The resulting structure shows that the first two cysteines are located in a region of continuous α -helix in HMW-GS Dx5, whereas HMW-GS Bx7 shows an α -helix structure interrupted between the first two cysteines by an inverse γ -turn. This break in the helix permits a possible intramolecular disulfide bond, in agreement with some of the experimental results (Köhler et al. 1993).

In the present work, we have investigated the structure of the N-terminal domain of HMW-GS Dy10 (Cheyenne cultivar, GenBank accession number X12929; Anderson et al.

1989), using fold-recognition and molecular modeling techniques. Note that, hereafter, this N-terminal domain will be referred to as Nt-Dy10. Nt-Dy10 is defined as encompassing the first 119 amino acids, excluding the signal peptide (the first 21 residues in the GenBank file). An analysis of the resulting models allows us to propose a number of hypotheses regarding the properties of Nt-Dy10 that can be tested experimentally and to discuss the cysteines of the HMW-GS Dy10 involved in intra- and interchain disulfide bridges.

Results

Sequence analysis

To obtain a more accurate partition of the sequence in terms of domains, we used the program SEG (Wootton and Federhen 1993), which is designed to segment amino acid sequences into subsequences of contrasting complexity. Figure 1 shows the result of this analysis for the sequence of HMW-GS Dy10. It confirms well what is known experimentally about the structure of the protein (Shewry et al. 1992). HMW-GS Dy10 is made up of three domains, an N-terminal globular domain having 119 residues, a small

Low complexity fragments		"regular" fragments
	1-119	EGEASRQLQCERELQESSLEACRQVVDQQL AGRLPWSTGLQMRCCQLRDVSAKCRSVAV SQVARQYEQTVVPPKGGSFYPGETTPLQQL QQGIFWGTSSQTVQGYYPGVTSRQGSYY
pgqasppqpgqgqgqgkwqepqgqgqg	120-145	
	146-152	WYPTSL
qapqgqgqgqgkqgqgqg	153-167	
	168-173	YYPTSL
qapqgqgqgqgqgqg	174-185	
	186-188	TSL
qhtgqrqgqgqgqgqgqgqgqgqgqgqg	189-215	
	216-221	YYPTSP
qg1gqgqgqgqgqgqgqgqgqgqgqg	222-242	
	243-262	HYPTSLQQPQCGQGHYLAS
qgqgqgqgqgqgqgqgqgqgqgqgqg	263-272	
	273-277	HYPAS
qgqgqgqgqgqgqgqgqgqgqgqgqg	278-287	
	288-292	HYPAS
qgqgqgqgqgqgqgqgqgqgqgqgqg	293-302	
	303-304	HY
pasqgqgqgqgqgqgqgqgqgqgqgqg	305-332	
	333-338	HYPASL
qgqgqgqgqgqgqgqgqgqgqgqgqg	339-347	
	348-352	HYPTS
1qg1gqgqgqgqgqgqgqgqgqgqgqg	353-395	
peqeqgqgqgqgqgqgqgqgqgqgqg		
	396-401	YYPTSL
qgqgqgqgqgqgqgqgqgqgqgqgqg	402-416	
	417-422	YYPTSL
qgqgqgqgqgqgqgqgqgqgqgqgqg	423-431	
	432-437	HYPASL
qgqgqgqgqgqgqgqgqgqgqgqgqg	438-511	
qgqgqgqgqgqgqgqgqgqgqgqgqg		
	512-516	HCPTS
pqqgqgqgqgqgqgqgqgqgqgqgqg	517-544	
	545-550	YYPTSV
qgqgqgqgqgqgqgqgqgqgqgqgqg	551-583	
qgqgqgqgqgqgqgqgqgqgqgqgqg		
	584-627	YDSPYHVSAREQQAASPMVAKAQQPATQLPT VCRMGGDALSASQ

Figure 1. SEG analysis of the HMW-GS Dy10 sequence. The left column shows low complexity regions, the right column regular regions, that is, regions that are likely to correspond to globular structures. The middle column is the sequence numbering.

C-terminal globular domain having ~50 residues, and a large, low-complexity domain connecting the two globular domains.

The presence of leucine and glutamate residues at the beginning of the sequence is reminiscent of a coiled-coil domain. There are several types of coiled-coil domains, for instance, the leucine zipper domain in transcriptional regulators or the coiled-coil domain observed in the hinge region of myosin. The leucine zipper motif allows the dimerization of DNA-binding proteins that are, except for this domain, regular globular proteins. Coiled-coil domains are often involved in protein oligomerization; for example, it is believed that they are involved in flagelin polymerization. To test for the existence of a coiled-coil domain, we used the program COIL (Lupas et al. 1991), which delineates coiled-coil domains in otherwise globular proteins. No coiled-coil domain was found in the sequence of HMW-GS Dy10.

