




Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <http://oatao.univ-toulouse.fr/27137>

Official URL: <https://doi.org/10.1016/j.ijbiomac.2020.09.136>

To cite this version:

Sahli, Line and Boire, Adeline and Solé-Jamault, Véronique and Rogniaux, Hélène and Giuliani, Alexandre and Roblin, Pierre  and Renard, Denis *New exploration of the γ -gliadin structure through its partial hydrolysis.* (2020) International Journal of Biological Macromolecules, 165 (Part A). 654-664. ISSN 0141-8130

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

New exploration of the γ -gliadin structure through its partial hydrolysis

Line Sahli ^a, Adeline Boire ^a, Véronique Solé-Jamault ^a, Hélène Rogniaux ^{a,b}, Alexandre Giuliani ^{c,d}, Pierre Roblin ^e, Denis Renard ^{a,*}

^a INRAE, UR1268 Biopolymers Interactions Assemblies, 44300 Nantes, France

^b INRAE, BIBS Platform, 44300 Nantes, France

^c DISCO Beamline, Synchrotron Soleil, L'Orme des Merisiers, 91192 Gif sur Yvette, France

^d UAR 1008, Transform, INRAE, BP 71627, F-44316, Nantes, France

^e Laboratoire de Génie Chimique and Fédération de Recherche FERMAT, 4 allée Emile Monso, 31030 Toulouse, France

ARTICLE INFO

Keywords:

Wheat storage proteins
Intrinsically disordered proteins
3D structure modelling

ABSTRACT

The partial enzymatic hydrolysis of wheat gliadins constitutes an interesting tool to unravel their structural specificity. In this work, the structure and conformation of γ gliadin were investigated through its limited chymotryptic digestion. Using a combination of computational, biochemical and biophysical tools, we studied each of its N and C terminal domains. Our results reveal that γ gliadin is a partially disordered protein with an unfolded N terminal domain surprisingly resistant to chymotrypsin and a folded C terminal domain. Using spectroscopic tools, we showed that structural transitions occurred over the disordered N terminal domain for decreasing ethanol/water ratios. Using SAXS measurements, low resolution 3D structures of γ gliadin were proposed. To relate the repeated motifs of the N terminal domain of γ gliadin to its structure, engineered peptide models PQQPY/F were also studied. Overall results demonstrated similarities between the N terminal domain and its derived model peptides. Our findings support the use of these peptides as general templates for understanding the wheat protein assembly and dynamics.

1. Introduction

Wheat gliadins are storage proteins that belong to the large prolamin family [1–5]. They are synthesized in the lumen of the endoplasmic reticulum (ER) and stored into micrometer sized organelles called protein bodies (PBs) [6–8]. Although these proteins are known to be essential during seed germination and growth, so far, their structure remains hypothetical.

Gliadins are monomeric proteins, soluble in an alcohol/water mixture or acidic solutions [1,9,10]. According to their electrophoretic mobility and isoelectric focusing, they are classified into three main types: α/β , γ (sulfur rich) and ω gliadins (sulfur poor) [2,9,11,12]. The amino acid sequences of sulfur rich gliadins consist of two separate domains: a N terminal domain composed of repeated sequences, and a non-repetitive C terminal domain containing disulfide bonds [3,4,10,13,14]. The sulfur poor gliadins sequence contains a long repetitive domain, lacking cysteines. Repetitive domains of whole gliadins are composed by interspersed repeats of motifs rich in proline, glutamine, and tyrosine (α/β , ω gliadins) or phenylalanine (γ gliadins) [13,15]. Previous spectroscopic studies have shown that their repeated domains result of poly proline II (PPII) helix and β turn structures, while their

sulfur rich domains are essentially composed of α helix [16–18]. Small Angle X-ray Scattering (SAXS) measurements of purified gliadins [19,20] or gliadins mixtures [21,22], in aqueous ethanol mixture [20,22] or acidic solutions [19], were performed. All the findings converge to describe gliadins as extended proteins with an ellipsoidal conformation [23]. In this present work, we intend to reconsider the structure of gliadins in the light of advances in the field of intrinsically disordered proteins (IDPs) [24–27]. A better understanding of their structure will allow to determine the contribution of these specific sequences in the self-assembly and storage of gliadins in wheat seed.

γ gliadin, comprising both a N terminal repetitive domain and a C terminal non-repetitive domain, has been chosen as a representative model of gliadins. Due to the polymorphism of wheat proteins, various isoforms of γ gliadins have been identified [9,12]. To avoid this molecular diversity, we only focus on the γ 44 isoform. As previously described [28], extracted γ 44 gliadin was enzymatically digested and released domains were purified separately. To unravel the importance of the repeated motifs in the aforementioned process, pentapeptides of (PQQPF) and (PQQPY) representing the consensus sequences, respectively found in γ gliadins and $\alpha/\beta/\omega$ gliadins, were designed. The influence of their length on their structure was also studied by repeating those 8 or 17 times. In this study, we showed that γ gliadin is partially disordered. With SAXS measurement, we proposed a partial 3D structure at low resolution of γ 44 gliadin and we discussed the limitations

* Corresponding author.

E-mail address: denis.renard@inrae.fr (D. Renard).

of this approach. We demonstrated strong similarities between the N terminal domain and the model peptides.

2. Material and methods

2.1. *In silico* analyses

The analysis of protein sequences was carried out using ExPASy (<https://web.expasy.org/protscale/>) and CIDER (<https://pappulab.wustl.edu/CIDER/>) tools. Structural predictions were performed by i Tasser software (https://zhanglab.ccmb.med.umich.edu/I_TASSER/). In the case of γ gliadin and its domains, referenced and reviewed P08453 accession was used (UniProtKB).

2.2. Purification of domains

γ 44 gliadin was extracted and purified from wheat gluten of cv. Hardi as previously described [29]. Domains were obtained by the chymotryptic digestion of the γ 44 gliadin [28]. γ 44 gliadin was first solubilized (10 mg/mL) in Tris HCl pH 8/ethanol mixture (60/40). Chymotrypsin contained in 10 mM CaCl₂ buffer, was added at enzyme/protein ratio of 1: 500. Enzymatic hydrolysis was carried out at room temperature, for 24 h, under continuous stirring. Hydrolyzed samples were then applied on a semi preparative column of Nucleosil C 18 (300 Å, 7 μ m, 250 \times 10 mm² Macherey nagel) equilibrated with deionized water containing 0.06% of trifluoroacetic acid (TFA). Elution was performed using an acetonitrile gradient containing 0.06% TFA (15–100%). The N terminal domain does not absorb at 280 nm, elution was therefore monitored at 215 and 280 nm. Purified domains were collected. Acetonitrile and TFA solvents were evaporated using a rotatory evaporator. Fractions were then freeze dried and stored at room temperature.

2.3. Synthesis and purification of peptides

(PQQPY)₈ and (PQQPF)₈ were synthesized and purified on an Inertsil ODS 3 column (4.6 \times 250 mm) by GenScript (Piscataway NJ, USA). The synthetic gene of (PQQPY)₁₇ was inserted on the pETb32 vector (Novagen) and expressed as a fusion protein with thioredoxin and His₆ tag as previously described [30,31]. Asp Pro residues were introduced between the recombinant peptide and the thioredoxin/His tag to allow their subsequent separation. Protein expression was performed in *Escherichia coli* BLR (DE3) pLys strain (Novagen) in the presence of 0.5 mM isopropyl β D 1 thiogalactopyranoside (IPTG). Glycerol stocks of the transformed bacteria (stored at -80°C) were used to inoculate the overnight Luria Bertani (LB) pre culture at 37°C . Cells were induced at 30°C in Terrific broth (TB) media containing 25 $\mu\text{g}/\text{ml}$ chloramphenicol and 50 $\mu\text{g}/\text{ml}$ ampicillin. After the expression, the cells collected by centrifugation (5000 $\times g$, 20 min, 4°C) were washed and re-suspended in Ni NTA binding buffer (20 mM Tris pH 8, 0.5 M NaCl, and 20 mM imidazole). Cells were lysed by sonication (3 cycles of 1.5 min, 4°C) and harvested at 20,000 $\times g$ for 20 min. Pellets were further resuspended with the Ni NTA binding buffer supplemented with 6 M of urea. Cell debris were removed by a final centrifugation step performed at 20,000 $\times g$ for 20 min. Resulting supernatants were purified by Ni NTA column (HisTrapTMHP). Proteins were eluted from the column using an imidazole gradient (20–500 mM). Collected fractions were dialyzed against deionized water and freeze dried. An acid cleavage was carried out to recover the repetitive polypeptide. The fusion proteins were dispersed in 70% formic acid at 37°C for 72h. The resulting samples were then loaded on the Nucleosil C 18 column (300 Å, 7 μ m, 250 \times 10 mm, Macherey nagel) equilibrated with 0.06% of TFA in deionized water. Elution was performed using an acetonitrile gradient containing 0.06% TFA (10–75%) in 90 min. Collected fractions were then dialyzed against deionized water, freeze dried and stored at room temperature.

2.4. Analytical characterization

For biochemical characterization, protein powders of γ 44 gliadin and its domains (150 μg) were solubilized in Tris HCl pH 8/ethanol mixture (60/40) (v/v) and applied on an analytical column of Nucleosil C 18 (300 Å, 5 μ m, 250 \times 4 mm, Macherey nagel) equilibrated with deionized water containing 0.06% of TFA. Elution was performed using an acetonitrile gradient containing 0.06% TFA (15–100%) in 60 min. Protein samples were diluted with Laemli buffer and heated at 95°C for 5 min for SDS PAGE analysis (4–12% Bis Tris Plus Gels, BoltTM, Invitrogen). After migration, electrophoresis gel was incubated overnight in Instant blue solution (Expedeon) to ensure gel coloration. In the case of the N terminal domain, a further silver nitrate staining was carried out. The gel was rinsed in distilled water and scanned.

2.5. Mass spectrometry (MS) analysis

Purity and molecular weight of proteins were determined using a SYNAPT G2Si HDMS mass spectrometer (Waters, France) with an electrospray source of ionization (ESI). Measurements were performed in the positive ion mode, on a mass to charge ratio (m/z) range of 800–3000. (PQQPY)₁₇ was analyzed at 0.1 mg/mL in a water acetonitrile mixture (50/50) (v/v) containing 0.1% formic acid (v/v). Domains were analyzed at 0.2 mg/mL in 60% of acetonitrile (v/v). Samples were infused at a flow rate of 5 $\mu\text{L}/\text{min}$ and spectra were recorded over 2 min. Spectrum processing was done using the MassLynxTM software (Waters, France).

2.6. Dynamic light scattering (DLS)

DLS measurements were performed using a zetasizer Nano Series (Nano ZS, Malvern instrument, Germany). Protein powders were dissolved in 0.05 M Na₂HPO₄/NaH₂PO₄ pH 7.2, 55% ethanol (v/v), under stirring, overnight, at room temperature. Residual undissolved material was removed by filtration on a 0.2 μm membrane filter (Sartorius, France). Protein samples (100 μL , 2.5 mg/mL) were analyzed in a quartz cuvette (Hellma analytics) with a path length of 10 mm. The measurements were performed in manual mode: 10 runs of 60 s were collected and repeated three times for each measurement (20°C). The correlation functions were analyzed using the CONTIN method.

