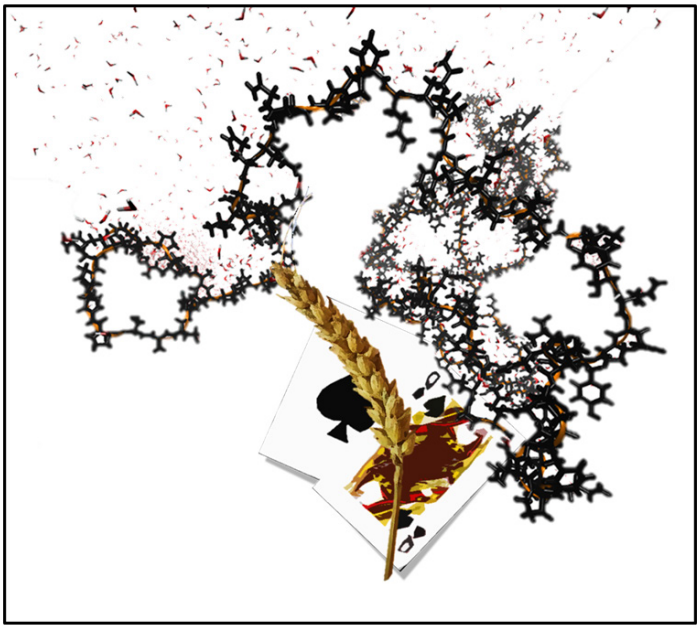
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2022:10
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Aggregation of gluten proteins - from wheat seed biology to hydrogels

Scientific modelling based primarily on Monte-Carlo and HPLC methods

Joel Markgren



Aggregation of gluten proteins - from wheat seed biology to hydrogels

Scientific modelling based primarily on Monte-Carlo and HPLC methods

Joel Markgren

Faculty of Landscape Architecture, Horticulture and Crop Production Science (LTV)

Department of Plant Breeding

Alnarp

SLU

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS

Alnarp 2022

Acta Universitatis agriculturae Sueciae
2022:10

Cover: α-gliadin molecule with water molecules, an ear of wheat and playing cards
(Illustration by Joel Markgren)

ISSN 1652-6880
ISBN (print version) 978-91-7760-895-0
ISBN (electronic version) 978-91-7760-896-7

Alnarp
Print: SLU Service/Repro, Alnarp 2022

Aggregation of gluten proteins - from wheat seed biology to hydrogels

Abstract

Gluten proteins are intrinsically disordered proteins that form extensive aggregated networks in wheat seeds, where they are stored as a nutrient source for the embryo. A modelling approach involving computational biology with Monte-Carlo algorithms and wet laboratory studies, including HPLC analysis, was applied to unravel the aggregational and hydrogel-forming properties of the gluten proteins. Two of the gluten proteins, "α-gliadin" and "low molecular weight glutenin subunits" (LMW-GS) were found to have similar size, folding of disordered, rigid and compact structures, elliptical shape and secondary structures of random coils and turns. Both proteins also share an evolutionarily conserved motif resulting in internal disulphide bonds, which were shown to be established through hydrophobic interactions, together with the inherent order of cysteines. In laboratory conditions and simulations, it was found that gliadins formed oligomers by hydrophobic interactions and cross-links by disulphide and lanthionine bonds at peptide sections in the C-terminal part of the protein. At the N-terminal part, the protein formed oligomers by liquid-liquid phase separation, polyproline II structures and β-sheets. Heat and alkaline treatment was shown to favour cross-linking by lanthionine, lysinoalanine and disulphide bonds among gliadins and increase their ability to absorb liquid. Thus the modelling approach successfully characterised the gluten proteins α-gliadin and LMW-GS, the mechanisms by which they form internal and external cross-links, how they merge into oligomers and how to increase their liquid absorption.

Keywords: gliadin, glutenin, super-absorbents, LMW-GS, HMW-GS, protein modelling, peptides, folding, lanthionine, lysinoalanine, disulphide

Author's address: Joel Markgren, Swedish University of Agricultural Sciences Alnarp, Department of Department of Plant Breeding, P.O. Box 190, 234 22 Lomma, Sweden.