We performed a secondary structure prediction of Nt-Dy10 using the program PSIPRED (McGuffin et al. 2000). On average, this program provides a percentage of correctly predicted residues of 75%, when three states are taken into consideration: α -helix, β -strand, and coils. Figure 2 shows the result of the secondary structure prediction. The structure of Nt-Dy10 appears to be mostly helical: There are three N-terminal helices and two C-terminal helices. A short β -strand is predicted in between the N-terminal and C-terminal helices.

Homology search in the databases

The next stage of our analysis was to search in databases for proteins that were homologous to Nt-Dy10. We started our analysis with sequence comparison techniques, using increasingly sensitive methods.

Thus, we first used BLAST (Altschul et al. 1990), then PSI-BLAST (Altschul et al. 1997), and finally HHMER, a program based on Hidden Markov Model techniques (Hughey and Krogh 1996) that performs a search against the PFAM database of domains (Bateman et al. 2002).

It has been shown (Brenner et al. 1998) that programs such as BLAST are not very effective for finding remote homologs whose amino acid sequences can have greatly diverged. More effective methods, instead of using a single sequence, make use of multiple aligned sequences (or profiles). Profiles contain more information than single sequences and therefore allow a more sensitive search. PSI-BLAST is a method based on BLAST that builds a profile iteratively. This program, as well as Hidden Markov Model techniques, has been shown to be three times more sensitive than BLAST (Park et al. 1998).

Using these three methods, only glutenins and D-hordeins were found (data not shown). We therefore turned toward fold-recognition methods. It is well known that the 3D structure of homologous proteins is better conserved

than the amino acid sequence: remote homologs can have as few as 10%–15% conserved residues and yet still have very similar 3D structures and often similar functions. The 3D structure is thus a better means of characterizing remotely homologous proteins than the sequence.

We have developed a fold-recognition method, called FROST (Marin et al. 2002a,b), and have shown that this method is twice as sensitive as PSI-BLAST (Marin et al. 2002a).

Table 1 shows the results of the search for the sequence of Nt-Dy10 against a database that contains representatives of all the known 3D structures. FROST consists of two filters called the 1D and 3D filters. The Core column corresponds to the code of known protein 3D structures of the Protein Databank (PDB). The Definition column summarizes the function of the corresponding protein, and the Id column indicates the percentage of conserved residues in the query and core sequences. The Sini column gives the alignment score. The most interesting column for our purpose is the column Ndist (normalized distance). Without going into details, let us say that this distance quantifies the degree of confidence we have in the fold recognition. We have shown that, for a normalized distance >3.2 , there is a $<1\%$ chance to wrongly assign a fold to a sequence. In Table 1, three proteins with the same fold appear above this threshold: the bifunctional corn Hageman factor inhibitor: PDB code 1BEA (Behnke et al. 1998); the bifunctional α -amylase/trypsin inhibitor from ragi seed: PDB codes 1BIP (Strobl et al. 1995) and 1TMQ (Strobl et al. 1998); and the 0.19 α -amylase inhibitor from wheat kernel: PDB code 1HSS (Oda et al. 1997). 1BIP and 1TMQ correspond to the same protein, whose 3D structure has been solved by X-ray crystallography, file 1TMQ, and by NMR spectroscopy, file 1BIP. In fact, the structure in file 1TMQ describes the complex yellow meal worm α -amylase and bifunctional α -amylase/trypsin inhibitor from ragi seed. Note that the results of both filters concord.

All these proteins belong to the cereal inhibitor family. In Release 40.7 of SWISS-PROT, we found 19 known members of this family (omitting short fragments). These pro-

Figure 2. Result of the PSIPRED secondary structure prediction of Nt-Dy10.

Table 1. FROST results for Nt-Dy10

ID filter					
Core	Ndist	Sini	Id	Definition	Rank
1TMQB0	5.3	-21.3	17.8%	Bifunctional α -amylase/trypsin inhib	1
1BIP_0	4.9	-81.9	16.5%	Bifunctional α -amylase/trypsin inhib	2
1BEA_0	4.8	-36.6	20.0%	Bifunctional corn Hageman factor inhibit	3
1HSSA0	3.7	-12.0	15.3%	0.19 α -amylase inhibitor	4
1MEK_0	2.1	-30.9	13.3%	Human protein disulfide isomerase, elect	5
2BBKL0	2.0	-26.8	12.7%	Methylamine dehydrogenase	6
1FRD_0	2.0	8.7	8.8%	Heterocyst [2fe-2s] ferredoxin	7
2RHE_0	2.0	-8.6	9.8%	Bence-Jones protein	8
2KAUB0	1.9	-13.5	11.4%	<i>Klebsiella aerogenes</i> urease	9
1BMP_0	1.9	-24.9	13.7%	Bone morphogenetic protein-7 morphogen	10
3D filter					
1BEA_0	4.2	-139.1	18.2%	Bifunctional corn Hageman factor inhibit	1
1BIP_0	4.0	-122.9	15.7%	Bifunctional α -amylase/trypsin inhib	2
1TMQB0	3.9	-105.4	17.2%	Bifunctional α -amylase/trypsin inhib	3
2BBKL0	1.1	-38.8	9.8%	Methylamine dehydrogenase	4
1BMP_0	1.0	0.9	10.8%	Bone morphogenetic protein-7 morphogen	5
2KAUB0	0.8	-73.0	9.7%	<i>Klebsiella aerogenes</i> urease	6
1FDR_0	0.8	-13.7	6.2%	Heterocyst [2fe-2s] ferredoxin	7
1HSSA0	0.6	22.7	12.1%	0.19 α -amylase inhibitor	8
2RHE_0	0.5	21.3	11.9%	Bence-Jones protein	9
1MEK_0	0.0	35.6	7.9%	Human protein disulfide isomerase, elect	10