2.7. Synchrotron radiation circular dichroism (SRCD)

Measurements were performed using the DISCO beamline at Soleil synchrotron (Gif sur Yvette, France) using the SRCD endstation [32].

The samples were prepared in 50 mM MOPS buffer pH 7.2, 25 mM NaCl at 20% or 55% ethanol (v/v). At 55% ethanol, protein samples are concentrated at 2 mg/mL, except (PQQPY)₁₇ concentrated at 0.72 mg/mL. At 20% ethanol, protein samples are concentrated as follows: 1.5 mg/mL for N terminal domain, 1.5 mg/mL for C terminal domain, 1.5 mg/mL for mixed domains (molar ratio 1:1), 0.53 mg/mL for (PQQPY)₁₇, 2 mg/mL for (PQQPY)₈ and 2 mg/mL for (PQQPF)₈. Each spectrum is the average of three acquisitions. The spectrum of buffer was subtracted from the protein spectrum using the CDtoolX software (<http://www.cdtools.cryst.bbk.ac.uk/>). Spectrum was smoothed using the Savitzky Golay filtering (order 3 out of 9 points). Content of secondary structures was determined using the BestSel software (<http://bestsel.elte.hu/index.php>).

2.8. Small angle X ray scattering (SAXS)

SAXS measurements were performed at the Chemical Engineering Laboratory (LGC Toulouse, France) with the XEUS 2.0 (Xenocs Company) composed of X ray micro source delivering at 8 keV a spot sized beam equal to 0.5 mm with intensity close to 30×10^6 photons/s. The sample to detector distance of the X rays was 1216.5 mm

providing a range of scattering vector starting from 0.005 \AA^{-1} to 0.5 \AA^{-1} . Protein samples were solubilized in 50 mM MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v), under stirring, overnight, at room temperature. Residual undissolved material was removed by centrifugation (30 min, $15,000 \times g$). Protein concentrations are as follows: 8.3 mg/mL for $\gamma 44$ gliadin, 8.4 mg/mL for N terminal domain, 5.0 mg/mL for C terminal domain and 9.3 mg/mL for mixed domains (molar ratio 1:1). The samples were exposed for 7200 s to the beam and the scattered intensity was collected on the 1 M Pilatus detector. The protein solvent was recorded and subtracted from the protein spectra. Data integration, reduction and averaging were performed using the FOXTROT software.

The theoretical R_g (\AA) for globular protein was calculated by the following equation [33]: $R_g = (3/5)^{1/2} R_h$, with R_h , the experimental hydrodynamic radius of the native protein.

The theoretical R_g (\AA) for a disordered protein was determined by the Flory's equation [34]: $R_g = R_0 N^\nu$ with N , the number of residues, $R_0 = 2.54 \pm 0.01 \text{ \AA}$ which is a constant that depends on the persistence length and ν a value comprised between $1/2$ and $3/5$. Under conditions where polymer polymer interactions are equal to polymer solvent and solvent solvent interactions, the polymer behaves as a Gaussian chain and the exponent ν is predicted to be $1/2$. An average value $\nu = 0.55$ was chosen corresponding to a theta solvent for the protein [35].

2.9. SAXS fitting using SasView v. 5.0

SAXS fitting using SasView v. 5.0 (<https://www.sasview.org/>) was used to fit experimental SAXS curves of N ter, C ter domains and $\gamma 44$ gliadin by geometrical models.

2.10. Ab initio modelling

The collected data of the entire protein was analyzed using the ATSAS program package (https://www.embl-hamburg.de/biosaxs/atsas_online/) and in particular the GASBOR program which consists in the ab initio reconstruction of a protein structure by a chain like ensemble of dummy residues and the MONSA program where a low resolution beads model is proposed using two phases and the three SAXS curves.

3. Results and discussion

3.1. Sequence properties and structural predictions of the γ gliadin domains

To determine the main features of the primary structure of γ gliadin domains, *in silico* analyses were performed. The P08453 accession (Uniprot KB) of γ gliadin is used for this purpose. Note that previous work has shown the high similarity between sequences of all the referenced γ gliadins found in the Uniprot KB database [36]. The amino acid sequences were run through CIDER [36] and ExPASy tools to be compared with a fully disordered protein (calpastatin P20810) and a globular protein (lysozyme, P83673). The fractions of charged residues (FCR) and the net charge per residue (NCPR) presented in Table 1 highlight

the weakly charged nature of the γ gliadin and its domains. The low charge content of γ gliadin and its truncated domains suggests that they belong to the weak polyelectrolytes/polyampholytes category which could adopt globules or tadpoles conformations (compact conformation) [37]. Similar results were recently obtained for α gliadin and low molecular weight glutenin [35]. Besides, the composition analysis of γ gliadin amino acid sequence suggests a significant enrichment in proline and glutamine as compared to other IDPs (see Supplementary Fig. S1). Both γ gliadin and C terminal domain are basic proteins with a theoretical isoelectric point (pI) ranging from 8 to 8.5. The N terminal domain has a theoretical pI close to 5. According to the hydropathy values, C terminal is suggested to be the most hydrophobic domain of γ gliadin. Resulting to its low complexity sequence, N terminal domain is predicted to be disordered (Table 1). Disorder predictors also suggest a disordered N terminal domain [36]. Further analyses through the tertiary and secondary structural predictor i Tasser [38] confirm these observations. The repetitive sequence is predicted to be unfolded while the C terminal domain is predicted to contain α helix structures. As the C terminal of the γ gliadin is predicted to be structured, a particular attention is paid to this protein domain (Fig. 1A). i Tasser simulations generate a C terminal model with a moderate confidence (C score = -1.79) (Fig. 1B). Based on these results, structural alignment program has been performed to match the C terminal model to all structures in the PDB library. According to i Tasser, the protein from the PDB that have the closest structural similarity with the C terminal domain is a globular protein, termed as Brazil nut 2S albumin (2LVF). Sequence homology between the non repetitive domain of α gliadin with 2S storage proteins was already highlighted and was suggested to be related to a common ancestral gene [39,40].

Finally, *in silico* results suggest that γ gliadin is a weakly charged protein, comprising two distinct domains of equivalent length that exhibit different physical chemical properties. Analyses converge to predict that γ gliadin is an hybrid protein which could adopt globular or tadpole conformations. The non repetitive C terminal domain is predicted to be globular while the repetitive N terminal domain is predicted to be disordered. These predictions are in agreement with previous SRCD results [36]. A high content of disordered structures (46.9%) has been determined in the $\gamma 44$ gliadin full length and have been attributed to its repetitive N terminal domain. However, even if the low sequence complexity of the N terminal domain is consistent with its hypothetical disorder, its low charge content contrast with IDPs [41]. Moreover, the absence of enzymatic hydrolysis within this domain, as previously reported by Legay et al. [28], raises some questions. Limited proteolysis experiments are usually used to probe the unfolded conformation of IDPs [42]. The N terminal domain should therefore exhibit hypersensitivity to proteases [25,42,43]. Ultimately, all these elements show the importance of experimentally studying the structure of each of the domains.

3.2. Composition of purified domain samples

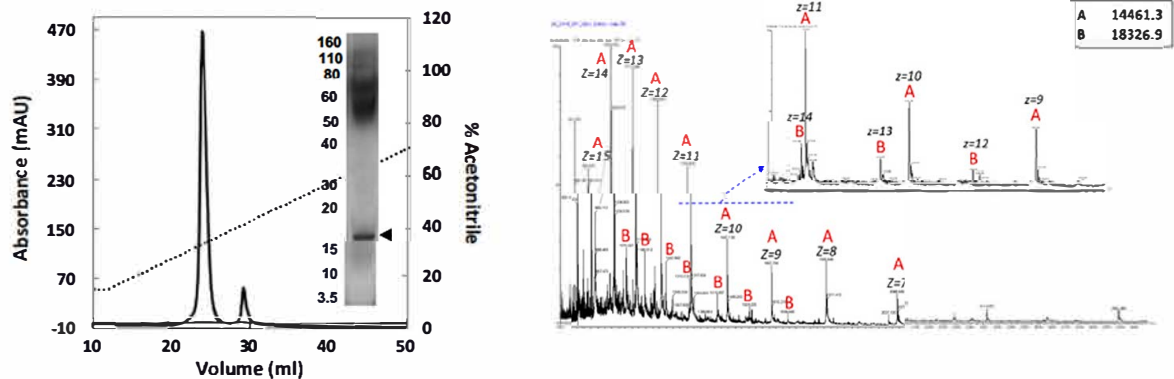
As previously stated, $\gamma 44$ gliadin has been extracted and purified from the wheat seed [29,36]. According to Legay et al. method [28], repetitive (N terminal) and non repetitive (C terminal) domains were

Table 1

Sequence properties of proteins. pI, isoelectric point; k, charge patterning parameter. k values close to zero indicate that oppositely charged residues are well-mixed, while k values close to one indicate that oppositely charged residues are segregated. FCR, fraction of charged residues. NCPR, net charge per residue which is the difference between the fractions of positively charged and negatively charged residues. Disorder promoting scores above 0.5 predict disordered sequences. Disorder promoting score below 0.5 predict ordered sequence. To compare with our proteins, two model proteins were also analyzed: calpastatin corresponds to a fully disordered protein model while lysozyme corresponds to a globular protein model.

Protein	Residues	pI	Hydropathy	k	FCR	NCPR	Proline fraction	Disorder promoting
Calpastatin	708	4.97	3.31	0.14	0.38	0.05	0.10	0.82
N-terminal domain	162	5.08	3.24	0.69	0.01	0	0.25	0.78
γ -Gliadin	326	8.20	8.20	0.21	0.04	0	0.17	0.70
C-terminal domain	164	8.50	4.37	0.24	0.07	0.01	0.09	0.62
Lysozyme	184	8.88	4.18	0.26	0.16	0.06	0.02	0.59

A/ N-terminal domain



B/ C-terminal domain

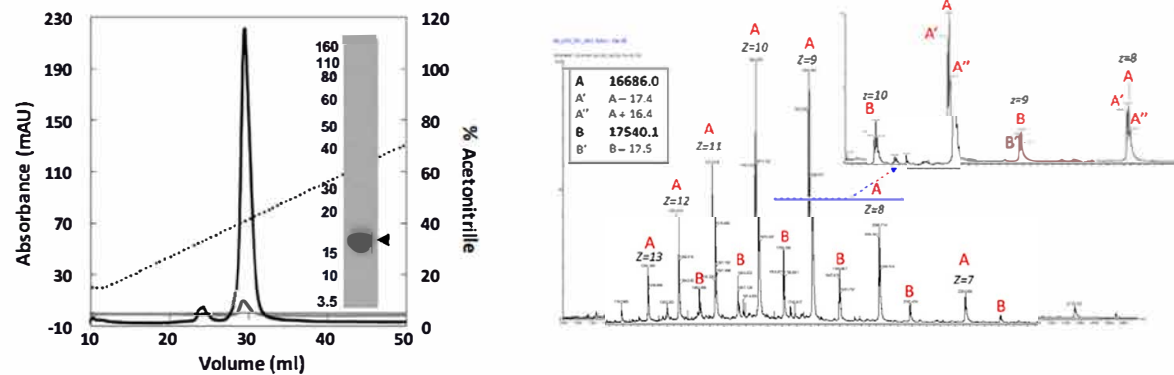


Fig. 2. Composition and purity of protein samples. A. (Left) Chromatographic profile of N-terminal domain at 215 (black) and 280 nm (grey), eluted with acetonitrile gradient (15–100%) containing 0.06% TFA. SDS-PAGE of N-terminal domain (black arrow at ~16–17 kDa) revealed by silver nitrate staining. The smear observed at 50–110 kDa is due to a bad destaining of the gel. (Right) ESI-MS spectrum of N-terminal domain analyzed at 0.2 mg/mL in 60% of acetonitrile (v/v). B. (Left) Chromatographic profile of C-terminal domain at 215 (black) and 280 nm (grey), eluted with acetonitrile gradient (15–100%) containing 0.06% TFA. SDS-PAGE of C-terminal domain (black arrow at ~16–17 kDa) revealed by Instant blue staining. (Right) ESI-MS spectrum of C-terminal domain analyzed at 0.2 mg/mL in 60% of acetonitrile (v/v).