Glutenproteiners aggregeringsförmåga - en studie om spännande fröbiologi och hydrogeler

Abstract

Glutenproteiner är oordnade proteiner som bildar omfattande aggregerade nätverk i vetefrön där de lagras som en näringskälla för embryot. För att beskriva proteinernas aggregerings-och hydrogelbildande egenskaper användes i denna avhandling ett modelleringsupplägg med beräkningsbiologi i form av Monte Carlo-algoritmer och laborationer med HPLC-analyser. De två glutenproteinerna ”α-gliadin” och ”Low Molecular Weight Glutenin Subunits” (LMW-GS) undersöktes och båda befanns ha likartade storlekar, elliptiska former, sekundära strukturer med slumpmässiga spiraler och böjningar, samt vikningar med oordnade, utdragna och kompakta strukturer. Båda proteinerna har även ett liknande och evolutionärt bevarat mönster av interna disulfid-bindningar som visade sig formas av hydrofoba interaktioner tillsammans med proteinernas nedärvda ordning av cysteiner. I labbmiljö bildar dessutom gliadiner oligomerer vid peptidsektioner i C-terminaldelen via hydrofobiska interaktioner och hoplänkningar av disulfid-och lanthionin-bindningar. Vid N-terminaldelen kan oligomerer hopfogas genom vätskefas-separering, polyproline II-strukturer och β-flak. Värme och alkaliska förhållanden gynnade hoplänkning av lanthionin-, lysinoalanin- och disulfid-bindningar bland gliadiner och ökade deras absorbering av vätskor. Våra modeller beskriver glutenproteinerna α-gliadin och LMW-GS och deras mekanismer för att bilda interna och externa ihoplänkningar, hur de sammanfogas till oligomerer och hur de ökar sin vätskeabsorbering.

Nyckelord: gliadin, glutenin, super-absorbenter, LMW-GS, HMW-GS, protein modellering, peptider, protein vikning, lanthionin, lysinoalanin, disulphid

Author’s address: Joel Markgren, SLU Alnarp, Inst. för Växtförädling, P.O. Box 190, 234 22 Lomma, Sverige

“Love is like a pineapple, sweet and undefinable.”

– Piet Hein

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Faiza Rasheed, Joel Markgren, Mikael Hedenqvist and Eva Johansson. (2020). Modeling to understand plant protein structure-function relationships—implications for seed storage proteins. *Molecules* 25 (4), 873.

II. Joel Markgren, Mikael Hedenqvist, Faiza Rasheed, Marie Skepö and Eva Johansson. (2020). Glutenin and gliadin, a piece in the puzzle of their structural properties in the cell described through Monte Carlo simulations. *Biomolecules* 10 (8), 1095.

III. Joel Markgren, Faiza Rasheed, Mikael Hedenqvist, Marie Skepö and Eva Johansson. Assessing clustering and cross-linking reactions of the seed storage protein α-gliadin. A combined experimental and theoretical approach. Manuscript (submitted).

IV. Joel Markgren, Antonio Capezza, Mikael Hedenqvist, Marie Skepö and Eva Johansson. The underlying mechanisms for crosslinking and functionalization of gluten proteins. Manuscript.

The contribution of Joel Markgren to the papers included in this thesis was as follows:

I. Participated in planning and conceptualisation and wrote separate sub-sections, prepared and performed the simulations for Figure 1.

II. Participated in idea generation, conceptualisation and planning of the manuscript. Planned and performed the all-atom and coarse-grained simulations and hydrophobic analysis. Participated in analysis of the results, wrote the original draft and prepared all diagrams.

III. Participated in idea generation, conceptualisation and planning of the manuscript. Planned and performed coarse-grained simulations and experiments with gliadin peptides and gliadin protein mixtures. Participated in analysis of the results, wrote the original draft and prepared most of the diagrams.

IV. Participated in idea generation, conceptualisation and planning of the manuscript. Planned and performed the coarse-grained simulations and all experimental work except FTIR analyses. Participated in analysis of the results, wrote the original draft and prepared all diagrams.