teins were detected with FASTA (Pearson 2000) using the sequence of 1BIP (IAAT_ELECO) as a query. Table 2 shows the function of these proteins when known (Strobl et al. 1998).

Modeling of the 3D structure of Nt-Dy10

Figure 3A presents a structural alignment of proteins 1HSS (IAA1_WHEAT), 1BIP (IAAT_ELECO), and 1BEA (ITRF_MAIZE) performed using the VAST program (Gibrat et al. 1996). The core of the structures consists of four conserved α -helices with a simple up-and-down topology. Other parts of the structure differ markedly.

The loop between helix 1 and helix 2 corresponds, for 1BIP and 1BEA, to the canonical inhibition site for trypsin

serine proteases (Bode and Huber 1992). This loop protrudes into the active site of the corresponding serine proteases in a substrate-like fashion. The corresponding loop in 1HSS has a very different conformation, but apparently 1HSS inhibits only α -amylase. The loop between helices 3 and 4 is also very different in both types of structures.

Once homologous proteins whose 3D structure is known have been identified, it is possible to build a 3D model for Nt-Dy10 using these known structures as templates.

The first stage of homology modeling, that is, the query protein sequence alignment with the sequence of the template, is critical. If this alignment is wrong, the resulting model will be wrong. To obtain the best possible model, we make use of all the information at hand: cereal inhibitor family multiple sequence alignment (data not shown),

Table 2. Function of the members of the cereal protease inhibitor family

Bifunctional ^a	Trypsin inhibitor	α -Amylase inhibitor ^b	α -Amylase inhibitor ^c	Unknown ^d
IAAT_ELECO	ITRF_MAIZE	IAA4_SORBI	IAAD_HORVU	IAA_HORVU
		IAA5_SORBI	IA01_WHEAT	RA17_ORYSA
		IAAA_HORVU	IA02_WHEAT	RA14_ORYSA
		IAA1_WHEAT	IA03_WHEAT	RAG2_ORYSA
		IAA2_WHEAT	IA16_WHEAT	RA05_ORYSA
		IAA5_WHEAT	IAAB_HORVU	

^a α -Amylase/trypsin inhibitor.

^b α -Amylase inhibitor active as monomers or homodimers.

^c α -Amylase inhibitor active as heterooligomers.

^d Without known inhibitory function.

Table 3. Conserved residues that form the core of the 3D structures

1BIP	1HSS	Nt-Dy10
Pro 16	Ala 16	Ser 18
Leu 17	Leu 17	Leu 19
Val 24	Leu 24	Val 26
Leu 48	Leu 45	Leu 48
Ile 51	Ile 48	Val 51
Ala 59	Ala 56	Ala 59
Val 60	Leu 57	Val 60
Leu 63	Met 60	Val 63
Leu 97	Ile 95	Thr 98
Val 98	Thr 96	Ser 99
Arg 56	Arg 53	Arg 56
Gln 90	Val 88	Gln 91

Generally speaking, the alignment of the first half of the sequence, up to the end of the third helix, is relatively straightforward. The number of conserved residues, such as the abovementioned core residues, but also cysteines and arginines, is sufficient, we believe, to correctly anchor the alignment. The second half of the structure is more problematic, in particular the location of the fourth helix. There is no residue conserved in this helix, and thus its precise location is more difficult to assess. This helix, potentially, plays an important role. In the homologous structure of 1HSS, hydrophobic residues lie on the external face of this helix and are involved in the dimer interface. In addition, none of the cysteines characteristic of the family appears to be conserved in this part of the sequence.

In their paper about the modeling of the N-terminal regions of HMW-GS Dx5 and Bx7 (first 50 residues) based exclusively on secondary structure predictions, Köhler et al. (1997) pointed out that the models they built had to be considered highly speculative.

Here, in contrast, we emphasize that fold-recognition techniques are now well established and constitute increasingly reliable techniques (CASP4 2001).

Benchmark tests with our method have also revealed that above a normalized distance of 4, on average, 75% of the query residues are perfectly aligned (with a shift of 0 with respect to their position in the true structural alignment), and a remaining 20% are aligned with a shift of at most 4 (Marin et al. 2002a). Therefore, we believe that, although all atomic details of the models are unlikely to be right, general features of the 3D structure should be essentially correct.