3.4. The $\gamma 44$ gliadin intrinsic disorder arises from its N terminal domain

To estimate the secondary structure content of the γ gliadin domains, synchrotron radiation circular dichroism spectra in far ultraviolet region were recorded in good solvent conditions (0.05 M MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v)) (Fig. 4). The spectroscopic study of the full length $\gamma 44$ gliadin was previously reported [36]. Spectral deconvolution of N terminal domain spectrum using Bestsel software [46] suggests the high disordered content of the N terminal domain (51.5% unordered structures, 2.5% α helix, 27.3% β strand, 18.7% turn). Surprisingly, the C terminal domain would also exhibit a high disorder content (41.8% unordered structures, 32.6% α helix, 9.4% β strand, 16.3% turn). On the other hand, the high α helical secondary structure proportion found in the C terminal moiety is highlighted. Signal of the mixed domains spectrum is greatly noisy (Fig. 4A). The N and C terminal spectra were averaged to construct a theoretical mixed domains spectrum and submitted to the Bestsel software. The predictions obtained (23.2% α helix, 16.3% β strand, 15.8% turn and 44.3% unordered structures) show no noticeable differences with the secondary structure content of the $\gamma 44$ gliadin full length (Fig. 4B).

Finally, spectroscopic data suggests that N terminal is a low structured domain and C terminal is a domain rich in α helix. These observations are in line with the previous *in silico* analysis. These results are also in agreement with the established structural data of the γ gliadins [16,17], but it should be noticed that the Bestsel tool does not allow to quantify polyproline II type helix, considering them as disordered structures. Concerning the high disorder content found in the C terminal domain, this is consistent with the high content of coil predicted by i Tasser (Fig. 1A). However, this may also be related to the partial

hydrolysis of the C terminal domain as shown in the Supplementary Information. The deconvolution of CD spectra into structural elements relevant for IDPs relies on datasets including denatured proteins [58]. The interpretation in terms of intrinsic disorder should therefore be taken with care.

3.5. N terminal domain appears to fold upon decreasing ethanol content

The evolution of secondary structures of proteins at 20% of ethanol was also investigated (0.05 M MOPS pH 7.2, 25 mM NaCl and 20% ethanol (v/v)) (Fig. 5). As $\gamma 44$ gliadin undergoes a phase separation under these conditions [36], SR CD has not been measured on the full length protein. Some structural transitions are observed when the water content of samples is increased (Fig. 5 & Table 2). N terminal domain tends to lose disordered structures in favor of β strands (from 51.5% unordered structures, 27.3% β strands at 55% ethanol (v/v) to 41.1% unordered structures, 38.9% β strands at 20% ethanol (v/v)). C terminal domain folds into α helix (from 32.6% at 55% ethanol (v/v) to 45.2% at 20% ethanol (v/v)), while disordered content slightly increases (from 41.8% at 55% ethanol (v/v) to 49.1% at 20% ethanol (v/v)). A loss in β structures (strands, turns) is observable and would explain the increase of α helix and disorder contents. Mixed domains still have a high disordered content at 20% ethanol (v/v) (15.9% α helix, 14.9% β strand, 15.1% turn and 54.1% unordered structures). Since data of mixed domains at 55% ethanol were obtained by theoretical approach, they could not be compared with our present results.

Finally, results tend to show structural transitions upon increasing aqueous content. The repetitive N terminal domain appears to fold into β strand structures while the non repetitive C terminal domain

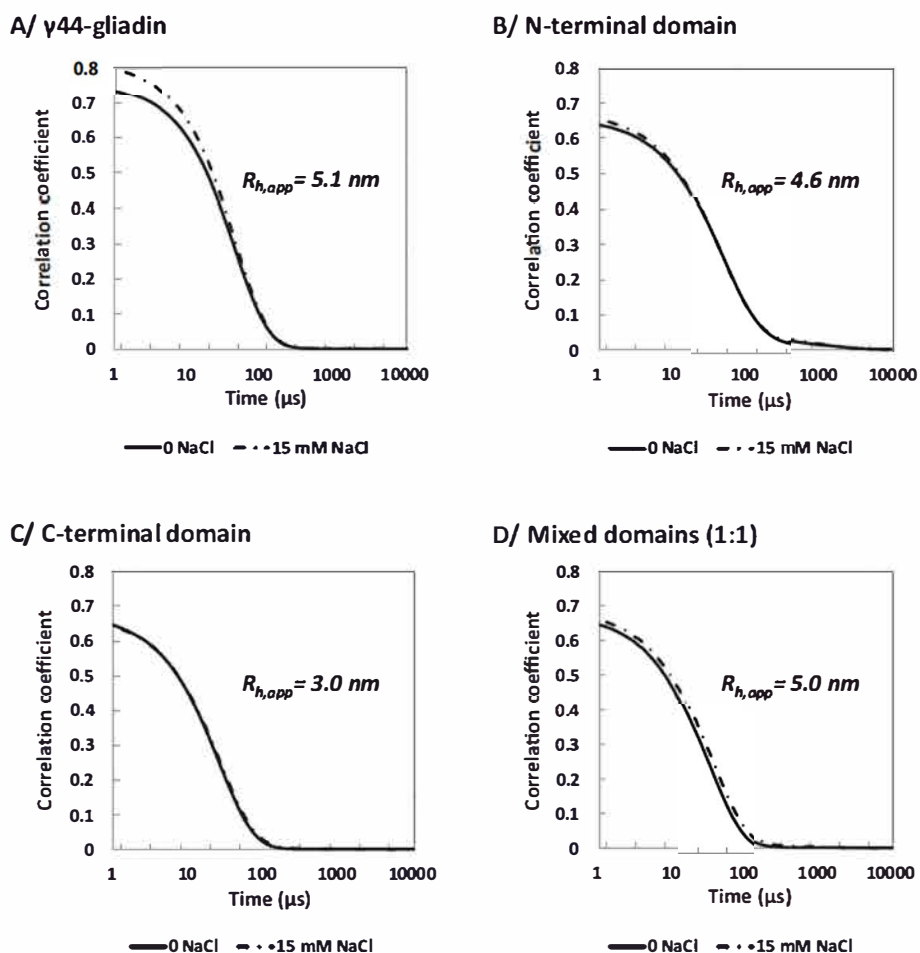


Fig. 3. Correlograms of γ 44-gliadin and its domains obtained by DLS. A. γ 44-gliadin. B. N-terminal domain. C. C-terminal domain. D. Mixed domains at molar ratio of 1:1. Correlograms are obtained at 2 mg/mL in 0.05 M MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v) (20 °C).

seems to gain α helix and disordered content in detriment to the β structures. The presence of disorder to order transitions within the N terminal may be consistent with the propensity of IDPs to this type of transition, due their flexibility and lack of structural constraint [25,47,48].

3.6. Structural features of γ 44 gliadin and its domains as revealed by SAXS

SAXS experiments were performed to determine the structure of γ 44 gliadin and its domains. The scattering curves of γ 44 gliadin and

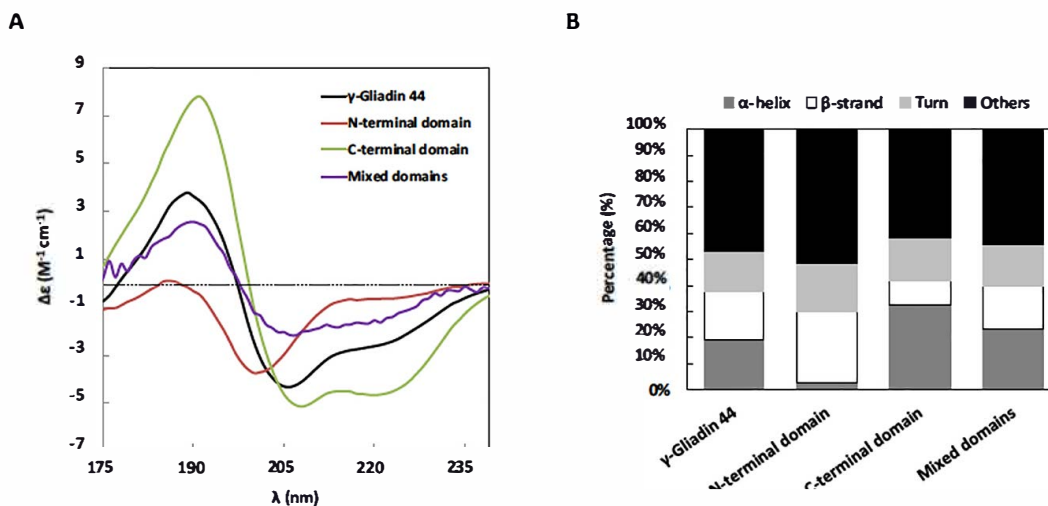


Fig. 4. Far-UV CD spectra of constructions under good solvent conditions. A. CD spectra of γ 44-gliadin at 2 mg/mL (black), N-terminal domain at 2 mg/mL (wine), C-terminal domain at 2 mg/mL (green) and mixed domains (1:1 molar ratio) at 2mg/mL (purple) in 0.05 M MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v). B. The histograms show secondary structures content obtained from the deconvolution of the spectra using BestSel software.