Abbreviations

CYS	Cysteine
ER	Endoplasmic reticulum
FTIR	Fourier-transform infrared
GS	Glutenin subunit
HMW	High molecular weight
IDP	Intrinsically disordered protein
LLPS	Liquid-liquid phase separation
LMW	Low molecular weight
MD	Molecular dynamics
PBC	Periodic boundary condition
R_{ee}	End-to-end distance
R_g	Radius of gyration
RP-HPLC	Reverse phase-high performance liquid chromatography
SAP	Super-absorbent polymers
SAXS	Small angle X-ray scattering
SE-HPLC	Size exclusion-high performance liquid chromatography

1. Introduction

1.1 Gluten proteins - seed storage proteins with an ability to build extensive networks

Gluten proteins are storage proteins in wheat seeds, where they are stored as nutrients for the embryo when sprouting (Seilmeier et al. 1991; Krishnan & Coe 2001; Wieser 2007; Tan-Wilson & Wilson 2012). These proteins are packed in protein bodies either as polymeric aggregates or as monomeric proteins in the seed. Their polymeric properties make them valuable in human food products and they are studied for potential uses as petroleum-free material substitutes for plastics and super-absorbents (Kuktaite & Ravel 2020, Capezza et al. 2020). Most graminaceous species have proteins that are homologous to the gluten proteins in wheat, sharing comparable primary structure, but those in wheat are known to form the most extensive networks (Shewry et al. 1984; Xu & Messing 2009; Souza et al. 2015).

There are four different seed proteins: albumins, globulins, glutenins and gliadins (Osborne 1908; Shewry et al. 1995; Wieser 2007). In wheat seeds, the storage proteins are of two types, gliadins and glutenins (Shewry et al. 1995; Wieser 2007). The gliadins are soluble in alcohol, monomeric in that they only form internal disulphide bonds (do not cross-link with other proteins) and relatively hydrophobic. The gliadins can be further differentiated into three different types: α/β, γ and ω. The α/β, and γ gliadins resemble each other, with repeats of proline and with glutamine, cysteines (CYS) and hydrophobic amino acids at comparable positions (Urade et al. 2017) (Figure 1). The ω-gliadins differ from the other gliadins because they

lack CYS and consist mainly of repetitive amino acids with short hydrophobic sections at the terminals.

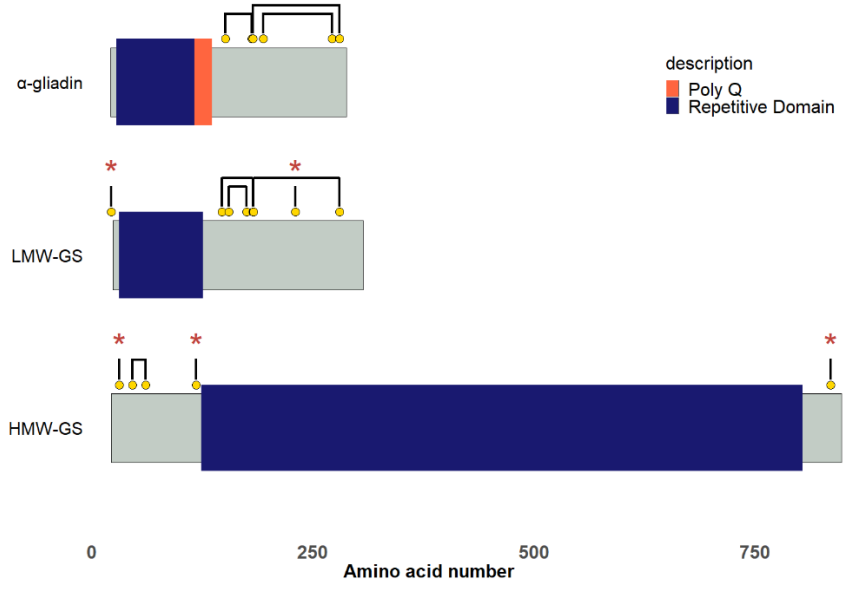


Figure 1. Schematic diagrams of α-gliadin, low molecular weight glutenin subunit (LMW-GS) and high molecular weight glutenin subunit (HMW-GS). Yellow dots indicate the position of cysteines and how they, in the seed, are connected (black lines) by disulphide bonds. Red stars indicate cysteines involved in inter-disulphide bonds. Poly Q: Polyglutamine

Glutenins are partly soluble in acid or alkali, larger than gliadins and form extensive networks by inter-disulphide bonds (Shewry & Tatham 1990). There are two subclasses of glutenin: high molecular weight glutenin subunits (HMW- GS) and low molecular weight glutenin subunits (LMW-GS). The HMW-GS are large molecules with extensive glutamine and glycine repeats and CYS positioned mainly at terminal parts (Anjum et al. 2007) (Figure 1). The LMW-GS resemble the gliadins in size, with CYS positioned at comparable conserved positions and proline and glutamine repeats (D'Ovidio & Masci 2004) (Figure 1).