Dimer interface

To check whether Nt-Dy10 is able to form a dimer in a way similar to its template 1HSS, we analyzed the accessible surface of the interface between the two monomers in the template protein and in the model Mod_1HSS.

For this purpose we used NACCESS (Hubbard et al. 1991) to calculate the accessible surface area of the dimer, on the one hand, and of the monomer, on the other hand. All residues showing a relative surface area decreasing by more than 10% between the monomer and the dimer were recorded. They are listed in Table 4.

The distribution of residues for 1HSS and Mod_1HSS is strikingly different. 1HSS exhibits a peculiar distribution of residues: seven are charged (Table 4, boldface), seven are hydrophobic (Table 4, italic), and seven are regular surface residues.

Two facts must be noticed. First, hydrophobic residues usually do not tend to cluster on the protein surface unless they are involved in a dimerization interface. Second, it is not favorable to bury charged residues unless they are able to create salt bridges. Figure 4 shows that this is precisely the case for 1HSS. The dimerization interface is made of three layers: a strip of regular surface residues sandwiched by hydrophobic residues (Val 83, Leu 86, Val 94, Val 101, Val 110), themselves surrounded by a crown of charged residues that all form salt bridges (Arg 96 of one chain with Glu 31 of the other one, Arg 80 with Asp 102 and Asp 106, Lys 112 with Glu 82).

As is apparent from Table 4, residues at the dimer interface for Mod_1HSS down to Val 103 are residues that are expected to be located at the protein surface for a regular globular protein. These data, therefore, seem to rule out the

Table 4. Dimerization interface

1HSS		Mod_1HSS	
Residue	%lost	Residue	%lost
Thr 87	11.0	Ala 31	10.0
Ala 115	11.3	<i>Leu 30</i>	10.5
Asn 25	11.8	Gly 105	12.7
Asp 102	13.0	Gln 29	13.3
<i>Val 29</i>	13.3	<i>Leu 87</i>	14.5
<i>Leu 34</i>	15.5	<i>Val 110</i>	15.8
Glu 31	18.4	<i>Trp 96</i>	17.5
Ala 93	23.3	Arg 33	20.0
Gln 23	23.5	<i>Trp 36</i>	20.5
<i>Val 101</i>	23.7	<i>Val 103</i>	20.4
Ser 90	25.3	Gly 116	26.8
Ala 108	26.6	Gln 92	27.2
Glu 82	26.6	Gln 89	27.4
Asp 106	27.8	Thr 98	27.5
<i>Val 83</i>	31.7	<i>Phe 95</i>	30.3
Lys 112	36.9	Ser 99	38.2
<i>Val 94</i>	37.4	Ser 117	39.6
Arg 80	44.2	Gln 115	41.5
Gly 26	47.3	Tyr 118	42.2
<i>Leu 86</i>	48.1	Gly 32	52.7
Tyr 109	50.6	Tyr 119	55.7
Gly 107	51.1	Thr 102	58.3
<i>Val 110</i>	52.0		
Arg 96	53.0

Figure 4. Space-filling representation of the dimer interface between chains A and B of the 1HSS structure. (Blue) Charged residues (Table 4); (red) hydrophobic residues; and (yellow) regular residues. Although hardly apparent in this view, the interface is rather flat.

formation of a dimer by Nt-Dy10, at least in a way similar to the 1HSS dimer.

Disulfide bridges

Proteins of the cereal inhibitor family contain, in general, 10 cysteines forming 5 disulfide bridges (sometimes a cysteine is missing, resulting in only 4 disulfide bridges).

Nt-Dy10 has 5 cysteines. According to the models, only one disulfide bridge can be created between Cys 22 and Cys 44. In the models, Cys 10 and Cys 55 are 15 Å apart for Mod_1BIP and Mod_1BEA and 10 Å for Mod_1HSS. Because the N terminus of the molecule is flexible, it might be possible for these two cysteines to get close and form a disulfide bridge.

The monitoring of the corresponding cysteine sulfur-atom distance along the molecular dynamics trajectories (data not shown) for Mod_1BIP and Mod_1BEA shows that this distance undergoes a sudden transition, down to 10 Å, after 250 psec for Mod_1BEA and after 280 psec for Mod_1BIP. The trajectory for Mod_1HSS is not long enough to observe a similar phenomenon.

Therefore, we cannot completely rule out the possibility of a disulfide bridge between Cys 10 and Cys 55. Two scenarios can be considered. In the first one there is only one intrachain disulfide bridge between Cys 22 and Cys 44.

Three cysteines—10, 45, and 55—remain available to make interchain disulfide bridges. In the second scenario there are 2 intrachain disulfide bridges connecting, respectively, Cys 22 to Cys 44, and Cys 10 to Cys 55. Only Cys 45 is free to create intermolecular disulfide bridges.