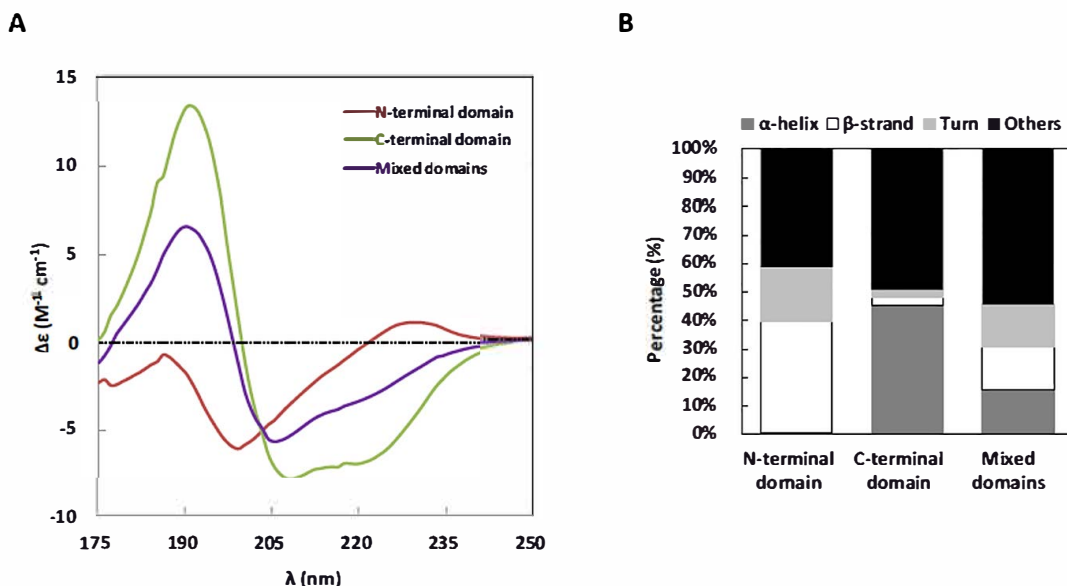


Fig. 5. Far-UV CD spectra of constructions upon increasing aqueous content. A. CD spectra of N-terminal domain at 1.5 mg/mL (wine), C-terminal domain at 1.5 mg/mL (green) and mixed domains (1:1 molar ratio) at 1.5 mg/mL (purple) in 0.05 M MOPS pH 7.2, 25 mM NaCl and 20% ethanol (v/v). B. The histograms show secondary structures content from the deconvolution of the spectra using BestSel software.

its derived domains are displayed in Fig. 6A where the absolute scattered intensity I (cm^{-1}) is plotted versus the scattering wave vector $\text{Log } q$ (\AA^{-1}). The Guinier analyses (for $qR_g < 1.0$) give a radius of gyration (R_g) of 62.5 \AA for $\gamma 44$ gliadin, 39.6 \AA for N terminal domain and 27.1 \AA for C terminal domain (Supplementary Information). The $\gamma 44$ gliadin R_g value is much higher than those previously determined in the literature (26.6 \AA in 1% acid acetic and 38 \AA in 70% ethanol (v/v)) [19,20]. However, R_g is strongly dependent on solvent conditions. It ranged from 43.8 to 69.7 \AA by decreasing water content in ethanol/water mixtures [49]. To deepen our analyses, we use the previously determined R_h values to calculate the theoretical R_g values. According to a globular model [33], a theoretical R_g of 23.2 \AA was determined for the C terminal domain. Considering the protein disorder, a R_g of 61.2 \AA was calculated for the $\gamma 44$ gliadin using Flory's equation [34] (Material and methods). The same equation was used to establish a theoretical R_g of 41.7 \AA for the N terminal domain. On the other hand, the relation dedicated to fully unfolded protein models [33] shows a considerable overestimation of the R_g of the N terminal domain (120 \AA). Finally, the experimental R_g values are in very close agreement with theoretical ones, meaning that the conformations and solvent conditions were fully taken into account in the equations used.

M_w^{app} of 51 kDa ($\gamma 44$ gliadin), 22 kDa (N terminal domain) and 41 kDa (C terminal domain) were estimated from the extrapolated scattering intensity at zero angle $I(0)$ (Fig. 6A). In comparison with the data derived from mass spectrometry, it clearly appears that molecular weights extrapolated from SAXS experiments are largely overestimated. This result may be justified by the presence of aggregates in the samples and the slight upturn of scattering intensity at low q values. The disorder of the N terminal domain and the order of the C terminal domain are visible in the shape of the pair distance

Table 2
Structural comparison at 55% and 20% of ethanol (v/v). Difference values (Diff) correspond to the differences of the values obtained at 55% and 20% of ethanol (v/v).

Ethanol %	N-terminal domain			C-terminal domain			Mixed domains		
	55%	20%	Diff	55%	20%	Diff	55%	20%	Diff
α -Helix	2.5	0.7	1.8	32.6	45.2	12.6	23.2	15.9	7.3
β -Strand	27.3	38.9	11.6	9.4	3.1	6.3	16.3	14.9	1.4
Turn	18.7	19.4	0.7	16.3	2.6	13.7	15.8	15.1	0.7
Others	51.5	41.1	10.4	41.8	49.1	7.3	44.3	54.1	9.8

distribution functions $p(r)$ (Fig. 6B). The maximum distance (D_{max}) found for $\gamma 44$ gliadin, N terminal domain and C terminal domain are respectively of 255 \AA , 150 \AA and 100 \AA . The large D_{max} of the N terminal domain suggests an extended conformation, while the lower D_{max} of the C terminal domain suggests a compact conformation. The sum of D_{max} of N ter and C ter domains is in agreement with the D_{max} calculated for the entire protein. The R_g values calculated from the $P(r)$ functions are respectively of 65.3 \AA for $\gamma 44$ gliadin, 41.7 \AA for N terminal domain and 27.7 \AA for C terminal domain. $P(r)$ functions displayed R_g values closer to the experimental and theoretical R_g values, in line with the lesser sensitivity of aggregation in the R_g calculation from $P(r)$ function. In the case of the full length $\gamma 44$ gliadin and the N terminal domain, Krakty plots displayed in Fig. 6C show the presence of a plateau from 0.1 \AA^{-1} , in agreement with their high disorder content ("unfolded" conformation). An increasing slope at high q values of -1.7 is obtained for the N ter domain, value characteristic of unstructured proteins. While a slope of -4.0 is found for the C ter domain, value characteristic of a globular shape. The presence of a maximum at 0.08 \AA^{-1} in the C terminal curve confirms the compact or globular conformation of this domain (Fig. 6B). Finally, a slope of -2.0 is found for the $\gamma 44$ gliadin, value intermediate between those of the N ter and the C ter domains.

Scattering data of the N terminal domain was fitted on the 0.005–0.5 \AA^{-1} scattering vector q range by a flexible cylinder model (or wormlike chain model) with a length of 392.9 \AA , a Kuhn length of 21.3 \AA and a radius of 3.5 \AA ($\chi^2 = 2.9$) while the C terminal domain was fitted by a tri axial ellipsoid model, with a minor, major and polar radii of 11.1, 22.9 and 54.6 \AA , respectively ($\chi^2 = 2.4$) (Supplementary Information). In addition, the fitting of the full length protein on the 0.01–0.5 \AA^{-1} scattering vector q range from a combination of the two form factors identified for the two domains, i.e. sum of tri axial ellipsoid plus wormlike chain models, gives a very reasonable result ($\chi^2 = 2.8$). However, as the cross term between the two form factors is neglected in the fitting procedure, the final result has no physical meaning and is considered as being not fully satisfying.

3.7. Low resolution *ab initio* models of $\gamma 44$ gliadin

To determine the overall shape of $\gamma 44$ gliadin, low resolution models were computed using an *ab initio* approach. GASBOR software

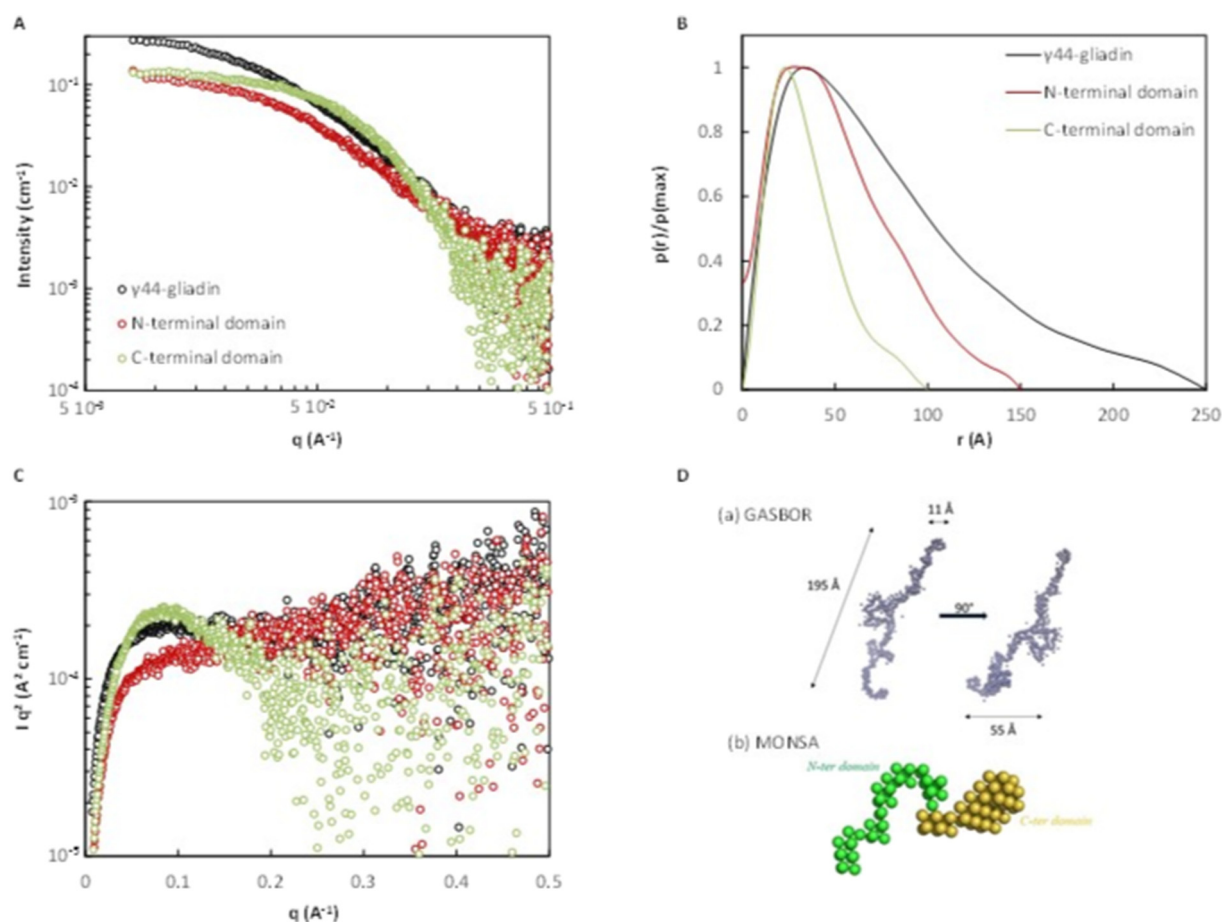


Fig. 6. SAXS analysis of $\gamma 44$ -gliadin and its domains. A. Scattering intensity $\text{Log } I(q)$ vs $\text{Log } q$ (\AA^{-1}) of $\gamma 44$ -gliadin at 8.3 mg/mL (blue), N-terminal domain at 8.4 mg/mL (orange) and C-terminal domain at 5 mg/mL (grey) in 0.05 M MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v). B. Pair distance distribution function $p(r)/P_{\text{max}}$ vs the distance r (\AA) calculated from the scattering curves using the PRIMUS software [53]. C. Kratky plots $I(q) \cdot q^2$ vs q (\AA^{-1}) of scattering curves displayed in A. D. Low resolution model of $\gamma 44$ -gliadin displayed by GASBOR (a) and MONSA (b) software.