Both gliadins and glutenins contain multiple sections of amino acids in repeating sequences, for example "PQPQPQP" (Shewry & Halford 2002) (Figure 1). Further, they often contain conserved sections (sequences that could be evolutionarily related), such as CYS placed in homologous positions (Shewry et al. 1995; José-Estanyol et al. 2004) and the presence of

blocks with hydrophobic amino acids (Reeck & Hedgcoth 1985; Henrissat et al. 1988; Urade et al. 2017; Wang et al. 2017). Neither of these two protein types has a stable protein structure and both are classified as intrinsically disordered (Rasheed et al. 2014; Roberts et al. 2015). The aim of this thesis work was to unravel the cross-linking and oligomerisation mechanisms among the gliadins and LMW-GS in wheat protein and how it affects absorption properties.

2. Background

2.1 A group of proteins with many functions

Wheat is one of the three major crops (together with maize and rice) produced around the globe for human consumption (*FAOSTAT* n.d.). It is used to make food products such as bread, pasta, pastries and noodles (Guy 1995). The preference for wheat is due to its relatively high yield, good adaption to a broad range of cultivation environments and convenience in grain storability and transportability (Shewry 2009; Wrigley 2016). The primary constituent of the wheat grain is starch (around 60%), while the protein concentration in the grain is 9-18% (Osborne et al. 1919). Approximately 80% of wheat grain proteins are seed storage proteins, often referred to as gluten proteins, which serve as an amino acid depot for the seed embryo. Grain gluten protein content and composition affect the performance of wheat in bread baking (Guy 1995; Shewry 2009). During kneading, wheat dough develops a network formed by the glutenins, where disulphide bonds link glutenins into a polymer (Lagrain et al. 2008, Weegels et al. 1996a). At the same time, the gliadins regulate dough viscosity by interacting and being included in the network (Fido et al. 1997; Belton 1999; Barak et al. 2014; Kuktaite et al. 2004). Together, the two types of proteins enable the dough to lift by elasticity and viscosity through coalescence of gas bubbles from the yeast (Ortolan & Steel 2017). During heating (baking), the proteins denature and solidify by cleaving and form new novel disulphide bonds, making the bread fluffy (Ortolan & Steel 2017).

Owing to their particular characteristics, gluten proteins can be a potential source of plastic-like materials when the proteins are mixed and heated under pressure. As during baking, heat and pressure can be applied to denature the proteins, which will cleave and form new disulphide bonds, making the proteins polymeric. When making a plastic-like material from gluten proteins, use of a plasticiser additive like glycerol can contribute to arrange the proteins into hexagonal structures that decrease in size with temperature (Rasheed et al. 2015; Kuktaite et al. 2016). Hence, an additive can affect the quality of gluten protein polymers by altering their internal arrangement and abilities to form cross-links (Kuktaite et al. 2011; Blomfeldt et al. 2012; Rombouts et al. 2013; Rasheed et al. 2014, 2015, 2018; Muneer et al. 2015; Andrade et al. 2018; Ceresino et al. 2020). Further, an additive can affect the material quality of the proteins by altering their conformational abilities (Kuktaite et al. 2011; Blomfeldt et al. 2012; Rombouts et al. 2013; Rasheed et al. 2014, 2015, 2018; Muneer et al. 2015; Andrade et al. 2018; Ceresino et al. 2020). However, gluten protein quality also depends on the underlying genetics and cultivation conditions for the wheat plant used for extraction, since the plant regulates the types and amounts of proteins found in the grain (Hussain et al. 2012, 2013; Johansson et al. 2013; Weegels el al. 1996b). Uncovering the molecular mechanisms of cross-linking and flocculation for gliadin and glutenins might enable better baking practices and the development of novel gluten materials.

Like super-absorbent polymers (SAP), gluten protein networks can hold significant amounts of liquid by forming a hydrogel (Capezza et al. 2019a). An increase in gluten absorption of up to 1000 times the protein's own weight can be achieved by adding the functionalisation agent ethylenediaminetetraacetic dianhydride (EDTAD) (Wu et al. 2016; Capezza et al. 2020b). Further increases in absorption can be achieved by enhancing cross-linking of the gluten proteins, enabling them to absorb non-polar fluids by capillary forces (Capezza et al. 2019b, 2020a). By exploring the cross-links in gluten protein hydrogels, it might be possible to refine their absorption properties. Functionalised gluten proteins could potentially serve as a future substitute for petroleum-based absorbents in hygiene products or polymeric products like petroleum-based plastics.