Discussion

Common viscoelastic properties with CM proteins

Some wheat proteins (SWISS-PROT IDs IA01_WHEAT, IA02_WHEAT, IA03_WHEAT, IA16_WHEAT) of the cereal inhibitor family, known as CM proteins (they were initially characterized by their solubility in a mixture of chloroform and methanol, hence their name), have been shown to be involved in the technological quality of wheat products like pasta (Gautier et al. 1989; Kobrehel and Alary 1989). Kobrehel and Alary (1989) found a significant correlation between the sulphydryl-disulfide content of a CM-rich fraction of gluten and the cooking quality of pasta. They underlined the “important functional role of [CM] proteins in the expression of the visco-elastic properties of gluten.” They proposed that during hydrothermic treatment, CM proteins can form disulfide bridges, thus creating a strong proteinaceous network that improves the cooking quality of pasta.

We think that it is not merely luck if CM proteins and the N-terminal domain of HMW glutenins share common viscoelastic properties, but can be readily explained by their underlying common fold, as we propose in this paper.

Possible modes of domain interactions

The principal function of models is to help us propose new experiments to test biological hypotheses. Here, regarding the interaction between HMW glutenin proteins that are responsible for mechanical properties of dough, we can consider two different possibilities of association (that are not necessarily exclusive).

Intra- and interchain disulfide bridges

The most obvious possibility, backed up by many experimental studies, is the creation of interchain disulfide bridges. However, different opinions prevail as to which cysteines are involved in intra- and interchain disulfide bridges.

According to Wieser and colleagues (Köhler et al. 1991, 1993; Keck et al. 1995), there exist two parallel intermolecular disulfide bridges between adjacent Cys 3 and Cys 4 with the same cysteines in another y-type subunit. Note that, in this section, for the sake of comparison with data of the literature, Nt-Dy10 cysteines are numbered as they appear in the sequence, that is, from 1 to 5. They also proposed,

tentatively, the existence of an interchain disulfide bond between Cys 2 of a y-type HMW-GS and Cys 5 of an x-type HMW-GS, although for this latter disulfide bridge two other possible types of connection exist.

Kasarda and colleagues (Tao et al. 1992) and Galili and colleagues (Shimoni et al. 1997) found only one interchain disulfide bridge connecting a y-type to an x-type HMW-GS. Note that these two teams used different experimental techniques to reach this conclusion. These experiments did not allow them to identify which cysteine of the y type was involved in the interchain disulfide bridge. Kasarda (1999), summarizing the experimental results available so far, proposed that the y-type HMW-GS contains two intrachain and one interchain disulfide bridges. He also acknowledged the inconsistency of his proposal with the results of the Wieser group.

The second scenario we proposed above, namely, that Cys 1–Cys 5 and Cys 2–Cys 3 form intrachain disulfide bonds, leaving Cys 4 free to create an interchain disulfide bridge, is fully compatible with the proposal of Kasarda.

However, this scenario is built on the assumption that Cys 1 and Cys 5 are able to create an intrachain disulfide bond. From the data obtained during the simulations this is possible but not yet proven.

The first scenario suggests that only Cys 2 and Cys 3 can create an intrachain disulfide bridge, leaving three cysteines free from making interchain bridges. This scenario is not compatible with data obtained by Köhler et al. (1991), because the intrachain bond involves two cysteines that are found by them to be engaged in interchain disulfide bridges. Let us note that the involvement of Cys 2 in an interchain bond with Cys 5 of an x-type HMW-GS is only one of three, equally likely, possibilities. The question, then, arises, regarding the work that shows the existence of the two parallel disulfide bridges (Köhler et al. 1991), of the possibility of an experimental artifact. In this work, a peptide made of two identical fragments, CCQQL (residues 44–48 of HMW-GS Dy10), linked by the two abovementioned parallel disulfide bonds was isolated. One might well imagine that, if the intrachain bond is broken in both subunits during the tryptic digestion, a new, artifactual, disulfide bridge could be made between the two newly freed Cys 3s (Cys 44s in the sequence) that are maintained in a close vicinity in the peptidic fragment created by the experiment.

The models provide ambiguous data regarding intra- and interchain disulfide bridges. It appears, however, that these data are easier to reconcile with the proposal of the Kasarda and Galili groups than with Wieser and coworkers' suggestions.

Domain dimerization

The second possibility concerns the direct association of chains to form oligomers. Proteins of the cereal inhibitor

family are known to exist as monomers, homodimers, or heterooligomers. Our analysis of the dimer interface in Mod_IHSS seems to rule out this possibility. We have tried to find other possible oligomerization interfaces, but we could not come up with anything significant (data not shown). We must emphasize, however, that it is, in general, rather difficult to locate such an interface without a priori clues, and we might well have overlooked it. It is therefore interesting, we think, to study experimentally whether HMW glutenin proteins are capable of association (mediated by the N-terminal domain) in the absence of disulfide bridges, under reducing conditions.