[50] was run using the reciprocal space mode (slow calculation). The symmetry of the protein was left as P1 (no symmetry, default value). The proposed model is shown in Fig. 6D. The $\gamma 44$ gliadin is predicted to be an elongated protein with an estimated length of 195 \AA (Fig. 6D, a). Similar conformation was obtained using the DAMMIF software [51] (data not shown). A strong structural similarity with the globular Brazil Nut 2S albumin (2LVF) has been previously suggested by i Tasser predictions. To confirm these predictions, the scattering profile of the Brazil Nut 2S albumin was calculated using CRY SOL [52] and compared (through its PDB structure) to the experimental scattering data of the C terminal domain. No good fitting was obtained between the Brazil

Nut 2S albumin theoretical curve and the C terminal experimental curve (Supplementary Information). The same observation was done when the C terminal domain scattering data was compared with the i Tasser model. These results suggest structural differences between the domain and the two models predicted by the i Tasser tool.

For comparison, $\gamma 44$ gliadin was submitted to MONSA software [54]. This multiphase dummy atom modelling tool allows the simultaneous fitting of multiple curves leading to *ab initio* shape determination of a full length protein from the two domains. The proposed model of $\gamma 44$ gliadin is consistent with the previous results by highlighting the worm like chain conformation of the N terminal domain (length of

Table 3
Structural parameters of $\gamma 44$ -gliadin and its domains derived from DLS and SAXS experiments.

Structural and hydrodynamic parameters	$\gamma 44$ -gliadin	N-ter domain	C-ter domain
Molecular weight (kDa) (from mass spectrometry)	38.6	14.4	16.6
Molecular weight (kDa) (from Guinier)	51	22	41
R_h^{DLS} (\AA) (from DLS)	51	46	30
$I(0)$ (cm^{-1})	0.29	0.12	0.13
R_g (\AA) (from Guinier)	62.5	39.6	27.1
$\frac{R_g}{R_h}$ (from Guinier)	1.22	0.86	0.90
R_g (\AA) (from $P(r)$)	65.3	41.7	27.7
$\frac{R_g}{R_h}$ (from $P(r)$)	1.28	0.91	0.92
R_g (\AA) theoretical (calculated)	61.2 ^a	41.7 ^a	23.2 ^b
D_{max} (\AA)	255	150	100
Form factor	Sum of wormlike chain and triaxial ellipsoid	Wormlike chain	Triaxial ellipsoid

^a According to Flory's equation [34].

^b According to a globular model equation [33].

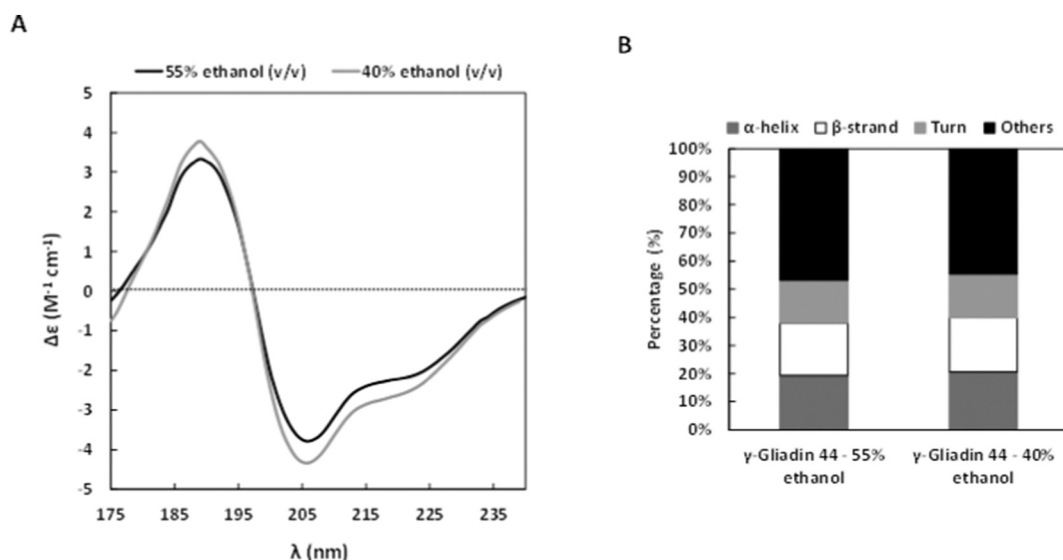


Fig. 7. Far-UV CD spectra of γ 44-gliadin in 55% and 40% ethanol (v/v). A. CD spectra of γ 44-gliadin in 55% (black) and 40% ethanol (v/v) (grey) at 2 mg/mL in 0.05 M MOPS pH 7.2, 25 mM NaCl. B. The histograms show secondary structures content from the deconvolution of the spectra using BestSel software.

154 Å) and the compactness of the C terminal domain (length of 74 Å) (Fig. 6D, b). Even if all the findings are consistent, it should be noted that *ab initio* calculations correspond to low resolution models that cannot necessarily reflect the reality. The absence of crystallographic data and cryo TEM observations on gliadins in solution do not allow the confirmation of these models.

Finally, SAXS experiments provide ensemble averaged structural information about the γ 44 gliadin and its isolated domains. All structural data obtained are summarized in Table 3. SAXS experiments are complementary to CD results. They also suggest that the N terminal domain is unordered in contrast to the globular like shape of the C terminal domain. In agreement with the previous *in silico* analysis, the low resolution models suggest that γ 44 gliadin could adopt a tadpole conformation with an ellipsoidal C terminal head. Note that the low resolution γ 44 gliadin model proposed in this study is in agreement with the first γ gliadin computer molecular model proposed in 1995 [55]. It is also consistent with a recent molecular model of a α gliadin obtained by Monte Carlo simulations [35]. In both models, an elongated shape was obtained with a tail in the vicinity of its N terminal domain. Interestingly, in the latter model, a ring like motif was also observed on snapshots when disulfide bonds were present. Similarly, a ring was obtained on the low resolution model of the C terminal domain obtained by GASBOR (Supplementary Information, Fig. S4).

More generally, it should be pointed out that the chimeric conformation of the protein, extended and globular mixed conformation, has underlined the limitations of the SAXS data processing tools. Due to its hybrid nature, it is very challenging to elucidate its 3D structure by taking into account its two domains. MONSA software was therefore the best compromise tool to probe the conformation of the entire protein from the two domains.

In this work, we carried out a detailed structural characterization of a γ gliadin. We proved its semi disordered nature by studying the

secondary and tertiary structure of each of its domains. These investigations were possible through a partial protein hydrolysis. The disordered N terminal domain is not hydrolyzed even after 24H of chymotrypsin digestion, its high flexibility being thought to promote the proteolytic targeting. On the other hand, the ordered protein domain is slightly hydrolyzed as shown in the Supplementary Information. All these elements could suggest that cleavage sites contained in the N terminal sequence of the γ 44 gliadin are not accessible to the chymotrypsin, in the proteolysis conditions used (40% ethanol (v/v)). To investigate the secondary structure of γ 44 gliadin in these specific conditions, further SRCD experiments were performed (Fig. 7). The resulted secondary structures content show no difference between 55% (19.4% α helix, 18.6% β strand, 15.2% turn and 46.9% unordered structures) and 40% of ethanol (v/v) (20.6% α helix, 19.2% β strand, 15.6% turn and 44.7% unordered structures). The secondary structure of the protein cannot therefore explain the absence of cleavage within the N terminal domain. Further SAXS experiments should be performed to determine the conformation adopted by the protein at 40% of ethanol. The three dimensional structure of the protein could limit the proteolysis reaction. Note that the chymotryptic insensitivity of the N terminal domain could also be related to its high proline content. Previous studies have reported that proline turns could induce conformational hardening of proteins [56,57]. It could be therefore concluded that the gaining of domains by partial proteolysis of the entire protein has indirectly provided valuable information about the γ 44 gliadin conformation.

3.8. Peptides models summarize the physicochemical properties and secondary structure of the N terminal domain

To determine whether (PQQPY)₈, (PQQPF)₈ and (PQQPY)₁₇ behave as the N terminal domain, we carried out biochemical (Supplementary Information) and structural assays. Their primary structures were first

Table 4

Sequence properties of model peptides. pI, isoelectric point; k, charge patterning parameter. k values close to zero indicate that oppositely charged residues are well-mixed, while k values close to one indicate that oppositely charged residues are segregated. FCR, fraction of charged residues. NCPR, net charge per residue which is the difference between the fractions of positively charged and negatively charged residues. Disorder promoting scores above 0.5 predict disordered sequences. Disorder promoting score below 0.5 predict ordered sequence.

Protein	Residues	pI	Hydropathy	k	FCR	NCPR	Proline fraction	Disorder promoting
(PQQPY) ₁₇	85	5.88	2.20	-	0	0	0.40	0.80
(PQQPY) ₈	40	5.92	2.20	-	0	0	0.40	0.80
(PQQPF) ₈	40	5.96	3.02	-	0	0	0.40	0.80

analyzed *in silico* through CIDER [36] and ExPASy tools (Table 4). Repeated peptides exhibit a theoretical pI close to that of the N terminal domain (-5.6). As expected, (PQQPY)₈, (PQQPF)₈ and (PQQPY)₁₇ are not charged peptides suggesting that they belong to the weak polyelectrolytes/polyampholytes category. As for the N terminal domain, their redundant sequences are predicted to be disordered. Regarding these predictions, peptides appear to share common physicochemical features with the N terminal domain.

The secondary structure content of model peptides were also recorded (0.05 M MOPS pH 7.2, 25 mM NaCl) at 55% and 20% ethanol (v/v) to be compared to the other γ 44 gliadin truncated forms (Fig. 8). At 55% ethanol, (PQQPF)₈ (53.6%), (PQQPY)₈ (46.9%) and (PQQPY)₁₇ (49.4%) display a high disorder content. The different lengths do not influence the peptide secondary structures. Note that the recombinant (PQQPY)₁₇ was only recorded from 190 nm, probably due to the contaminants present as aforementioned that strongly absorbed at $\lambda < 190$ nm. Since (PQQPY)₈ and (PQQPY)₁₇ spectra are too noisy at 20% of ethanol, only (PQQPF)₈ spectrum has been analyzed. In comparison to the previous SRCD results (55% ethanol (v/v)), spectral deconvolution of (PQQPF)₈ at 20% ethanol (v/v) shows a remarkable structuration in β strand (+9.7% of β strand and -5.6% of unordered structures). This induced folding into β strand structures at lower ethanol content was also observed in the case of the N terminal domain (Table 2).