2.2 Gluten protein biology

The gliadins and LMW-GS are produced by ribosomes in the wheat cell's endoplasmic reticulum (Kim et al. 1988; Levanony et al. 1992; Rubin et al. 1992; Loussert et al. 2008; Tosi et al. 2009, 2011). The gliadins are produced as monomeric proteins, while the LMW-GS are polymers where both protein types have internal disulphide bonds, but only the LMW-GS have inter-disulphide bonds (Lombardi et al. 2012). Both protein types form protein bodies containing either one protein type or a mixture of gliadins and LMW-GS that are later transported in the cell (Rubin et al. 1992; Loussert et al. 2008). The protein bodies are transported to the vacuole using two different routes, passing through either the endoplasmic reticulum lumen or the Golgi organelle (Rubin et al. 1992; Arcalis et al. 2004; Loussert et al. 2008; Tosi et al. 2009). At the vacuole, the protein bodies can coalesce into larger bodies and are stored until seed germination (Tosi et al. 2011). However, the mechanisms of disulphide bond formation in gluten proteins and of formation of protein bodies are still not entirely understood. When producing the gluten proteins, the cell also produces folding assisting proteins like foldases and chaperons (Li et al. 1993; DuPont et al. 1998; Vitale & Ceriotti 2004; Vitale & Boston 2008). These folding assistances proteins might help the gluten proteins in forming cross-links, pack them into protein bodies and signal them for transportation.

2.3 Challenges in studying gluten proteins

In the wheat plant, the proteins are mixed up in protein bodies and difficult to separate by extraction due to their often similar size, polarity, morphology and aggregation (Rubin et al. 1992; Tosi et al. 2009; Arcalis et al. 2019). These issues in retrieving gluten protein samples of a pure type present challenges when studying the properties of the proteins.

Instead of harvesting gluten proteins from wheat, it is possible to express the desired protein genes in another organism. Some gluten proteins have been sequenced and successfully produced in organisms like *Xenopus oocytes*, tobacco and bacterial systems (Greene 1981; Altschuler & Galili 1994; Senger et al. 2005; Francin-Allami et al. 2013). However, several studies report challenges in producing significant protein yields, due to issues

with the host system (Bartels et al. 1985; Norrander et al. 1985; Galili 1989; Dowd & Bekes 2002).

Producing synthetic gluten peptide sections is an alternative to biological methods. It is possible to tailor-make synthetic proteins, but the degree of effort required will depend on the sequence difficulty (Kent 2009; 2017). Hypothetically, it might be possible to synthesise complete gluten proteins with current methods, but their size and hydrophobic, repetitive and self-assembling sequences are seen as "problematic" (Paradís-Bas et al. 2016; Mueller et al. 2020). No examples of synthetic gluten proteins are known to date, but peptides from gluten proteins have been used in some studies, often in the context of coeliac disease research (De Re et al. 2013).

2.4 Gluten proteins have an intrinsically disordered structure

Unlike structured proteins, intrinsically disordered proteins (IDPs) like the gluten proteins continuously fold into different shapes without reaching a potential energy minimum that stabilises the structure, as illustrated by the example in Figure 2 (Dill & Chan 1997; Uversky 2016).

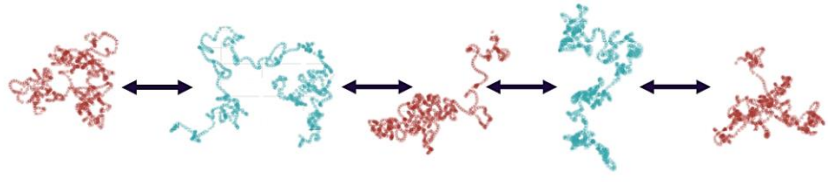


Figure 2. Schematic diagram of the intrinsic disorder of α-gliadin, where the protein chain constantly alters through different conformations.

A significant amount of research has been conducted on IDPs, due to their occurrence in several fatal and harsh human diseases like Alzheimer's, cancer and diabetes (Uversky et al. 2008; Scollo & La Rosa 2020). In the recent past, interest in IDPs in plant biology has increased and it has been shown that they may be more involved in several biological processes than previously known (Pazos et al. 2013; Sun et al. 2013). Previous studies on IDPs have focused on their physiochemical behaviours, including aggregational, structural and dynamic properties. Some of the properties of

IDPs depend on the amounts of charged amino acids present and chain length, and their behaviours can alter under changes in environmental factors such as crowding, temperature, pH, solvent, particle interactions and salinity (Henriques et al. 2015, 2018; Cragnell et al. 2016, 2018, 2019; Henriques & Skepö 2016; Fagerberg et al. 2019, 2020; Jephthah et al. 2019; Rieloff et al. 2019; Rieloff & Skepö 2021).