Potential function of the N-terminal domain

As indicated by Strobl et al. (1998) and summarized in Table 2, proteins of the cereal inhibitor family inhibit two important digestive enzymes of animals (trypsin and α -amylase) and are thought to be involved, *in vivo*, in a plant defense mechanism against predatory insects.

Proteins of the cereal inhibitor family and the N-terminal domain of HMW-GS Dy10 are remote homologs. This implies that the present function of the N-terminal domain of HMW-GS Dy10 and the function of proteins of the cereal family are more likely to have diverged over time. In spite of these questions, we, nevertheless, think that it would be interesting to test experimentally whether this domain is capable of inhibiting trypsin and/or α -amylase, at least *in vitro*.

An analysis of the models and related alignments leads us to believe that it is unlikely that Nt-Dy10 exhibits an α -amylase or trypsin inhibitory function. However, some engineering of the sequence could restore these functions as described below.

The α -amylase inhibitory site of 1BIP has been shown to involve principally two segments, Ser 1–Ala 11 and Pro 52–Cys 55 (Strobl et al. 1998). The first residue, Ser 1, plays a critical role because it is capable of interacting with all three acidic residues of the α -amylase active site. It is known that N-terminal elongation of wheat α -amylase inhibitor (IAA2_WHEAT) diminishes or abolishes inhibition of α -amylase (Garcia et al. 1991). This is confirmed by the fact that most proteins, IAA_HORVU, RA17_ORYSA, RA14_ORYSA, RAG2_ORYSA, RA05_ORYSA, which have no known inhibitory function according to Table 2, exhibit an N-terminal elongation. Nt-Dy10 also has such an N-terminal elongation, and one might be able to recover a detectable α -amylase inhibitory effect only by trimming the first 4 residues of Nt-Dy10.

The trypsin-binding loop of 1BIP is located between helices 1 and 2 and corresponds to the sequence: Gly 32–Pro 33–Arg 34–Leu 35–Ala 36–Thr 37. It adopts the canonical conformation observed in other serine protease inhibitors (Bode and Huber 1992). In Nt-Dy10, on the other hand, the

loop (Gly 32-Arg 33-Leu 34-Pro 35-Trp 36-Ser 37) is shorter and, in Mod_1BIP at least, adopts a different conformation. To test whether Nt-Dy10 is able to inhibit trypsin, it might be necessary to insert a proline in the loop between Gly 32 and Arg 33 and possibly replace Pro 35 and Trp 36 by the corresponding residues (Ala and Thr) in the 1BIP loop.

Of course, these engineering experiments would not provide a clue to the present function of Nt-Dy10 but would confirm experimentally that the overall 3D organization of this domain is correct.

Conclusion

Using a fold-recognition technique, FROST, we identified a fold compatible with the N-terminal domain of HMW-GS Dy10. This fold corresponds to the one adopted by proteins of the cereal inhibitor family. Starting from the known structure of three members of the family: the bifunctional α -amylase/trypsin inhibitor of ragi seed, the bifunctional corn Hageman factor inhibitor, and the 0.19 α -amylase inhibitor, we built three models of this N-terminal domain. We analyzed these models and discussed their properties in relation to the possible modes of interaction of y-type HMW-GSs in glutenin polymers.

We believe that our models can be a help in the design of new experiments aimed at the elucidation of the relationship between the protein structure and its (biomechanical) properties.

The three models of the N-terminal domain of HMW-GS Dy10 can be downloaded from our Website, <http://www-mig.jouy.inra.fr/mig/downloads.html>.

Materials and methods

Sequence analysis was performed using SEG (Wootton and Federhen 1993) for detecting segments of low complexity in the HMW-GS γ 10 sequence, COILS (Lupas et al. 1991) for searching for coiled-coils, and PSI-PRED (McGuffin et al. 2000) for predicting the secondary structure of the N-terminal domain. These three methods were used with default parameters.

Homology searches were carried out with BLAST (Altschul et al. 1990) and PSI-BLAST (Altschul et al. 1997) against the NCBI nonredundant database. Default parameters were used for BLAST and PSI-BLAST except that, for the latter, the threshold for including homologous proteins in the profile was set to 0.0005 and the number of iterations (runs) was set to 20, as recommended by Park et al. (1998). Domain search was performed with HHMER (Hughey and Krogh 1996) against the PFAM database of domains (Bateman et al. 2002) with default parameters.

Fold recognition was performed with FROST (Marin et al. 2002a,b) with a database containing ~1200 cores.

Structural alignments between 1BIP, 1BEA, and 1HSS were performed with VAST (Gibrat et al. 1996).

Models were built with MODELLER (Sali and Blundell 1993) starting from the alignments shown in Figure 3B. For each model,

10 structures were generated, and the one with the best objective function was kept.