Taken together, all these data demonstrate that peptides exhibit structural and behavioral similarities with the repetitive γ 44 gliadin N terminal domain. Other investigations on these simplified sequences could enable to understand the aforementioned chymotrypsic insensitivity of the domain. As an instance, mutation of the peptide prolines would confirm their potential role in the hypothetical conformational hardening of the N terminal. More broadly, our engineered peptides, inspired by the repetitive sequences of all gliadins, may act as mimicking models to study the self assembly of wheat proteins in further experiments. It is still unclear how the intrinsic disorder of the N terminal domain contributes to the biological function of wheat gliadins. It may contribute to the tendency of gliadins to undergo liquid liquid phase separation highlighted in previous studies [22,36]. Also, it may enhance specific interactions with the membrane of the endoplasmic reticulum during their synthesis in wheat seed [64]. Engineered peptides mimicking N terminal domain may help to better understand the driving force of gliadins self assembly occurring during the phases of accumulation and storage in wheat seed.

4. Conclusion

Through partial hydrolysis and combination of bioinformatics, physicochemical and biophysical tools, we were able to revisit the structure of a wheat gliadin model. We demonstrated the semi disordered conformation of the γ 44 gliadin by analyzing each of its domains. Structural assessments underlined the difficulty to elucidate the three dimensional structure of a hybrid protein. An approach combining SAXS curves fitting using geometrical models and *ab initio* computations revealed that γ 44 gliadin could be ascribed to an elongated protein with a tadpole conformation and an estimated maximum distance of 200-250 Å. The enzymatic reaction, exhibiting a disordered N terminal resistant domain, highlighted the particularity of the studied protein and contributed to unravel its molecular features.

The structural characterization of repetitive polymers like architecture has shown their strong similarities with the repeated gliadin sequences from which they derive. Our findings support their use as general templates for understanding the wheat protein assembly and dynamics.

CRedit authorship contribution statement

Line Sahli: methodology, acquisition of data, analysis and interpretation of data, drafting the manuscript.

Adeline Boire: critical reading of the manuscript, supervision of Line Sahli's work.

Véronique Solé Jamault: purification of wheat proteins.

Hélène Rogniaux: mass spectrometry data acquisition and analysis.

Alexandre Giuliani: supervision at SOLEIL synchrotron beamline for SRCD experiments.

Pierre Roblin: SAXS data analysis.

Denis Renard: critical reading of the manuscript, supervision of Line Sahli's work

Acknowledgments

This work was carried out with the financial support of the regional programme "Food for Tomorrow/Cap Aliment, Research, Education and Innovation in Pays de la Loire", which is supported by the French Region Pays de la Loire and the European Regional Development Fund (FEDER).

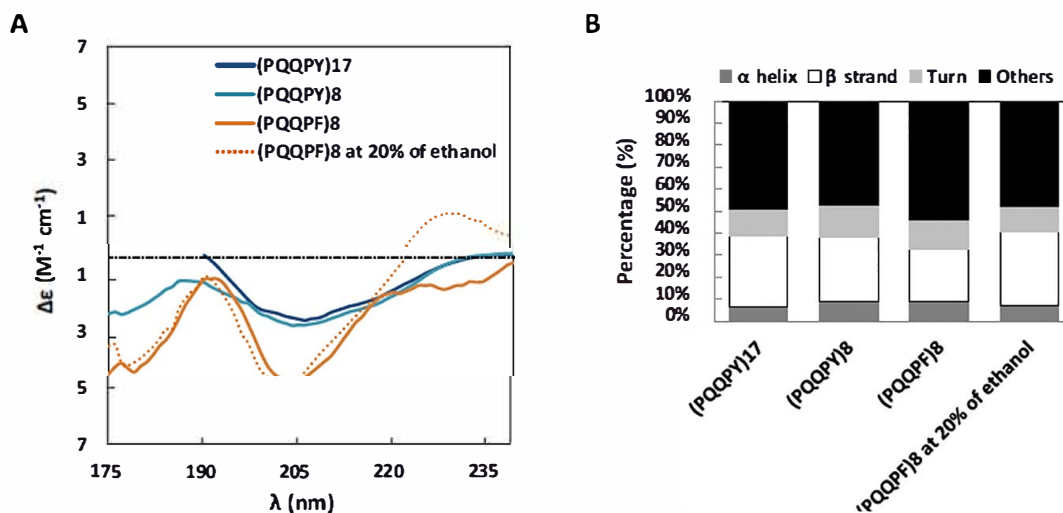


Fig. 8. Far-UV CD spectra of model peptides. A. CD spectra of (PQQPY)₁₇ at 0.72 mg/mL (dark blue), (PQQPY)₈ at 2 mg/mL (light blue) and (PQQPF)₈ at 2 mg/mL (orange) in 0.05 M MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v). The spectra of (PQQPF)₈ at 2 mg/mL (orange) in 20% ethanol (v/v) is represented in orange dotted lines. B. The histograms show secondary structures content from the deconvolution of the spectra using BestSel software.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2020.09.136>.

References

- [1] T.B. Osborne, *Vegetable Proteins*, Longmans, London, 1924 2.
- [2] P.R. Shewry, A.S. Tatham, The prolamin storage proteins of cereal seeds: structure and evolution, *Biochem. J.* 267 (1990) 1–12.
- [3] Shewry, Seed storage proteins: structures and biosynthesis, *Plant Cell Online* 7 (1995) 945–956.
- [4] P.A. Sabelli, P.R. Shewry, Characterization and organization of gene families at the Gli-1 loci of bread and durum wheats by restriction fragment analysis, *Theor. Appl. Genet.* 83 (1991) 209–216.
- [5] P. Shewry, What is gluten—why is it special? *Front. Nutr.* 6 (2019).
- [6] G. Galili, Y. Altschuler, H. Levanony, Assembly and transport of seed storage proteins, *Trends Cell Biol.* 3 (1993) 437–442.
- [7] P. Tosi, et al., Trafficking of storage proteins in developing grain of wheat, *60* (2009) 979–991.
- [8] P. Tosi, Trafficking and deposition of prolamins in wheat, *J. Cereal Sci.* 56 (2012) 81–90.
- [9] Y. Popineau, F. Pineau, Fractionation and characterisation of γ -gliadins from bread wheat, *J. Cereal Sci.* 3 (1985) 363–378.
- [10] P. Shewry, B.J. Mifflin, D.D. Kasarda, The structural and evolutionary relationships of the prolamin storage proteins of barley, rye and wheat, *Philos. Trans. R. Soc. B Biol. Sci.* 304 (2006) 297–308.
- [11] H. Wieser, Chemistry of gluten proteins, *Food Microbiol.* 24 (2007) 115–119.
- [12] C. Larre, Y. Popineau, W. Loisel, Fractionation of gliadins from common wheat by cation exchange FPLC, *J. Cereal Sci.* 14 (1991) 231–241.
- [13] P. Shewry, N. Halford, Cereal seed storage proteins: structures, properties and role in grain utilization, *53* (2002) 947–958.
- [14] C. Benitez-Cardoza, H. Rogniaux, Y. Popineau, J. Guéguen, Cloning, bacterial expression, purification and structural characterization of N-terminal-repetitive domain of γ -gliadin, *Protein Expr. Purif.* 46 (2006) 358–366.
- [15] K.A. Feeney, et al., Synthesis, expression and characterisation of peptides comprised of perfect repeat motifs based on a wheat seed storage protein, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1546 (2001) 346–355.
- [16] A. Tatham, P. Shewry, The conformation of wheat gluten proteins. The secondary structures and thermal stabilities of α -, β -, γ - and ω -gliadins, *J. Cereal Sci.* 3 (1985) 103–113.
- [17] Tatham, Masson, Popineau, Conformational studies of peptides derived by the enzymic hydrolysis of a gamma-type gliadin, *J. Cereal Sci.* 11 (1990) 1–13.
- [18] S.M. Gilbert, et al., Expression and characterisation of a highly repetitive peptide derived from a wheat seed storage protein, *Biochim. Biophys. Acta, Proteins Proteomics* 1479 (2000) 135–146.
- [19] Thomson, Miles, Tatham, Shewry, Molecular images of cereal proteins by STM, *Ultramicroscopy* 42–44 (1992) 1204–1213.
- [20] N. Thomson, et al., 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 Protein Struct. Mol. Enzymol.* 1430 (1999) 359–366.
- [21] N. Sato, et al., Molecular assembly of wheat gliadins into nanostructures: a small-angle X-ray scattering study of gliadins in distilled water over a wide concentration range, *J. Agric. Food Chem.* 63 (2015) 8715–8721.
- [22] A. Boire, C. Sanchez, M.H. Morel, M.P. Lettinga, P. Menu, Dynamics of liquid-liquid phase separation of wheat gliadins, *Sci. Rep.* 8 (2018) 1–13.
- [23] R. Urade, N. Sato, M. Sugiyama, Gliadins from wheat grain: an overview, from primary structure to nanostructures of aggregates, *Biophys. Rev.* 10 (2018) 435–443.
- [24] A.K. Dunker, et al., Intrinsically disordered protein, *J. Mol. Graph. Model* 19 (2001) 26–59.
- [25] P. Tompa, Intrinsically unstructured proteins, *Trends Biochem. Sci.* 27 (2002) 527–533.
- [26] H.J. Dyson, P.E. Wright, N.T. Pines, Intrinsically unstructured proteins and their functions, *Mol. Cell Biol.* 6 (2005) 197–208.
- [27] J. Zhou, C.J. Oldfield, W. Yan, B. Shen, A.K. Dunker, Intrinsically disordered domains: sequence \rightarrow disorder \rightarrow function relationships, *Protein Sci.* 28 (2019) 1652–1663.
- [28] C. Legay, Y. Popineau, S. Bérot, J. Guéguen, Comparative study of enzymatic hydrolysis of α/β and γ -gliadin, *Mol. Nutr. Food Res.* 41 (1997) 201–207.
- [29] A. Banc, et al., Structure and orientation changes of ω - and γ -gliadins at the air-water interface: a PM-IRRAS spectroscopy and Brewster angle microscopy study, *Langmuir* 23 (2007) 13066–13075.
- [30] K. Elmorjani, et al., Synthetic Genes Specifying Periodic Polymers Modelled on the Repetitive Domain of Wheat Gliadins: Conception and Expression, vol. 246, 1997 240–246.
- [31] S. Source, A. Nisole, J. Gu, Y. Popineau, K. Elmorjani, High microbial production and characterization of strictly periodic polymers modelled on the repetitive domain of wheat gliadins, *Biochem. Biophys. Res. Commun.* 312 (2003) 989–996.
- [32] M. Réfrégiers, et al., DISCO synchrotron-radiation circular-dichroism endstation at SOLEIL, *J. Synchrotron Radiat.* 19 (2012) 831–835.
- [33] D.K. Wilkins, et al., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques, *Biochemistry* 38 (1999) 16424–16431.
- [34] P. Flory, Principles of polymer chemistry, *J. Am. Chem. Soc.* 76 (1953) Cornell University Press.
- [35] J. Markgren, M. Hedenqvist, F. Rasheed, M. Skepö, E. Johansson, Glutenin and gliadin, a piece in the puzzle of their structural properties in the cell described through Monte Carlo simulations, *Biomolecules* 10 (2020) 1095.
- [36] L. Sahli, D. Renard, V. Solé-jamault, A. Giuliani, A. Boire, Role of protein conformation and weak interactions on γ -gliadin liquid-liquid phase separation, *Sci. Rep.* 9 (2019) 1–13.
- [37] R.K. Das, K.M. Ruff, R.V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins, *Curr. Opin. Struct. Biol.* 32 (2015) 102–112.
- [38] A. Roy, A. Kucukural, Z. Yang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protoc.* 5 (2010) 725–738.
- [39] M. Kreis, B.G. Forde, S. Rahman, B.J. Mifflin, P.R. Shewry, Molecular evolution of the seed storage proteins of barley, rye and wheat, *J. Mol. Biol.* 183 (1985) 499–502.
- [40] M. Kreis, P.R. Shewry, Unusual features of cereal seed protein structure and evolution, *BioEssays* 10 (1989) 201–207.
- [41] P. Romero, et al., Sequence complexity of disordered protein, *Proteins Struct. Funct. Genet.* 42 (2001) 38–48.
- [42] A. Fontana, et al., Probing protein structure by limited proteolysis, *51* (2004) 299–321.
- [43] V. Receveur-Bréhot, J.M. Bourhis, V.N. Uversky, B. Canard, S. Longhi, Assessing protein disorder and induced folding, *Proteins Struct. Funct. Genet.* 62 (2006) 24–45.
- [44] R. Pitt-Rivers, F.S. Ambesi-Impombato, The binding of sodium dodecyl sulphate to various proteins, *Biochem. J.* 109 (1968) 825–830.
- [45] L.M. lakoucheva, et al., Aberrant mobility phenomena of the DNA repair protein XPA, *Protein Sci.* 10 (2001) 1353–1362.
- [46] A. Micsonai, et al., BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, *Nucleic Acids Res.* 46 (2018) W315–W322.
- [47] V. Uversky, Natively unfolded proteins: a point where biology waits for physics, *Protein Sci.* 11 (2002) 739–756.
- [48] R.B. Berlow, H.J. Dyson, P.E. Wright, Functional advantages of dynamic protein disorder, *FEBS Lett.* 589 (2015) 2433–2440.
- [49] A.M. Orecchioni, C. Duclairoir, D. Renard, E. Nakache, Gliadin characterization by sans and gliadin nanoparticle growth modelization, *J. Nanosci. Nanotechnol.* 6 (2006) 3171–3178.
- [50] D.I. Svergun, M.V. Petoukhov, M.H. Koch, Determination of domain structure of proteins from x-ray solution scattering, *Biophys. J.* 80 (2001) 2946–2953.
- [51] D. Franke, D.I. Svergun, DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *J. Appl. Crystallogr.* 42 (2009) 342–346.
- [52] Svergun, Barberato, Koch, CRYSOLE - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, *J. Appl. Crystallogr.* 28 (1995) 768–773.
- [53] P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch, D.I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *J. Appl. Crystallogr.* 36 (2003) 1277–1282.
- [54] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys. J.* 76 (1999) 2879–2886.
- [55] R. D'Ovidio, M. Simeone, S. Masci, E. Porceddu, D.D. Kasarda, Nucleotide sequence of a gamma-type glutenin gene from a durum wheat: correlation with a gamma-type glutenin subunit from the same biotype, *Cereal Chem.* 72 (1995) 443–449.
- [56] T.N. Kirkland, F. Finley, K.I. Orsborn, J.N. Galgiani, Evaluation of the proline-rich antigen of *Coccidioides immitis* as a vaccine candidate in mice, *Infect. Immun.* 66 (1998) 3519–3522.
- [57] G. Fischer, H. Bang, E. Berger, A. Schellenberger, Conformational specificity of chymotrypsin toward proline-containing substrates, *Biochim. etBiophysicaActa* 791 (1984) 87–97.
- [58] N. Sreerama, S.Y. Venyaminov, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem.* 287 (2000) 243–251.
- [64] A. Banc, et al., Exploring the interactions of gliadins with model membranes: effect of confined geometry and interfaces, *Biopolymers* 91 (2009) 610–622.