Intrinsically disordered proteins can have an average type of shape, depending on their amino acids and the surrounding environment. If an IDP carries a high amount of similarly charged residues, the protein becomes expanded or swollen by charge repulsion (Das & Pappu 2013; Das et al. 2015). If it has many charged residues of mixed type, the protein will be contracted. Overall, fewer charged residues will also lead to a compact or tadpole-like appearance (Das & Pappu 2013; Das et al. 2015). Altering the surrounding salinity can alter the structural behaviour of the proteins, since an increase in ion particles will scramble the electrostatics between charged amino acids (Wohl et al. 2021). In some instances, IDPs fold into a global energy minimum when interacting with specific ligands or other proteins (Arai 2018). However, gluten proteins are destabilised by their composition of several bulky, secondary structure-breaking and flexible amino acids, and they may find a stable fold only in outer space (Shewry et al. 1986; Aibara 1995; Theillet et al. 2013; Uversky 2015).

When observed by microscopy, glutenins and gliadins are mostly apparent as larger protein bodies (Shewry & Halford 2002; Arcalis et al. 2004; Tosi et al. 2009; Tosi 2012). In more detail, based on spectroscopic studies and simulations, HMW-GS are predicted to appear like elastic coils, due to their similar properties and similar amino acid composition to elastomeric proteins (Tatham & Shewry 2000; Parchment et al. 2001). Gliadins have been observed as elliptic structures in both small-angle X-ray scattering (SAXS) and electron microscope images (Shewry et al. 1997; Thomson et al. 1999; Sato et al. 2015; Herrera et al. 2018).

3. Scientific aims

This thesis examined the cross-linking, aggregating and oligomer-assembling properties of gluten proteins in terms of biological and water absorption aspects. Special attention was devoted to studying the α-gliadin protein, which serves as a model protein for gliadins.

Specific objectives of the work presented in this thesis were to:

- Review modelling and simulations among plant proteins with emphasis on seed storage proteins (Paper I).
- Describe the structure and foldings of gliadins and LMW-GS and their differences in forming polymeric disulphide cross-links in wheat cells (Paper II).
- Outline the mechanisms and properties of cross-linked aggregated (covalent binding) and clustered (weak forces) gliadins (Paper III).
- Assess the swelling mechanisms of gluten proteins due to cross-linking (Paper IV).

4. Computational Methods

4.1 Overview of computational methods

To describe the gluten proteins, molecular models of them were created and simulated with computational Monte Carlo algorithms. A comparison was made to validate the simulation results against available α-gliadin or LMW-GS descriptions found in literature sources or obtained in wet laboratory experiments in-house. This approach was chosen to provide valuable estimates of molecular details that could not be uncovered with available laboratory techniques. In addition, have previous simulation studies of gluten proteins had similar concepts where the simulations are thought to further explain laboratory or literature results (Masci et al. 1998; Arêas & Cassiano 2001; Cazalis et al. 2003; Yaşar et al. 2003; Herrera et al. 2018; Mioduszewski & Cieplak 2021).

4.2 Modelling

4.2.1 All-atom model

The proteins were explicitly modelled using the simulation and modelling software PROFASI, with all atoms represented when describing the gliadins and LMW-GS in the cell (Paper II), as illustrated in Figure 3. The force field "Lund FF08", which is designed to model and simulate disordered, pathogenic and aggregating Alzheimer proteins (Irbäck & Mohanty 2006; Irbäck et al. 2009), was used to regulate how the different modelled particles interacted. For this force field, all interactions are pairwise and include interactions for electrostatic, excluded volume, formation of hydrogen bonds and hydrophobicity (Irbäck & Mohanty 2006; Irbäck et al. 2009).