Molecular dynamics simulations were carried out with AMBER6 (Cornell et al. 1995). Each model was solvated in a box of equilibrated water molecules, the size of the box being calculated such that there are at least 10 Å between any protein atom and the box edges. The whole system was then minimized. Starting from the minimized structure, a molecular dynamics simulation was run for 300 psec (200 psec for Mod_1HSS) using periodic boundary conditions at constant temperature ($T = 298$ K) and constant volume. We used a time step of 1 fsec. Conformations along the trajectory were saved every 0.1 psec.

Calculation of the accessible surface area for the 1HSS dimer was performed with NACCESS (Hubbard et al. 1991) using standard parameters.

Acknowledgments

We thank M.-F. Gautier for helping us to trace back the cryptic comment in SWISS-PROT about the role of CM proteins in the technological quality of wheat products, to the relevant bibliography.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Anderson, O. and Greene, F. 1989. The characterization and comparative analysis of high molecular weight glutenin genes from genomes A and B of a hexaploid bread wheat. *Theor. Appl. Genet.* **77**: 689–700.
- Anderson, O., Greene, F., Yip, R., Halford, N., Shewry, P., and Malpica, R. 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.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Ewinger, L., Eddy, S., Griffiths, J., Howe, K., Marshall, M., and Sonnhammer, E. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30**: 276–280.
- Behnke, C., Yee, V., Trong, I., Pedersen, L., Stenkamp, R., Kim, S., Reeck, G., and Teller, D. 1998. Structural determinants of the bifunctional corn Hageman factor inhibitor: X-Ray crystal structure at 1.95 Å resolution. *Biochemistry* **37**: 15277–15288.
- Bode, W. and Huber, R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**: 433–451.
- Brenner, S., Chothia, C., and Hubbard, T. 1998. Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc. Natl. Acad. Sci.* **95**: 6073–6078.
- CASP4: Critical Assessment of Techniques for Protein Structure Prediction. 2001. Progress report on 4th meeting. *Proteins Suppl.* **5**; <http://predictioncenter.llnl.gov>.
- Cornell, W., Cieplak, P., Bayly, C., Gould, I., Merz, J., Ferguson, D., Spellmeyer, D., Fox, T., Caldwell, J., and Kollman, P. 1995. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**: 5179–5197.
- Garcia, M., Carbonero, P., and Garcia, O. 1991. Site-directed mutagenesis and expression in *Escherichia coli* of WMAI-1, a wheat monomeric inhibitor of insect α -amylase. *Plant Mol. Biol.* **17**: 1005–1011.
- Gautier, M., Alary, R., Kobrehel, K., and Joudrier, P. 1989. Chloroform/methanol soluble proteins are the main components of *Triticum durum* sulfur-rich glutenin fractions. *Cereal Chem.* **66**: 535.
- Gibrat, J.-F., Madej, T., and Bryant, S.H. 1996. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6**: 377–385.
- Gupta, R., Popineau, Y., Lefebvre, J., Corne, M., Lawrence, G., and