Supplementary information

Compositional profile

The fasta sequence of γ 44-gliadin was analyzed through the webserver Composition Profiler (<http://www.cprofiler.org/>¹) to detect bias in its amino acid composition as compared to a set of proteins that display a well-defined 3D structure (PDB select 25). The analysis reveals that γ 44-gliadin is significantly enriched in proline and glutamine (**Figure S1A**). Also, it is depleted in charged amino acids (D,E, K, R). For comparison, the protein database Disprot is compared to the PDB Select 25 database (Figure S1B). It shows that Disprot proteins are enriched in disorder promoting amino acids (in red) and depleted in order promoting amino acids (in blue). Therefore, γ 44-gliadin may be a specific case of partially disordered proteins.

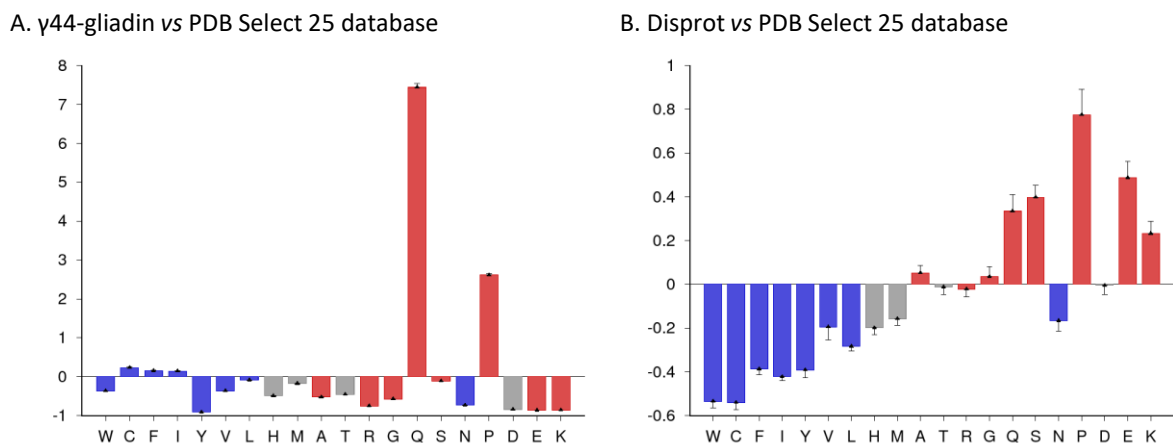


Figure S1. Composition profiles of A. γ 44-gliadin against PDBS25 dataset, and B. Disprot dataset against PDBS25 dataset. Red: disorder-promoting amino acids ; Blue: order-promoting amino acids ; Grey: neutral in regards to disorder.

Hydrolysis and purification of γ 44-gliadin domains

To isolate N and C terminal domains of the γ 44-gliadin, chymotryptic digestion of the protein was monitored in Tris-HCl pH 8/ethanol mixture (60/40), at room temperature, during 24H. The enzymatic kinetic was followed by polyacrylamide gel electrophoresis (SDS-PAGE 10%) under reducing and non-reducing conditions (Figure S1) as previously described²⁸. The appearance of two bands with a very close molecular weight (\sim 17 kDa) is observed after 1H hydrolysis. Samples after reduction revealed that observed bands correspond to the C-terminal domain, which is known to contain all cysteine residues of the γ -gliadin^{3,13}. These observations are consistent with the previous mass spectrometry data that identified the presence of two C-terminal populations (16.6 kDa and 17.5 kDa). The bands observed at 3 kDa and 14 kDa appearing after reduction correspond to fragments of the C-terminal domain. These findings show a partial hydrolysis of the C-terminal domain caused by the chymotrypsic digestion and maintained by the disulfide bonds. For unclear reasons, but certainly due to the particularity of its sequence (no basic amino acids)⁴⁴, N-terminal domain is not revealed by Coomassie staining, and therefore not visible by SDS-PAGE (\sim 14kDa) (**Figure S2A**). The injection of the γ 44-gliadin hydrolysates on C-18 column reveals the presence of

¹ Vacic V., Uversky V.N., Dunker A.K., and Lonardi S. "Composition Profiler: A tool for discovery and visualization of amino acid composition differences". *BMC Bioinformatics*. 8:211. (2007)

two major peaks corresponding to the N and C terminal domains, respectively eluted at 33% and 45% of acetonitrile (0.06% TFA) (**Figure S2B**). Note that the N-terminal domain, mainly composed of phenylalanines, do not absorb at 280 nm (**Figure S2B**, grey curve).

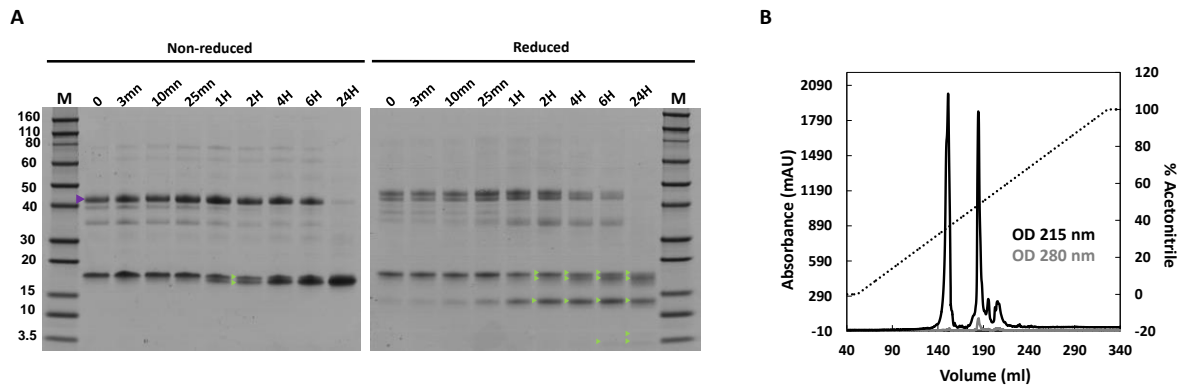


Figure S2: Hydrolysis of γ 44-gliadin by chymotrypsin reaction. A. Enzymatic kinetics of chymotrypsin followed by SDS-PAGE 10% under non reduced (*left*) and reduced (10 mM dithiothreitol) (*right*) conditions. **B.** Chromatographic profile of the γ 44-gliadin hydrolysates, injected on a semi-preparative column of Nucleosil C-18, observed at 215 (black) and 280 nm (grey), eluted with acetonitrile gradient (0-100% - 0.06% TFA).

Guinier analyses of SAXS experiments

To determine the structure of γ 44-gliadin and its domains, SAXS experiments were performed. The Guinier plots of γ 44-gliadin and its derived domains are displayed in **Figure S3** where $\ln I$ (cm^{-1}) is plotted vs q^2 (\AA^{-2}). The Guinier analyses (for $qR_g < 1.0$) gave a radius of gyration (R_g) of 62.5 \AA for γ 44-gliadin, 39.6 \AA for N-terminal domain and 27.1 \AA for C-terminal domain.