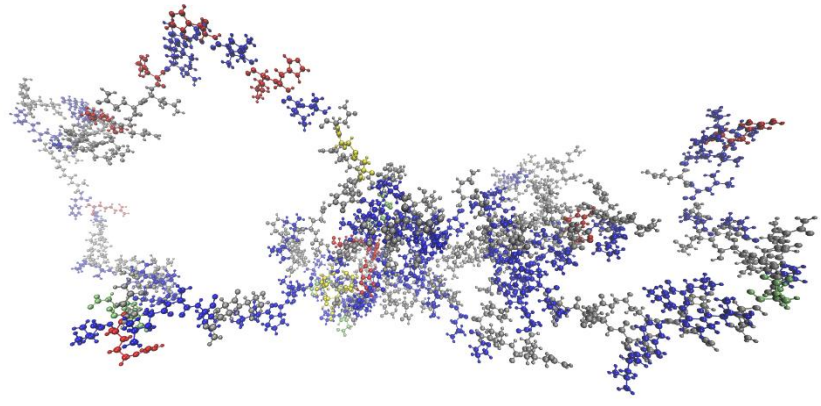


Figure 3. Screenshot of an all-atom model of α-gliadin simulated using the PROFASI software.

4.2.2 Coarse-grained model

Coarse-grained molecular models were used to describe the impact of protein/peptide salinity, oligomerising behaviours and swelling (Papers II, III, IV), where the term coarse-grained indicates simplification of the model compared to the all-atom form (Karimi-Varzaneh & Müller-Plathe 2012). Some of the ideas with simplification is to filter and capture essential protein features from the simulations. In the coarse-grained models used, the amino acids were represented by specific beads depending on their characteristics and the beads formed a necklace-like chain, as illustrated in Figure 4. Four different types of beads (neutral, positively charged, negatively charged, hydrophobic) were used to express the properties of the different amino acids. This molecular modelling approach was based on the extensive work performed previously on the disordered protein histatin (Cragnell et al. 2016, 2018; Rieloff et al. 2019), and used the MOLSIM simulation package (Jurij & Per 2015).

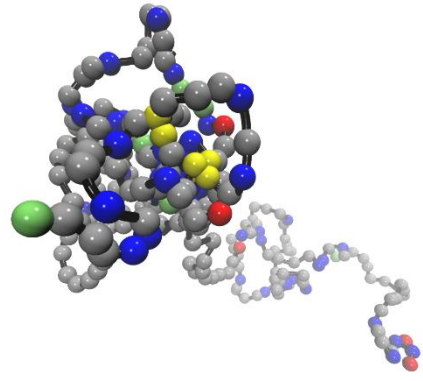


*Figure 4***.** Screenshot of a coarse-grained model of α-gliadin simulated using the MOLSIM software.

4.2.3 Coarse-grained model potentials

The interactions of the coarse-grained beads were described by potentials derived from previous work on histatin (Cragnell et al. 2016, 2018; Rieloff et al. 2019). The potentials were summarised pairwise and included bonded

interactions (the function of covalent bonds) and non-bonded interactions. Bonded interactions described the stretchiness and bond angles for covalent bonds. Non-bonded interactions described the impermeable surface of the beads, electrostatic interactions between charged groups and short-range attraction.

Amino acids that are considered hydrophobic according to the Kyte-Doolittle index (value higher than zero) were represented by a hydrophobic bead and regulated by a hydrophobic potential (Paper III & IV). This potential was described similarly to the short-range attraction.

4.2.4 Water and salinity

All molecular models were submerged in a hypothetical saline water solution implicitly described by the force field or potential. For the all-atom model a salinity of approximately 450-500 mM was used, while for the coarse-grained model a salinity of 150 mM was used. When testing the importance of charged groups for protein folding (Paper II), a range of salinities, corresponding to 10, 80, 500, or 1000 mM, were tested for the coarse-grained model.

4.3 Simulations

4.3.1 Monte Carlo algorithm

Monte Carlo and Metropolis algorithms were used to simulate potential conformations of proteins/peptides (Metropolis et al. 1953). These algorithms search for low-energy conformations according to the potential/force field used, as illustrated in Figure 5. This search is done by iteratively and randomly moving particles and thereby shifting the conformation or the relative position between multiple proteins/peptides. If the move leads to lower energy, it is accepted and sampled, and the search continues. On the other hand, if the move leads to higher energy, it is either rejected or accepted and sampled based on a probability criteria, and the search continues.