- MacRitchie, F. 1995. Biochemical basis of flour properties in bread wheats. II. Changes in polymeric protein formation and dough gluten properties associated with the loss of low molecular or high molecular glutenin subunits. *J. Cereal Sci.* **21**: 103–116.
- Hubbard, S., Campbell, S., and Thornton, J. 1991. Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.* **220**: 507–530.
- Hughey, R. and Krogh, A. 1996. Hidden Markov models for sequence analysis: Extension and analysis of the basic method. *Comput. Appl. Biosci.* **12**: 95–107.
- Humphris, A., McMaster, T., Miles, M., Gilbert, S., Shewry, P., and Tatham, A. 2000. Atomic force microscopy (AFM) study of HMW subunits of wheat glutenin. *Cereal Chem.* **77**: 107–110.
- Kasarda, D.D. 1999. Glutenin polymers: The in vitro to in vivo transition. *Cereal Foods World* **44**: 566–571.
- Kasarda, D.D., King, G., and Kumosinski, T.F. 1994. Comparison of β -spiral structures in wheat high molecular weight glutenin subunits and elastin, by molecular modeling. In *Proceedings of the Symposium on Molecular Modeling* (eds. T.F. Kumosinski and M.N. Liebman). ACS, 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. Unters. Forsch.* **200**: 432–439.
- Kobrehel, K. and Alary, R. 1989. The role of a low molecular weight glutenin fraction in the cooking quality of durum wheat pasta. *J. Sci. Food Agric.* **47**: 487–500.
- Köhler, P., Belitz, H., and Wieser, H. 1991. Disulphide bonds in wheat gluten: Isolation of a cysteine peptide from glutenin. *Z. Lebensm. Unters. Forsch.* **192**: 234–239.
- . 1993. Disulphide bonds in wheat gluten: Further cysteine peptides from high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin and from γ -gliadins. *Z. Lebensm. Unters. Forsch.* **196**: 239–247.
- Köhler, P., Keck-Gassenmeier, B., Wieser, H., and Kasarda, D.D. 1997. Molecular modeling of the N-terminal regions of high molecular weight glutenin subunits 7 and 5 in relation to intramolecular disulfide bond formation. *Cereal Chem.* **74**: 154–158.
- Lupas, A., Van, D., and Stock, J. 1991. Predicting coiled coils from protein sequences. *Science* **252**: 1162–1164.
- Marin, A., Pothier, J., Zimmermann, K., and Gibrat, J.-F. 2002a. FROST: A filter based fold recognition method. *Proteins* **49**: 493–509.
- . 2002b. Protein threading statistics: An attempt to assess the significance of a fold assignment to a sequence. In *Protein structure prediction: Bioinformatics approach* (ed. I.F. Tsigelny), Chap. 9, pp. 227–262. International University Line, La Jolla, CA.
- Matsushima, N., Creutz, C., and Kretsinger, R. 1990. Polyproline, β -turn helices. Novel secondary structures proposed for the tandem repeats within rhodopsin, synaptophysin, synexin, gliadin, RNA polymerase II, hordein, and gluten. *Proteins* **7**: 125–155.
- Matsushima, N., Danno, G., Sasaki, N., and Izumi, Y. 1992. Small-angle X-ray scattering study by synchrotron orbital radiation reveals that high molecular weight subunit of glutenin is a very anisotropic molecule. *Biochem. Biophys. Res. Commun.* **186**: 1057–1064.
- McGuffin, L., Bryson, K., and Jones, D. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404–405.
- Miles, M.J., Carr, H.J., McMaster, T.C., l'Anson, K., Belton, P.S., Morris, V.J., Field, J.M., Shewry, P.R., and Tatham, A.S. 1991. Scanning tunneling microscopy of a wheat seed storage protein reveals details of an unusual supersecondary structure. *Proc. Natl. Acad. Sci.* **88**: 68–71.
- Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T., and Morimoto, T. 1997. Tertiary and quaternary structures of 0.19 α -amylase inhibitor from wheat kernel determined by X-ray analysis at 2.06 Å resolution. *Biochemistry* **36**: 13503–13511.
- Parchment, O., Shewry, P., Tatham, A., and Osguthorpe, D. 2001. Molecular modeling of unusual structure in elastomeric wheat seed protein. *Cereal Chem.* **78**: 658–662.
- Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., and Chothia, C. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J. Mol. Biol.* **284**: 1201–1210.
- Payne, P. 1987. Genetics and wheat storage protein and the effect of allelic variation on breadmaking quality. *Ann. Rev. Plant Physiol.* **38**: 141–153.
- Payne, P., Corfield, K., and Blackman, J. 1979. Identification of a high-molecular-weight subunit of glutenin whose presence correlates with bread-making quality in wheats of related pedigree. *Theor. Appl. Genet.* **55**: 153–159.
- Pearson, W. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* **132**: 185–219.
- Popineau, Y., Cornec, M., Lefebvre, J., and Marchylo, B. 1994. Influence of high molecular glutenin subunits on glutenin polymers and rheological properties of gluteins and gluten sub-fractions of near-isogenic lines of wheat Sicco. *J. Cereal Sci.* **19**: 231–241.
- Sali, A. and Blundell, T. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**: 779–815.
- Shewry, P. and Tatham, A. 1997. Disulphide bonds in wheat gluten proteins. *J. Cereal Sci.* **25**: 207–227.
- Shewry, P., Halford, N., and Tatham, A. 1992. High molecular weight subunits of wheat glutenin. *J. Cereal Sci.* **15**: 105–120.
- Shimoni, Y., Blechl, A., Anderson, O., and Galili, G. 1997. A recombinant protein of two high molecular weight glutenins alters gluten polymer formation in transgenic wheat. *J. Biol. Chem.* **272**: 15488–15495.
- Strobl, S., Muhlhahn, P., Bernstein, R., Wiltsccheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R., and Holak, T. 1995. Determination of the three-dimensional structure of the bifunctional α -amylase/trypsin inhibitor from ragi seeds by NMR spectroscopy. *Biochemistry* **34**: 8281–8293.
- Strobl, S., Maskos, K., Wiegand, G., Huber, R., Gomis, R., and Glockshuber, R. 1998. A novel strategy for inhibition of α -amylases: Yellow meal worm α -amylase in complex with the Ragi bifunctional inhibitor at 2.5 Å resolution. *Structure* **6**: 911–921.
- Tao, H., Adalsteins, A., and Kasarda, D. 1992. Intermolecular disulfide bonds link specific high-molecular-weight glutenin subunits in wheat endosperm. *Biochim. Biophys. Acta* **1159**: 13–21.
- Thomson, N., Miles, M., Popineau, Y., Harries, J., Shewry, P., and Tatham, A. 1999. Small angle X-ray scattering of wheat seed-storage proteins: α -, γ - and ω -gliadins and the high molecular weight (HMW) subunits of glutenin. *Biochim. Biophys. Acta* **1430**: 359–366.
- Wootton, J. and Federhen, S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Computers Chem.* **17**: 149–163.