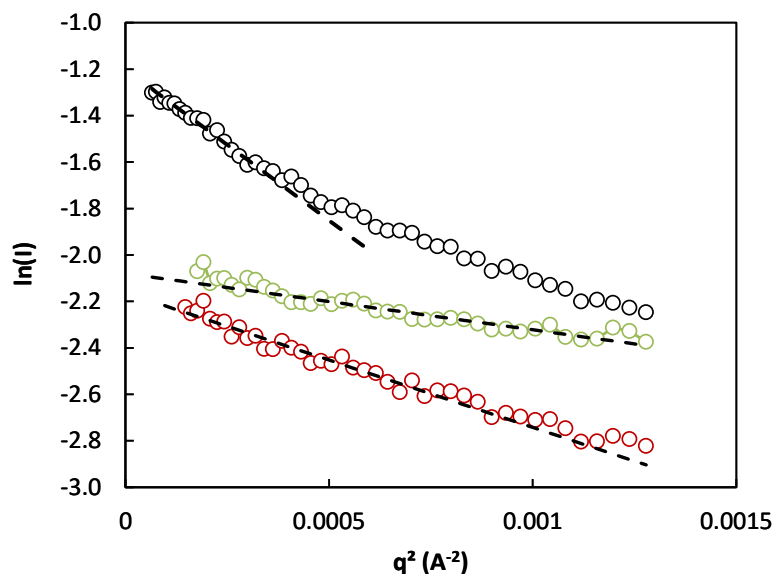


Figure S3: Guinier analyses. Guinier plots $\ln I(q)$ vs q^2 (\AA^{-2}) of γ 44-gliadin at 8.3 mg/mL (*black*), N-terminal domain at 8.4 mg/mL (*red*) and C-terminal domain at 5 mg/mL (*green*) in 0.05 M MOPS pH 7.2, 25 mM NaCl and 55% ethanol (v/v). Dashed lines stand for the Guinier fit.

Fitting SAXS curves using geometrical models

SAXS curves of N-ter and C-ter domains were preliminary fitted using SasView v. 5.0 in order to find the best realistic geometrical models. C-ter domain was best fitted using a triaxial ellipsoid model while N-ter domain was best fitted using a wormlike chain or flexible cylinder model (**Figure S4, A & B**). Fitting curves of γ 44-gliadin, N-ter and C-ter domains from MONSA computing were also displayed (**Figure S4, C**), given some confidence to the *ab initio* modelling.

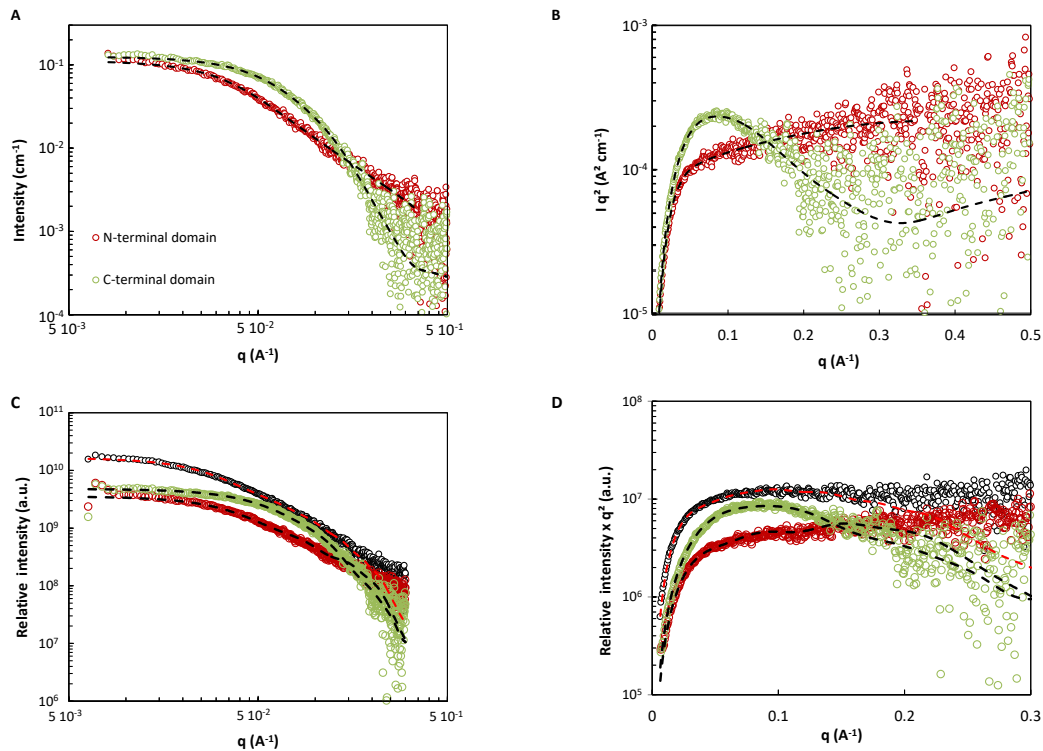


Figure S4: SAXS fitting. Geometrical models computed using SasView (A, B) for the N-ter (*red*) and C-ter (*green*) domains of γ 44-gliadin and using MONSA (C, D) for the N-ter (*red*) and C-ter (*green*) domains of γ 44-gliadin and γ 44-gliadin (black). Data are plotted as Intensity vs q (A, C) and as Kratky plot (B, D).

Ab initio model of the C-terminal domain obtained using GASBOR tool

To determine the overall shape of the C-terminal domain, a low-resolution model was computed using an *ab initio* approach similarly to the γ 44-gliadin. GASBOR software was run using the reciprocal space mode (slow calculation). The symmetry of the proteins was left as P1 (no symmetry, default value). The proposed models are shown in **Figure S5**. The C-terminal domain adopts a compact ellipsoidal conformation with an estimated length of 85 Å. The dimensions obtained are very similar to the one obtained by computer simulation on the amino acid sequence of a γ -type glutenin, closed to γ -gliadin's one.

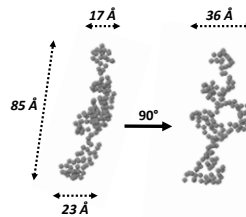


Figure S5. Low resolution model of the C-terminal domain of γ 44-gliadin displayed by GASBOR.

Assessment of the C-terminal domain model proposed by i-Tasser predictions

i-Tasser predicts a strong structural similarity between the C-terminal domain of the γ -gliadin and the globular Brazil Nut 2S albumin (2VLF). Simulations have also generate a C-terminal model with a moderate confidence (C-score= -1.79). To confirm all these predictions, the scattering profiles of the Brazil Nut 2S albumin and the i-Tasser model were compared to the experimental scattering data of the C-terminal domain using CRY SOL program⁵¹. Results show no good fitting between the Brazil Nut 2S albumin theoretical curve and the C-terminal experimental curve (**Figure S6**, *black curve*). The same observation was made in the case of the proposed i-Tasser model (**Figure S6**, *red curve*).

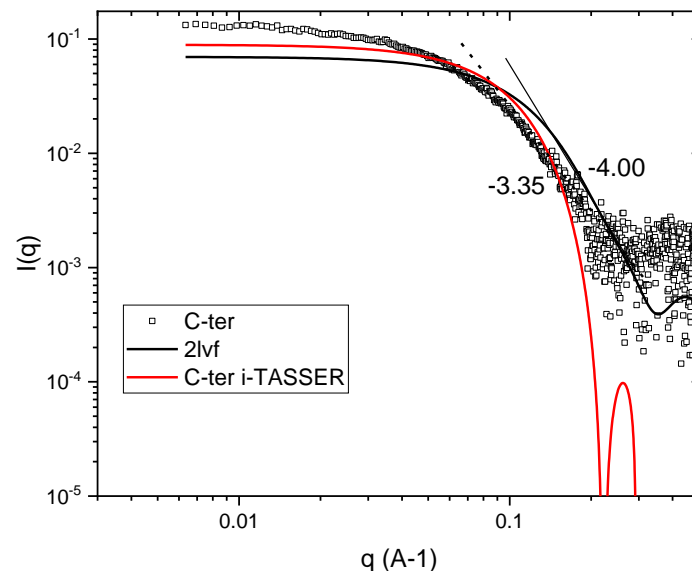


Figure S6: Comparison of the calculated scattering intensity of Brazil Nut 2S albumin and i-Tasser model to the C-terminal domain using CRY SOL software.

Biochemical characterization of the recombinant (PQQPY)₁₇

(PQQPY)₈ and (PQQPF)₈ peptides were chemically synthesized while (PQQPY)₁₇ was obtained by heterologous expression³⁰. As with gliadin and its domain, analytical chromatography and spectrometry analyses were performed to determine the composition of peptide samples. The recombinant (PQQPY)₁₇ is eluted at ~33% of acetonitrile and exhibits a very unusual electrophoretic migration (smear), presumably due to its very highly redundant sequence (**Figure S7, left**). Mass spectrometry data display a molecular weight of 10.4 kDa (**Figure S7, right**). The protein content of the sample is estimated to be greater than 70%. Despite the presence of non-protein contaminants, which most probably correspond to salts and/or urea used during the purification, (PQQPY)₁₇ sample has an excellent protein purity (~100%). As aforementioned, sequences repeated 8 times were provided and analyzed by Genscript. (PQQPY)₈ and (PQQPF)₈ were respectively eluted at ~38% and ~47% of acetonitrile with an estimated purity of 88.5% and 95.6% (information provided by Genscript).

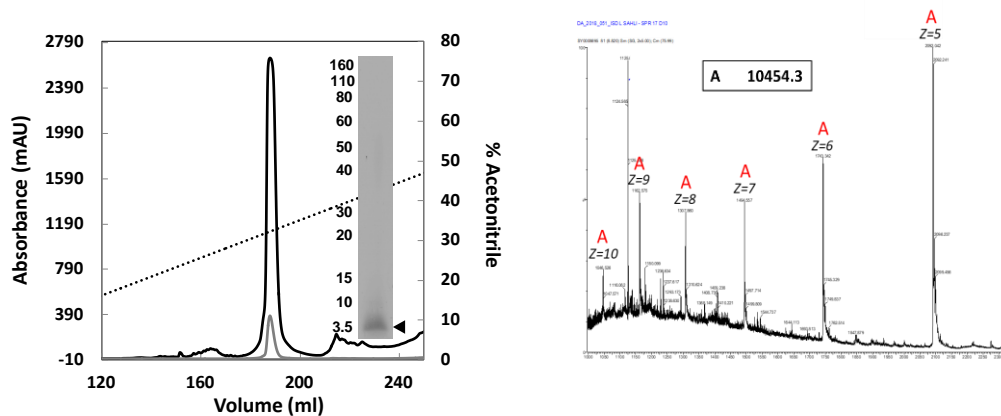


Figure S7: Composition and purity of the (PQQPY)₁₇ sample. (Left) Chromatographic profile of (PQQPY)₁₇ at 215 (black) and 280 nm (grey), eluted with acetonitrile gradient (10–75%) containing 0.06% TFA. SDS-PAGE of (PQQPY)₁₇ (black arrow at ~3.5 kDa) revealed by Instant blue staining. (Right) ESI-MS spectrum of (PQQPY)₁₇ at 0.1 mg/mL in water-acetonitrile mixture (50/50) (v/v) containing 0.1% formic acid (v/v)