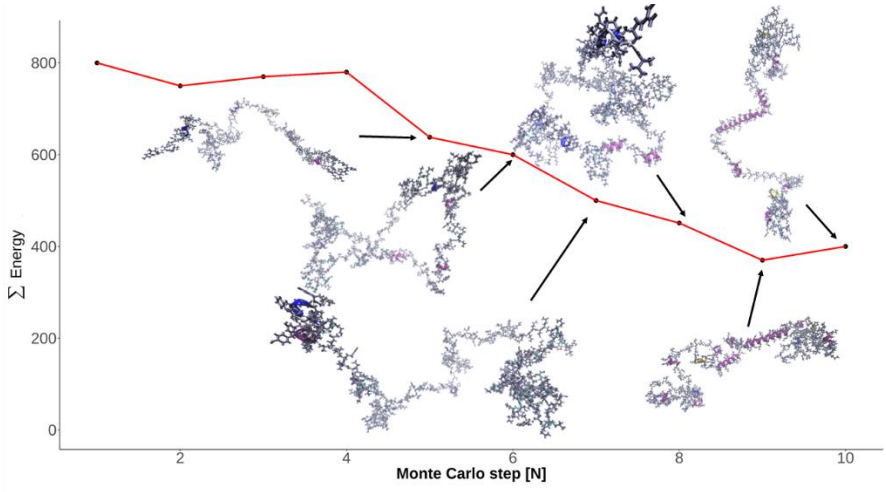


Figure 5. Energy values obtained in a simplified and demonstrative Monte Carlo simulation of α-gliadin and the protein conformations for each data point. The red line shows the randomly altered chain as the algorithm searches for an energy minimum.

4.3.2 Simulated annealing algorithm

Monte Carlo simulations can risk becoming stuck and continuously sampling a few closely related conformations, since moves resulting in higher energy may block the algorithm from accessing other low-energy conformations. To avoid this issue, the all-atom model simulations in Paper II also used a simulated annealing algorithm that altered the criterion determining whether moves that lead to a higher energy should be accepted or not. In a pre-set cycle, moves resulting in high energy are accepted, and the cycle also has a gradient for the energy difference for acceptance. From a conformational perspective, the algorithm enables the model to repeatedly open up and refold to a different type of fold (Kirkpatrick et al. 1983; Okamoto 1998).

4.3.3 Simulation parameters

All simulations were performed in virtual cubic boxes with canonical ensemble conditions (NVT), which involve a constant amount of particles (N), volume (V) and temperature (T). Periodic boundary conditions (PBC) at all orthological box sides were used to prevent the models interacting with

the box edge. The simulations were equilibrated before the production runs were executed. To verify that the sampling of the algorithms was performed correctly, the resulted probability distributions for the model's radius of gyration (R_g) were analysed.

4.3.4 Simulation analysis

Information regarding protein/peptide size, folding, shape, solvent behaviour, secondary structure and contacts was extracted from the simulation results. In all simulations, size was estimated through the model's R_g value. In Papers II, III & IV folding was estimated in the models through simulated SAXS analysis using wavelength functions (Schneidman-Duhovny et al. 2013, 2016; Cragnell et al. 2018). The shapes of the models in Paper II were estimated using an shape index based on R_g and end-to-end distance (R_{ee}) (Cragnell et al. 2018), and their solvent behaviours were calculated by Flory's power law (Flory 1949). In Paper III & IV, the shape was described by R_{ee} for peptide sections in the gliadin protein. An asperity index was calculated for protein and peptide aggregates in Paper III (Rieloff et al. 2019). An algorithm called "STRIDE" was used to predict the secondary structure for α-gliadin and LMW-GS in Paper II, based on hydrogen bonding and backbone angle data from known protein structures (Frishman & Argos 1995). To map contacts, a distance of 5Å or closer was sampled as contact between amino acids in Paper II and III.

4.4 Internet-based prediction services

In Paper III, two internet-based prediction tools, "aggrescan" and "PSPredictor", were used to predict the probability of protein aggregation and liquid-liquid phase separation (LLPS). The aggrescan tool is an algorithm that compares amino acid sequences to sequences that are empirically known to aggregate in aggregating β-amyloid peptides (Groot et al. 2005, 2006; Conchillo-Solé et al. 2007). From the comparison, the algorithm predicts potential aggregating sites in the sequence as so-called hot-spots. The PSPredictor tool uses a machine-learning algorithm trained with protein sequences known to phase-separate and a negative dataset with sequences known not to phase-separate (Sun et al. 2019).

5. Experimental Methods

5.1 Sample material

To assess the properties of gliadins (Papers III & IV), crude gluten powder, extracted gliadin mixtures, *E. coli*-produced fractions of α-gliadin and synthetic α-gliadin peptides were used. A wheat α-gliadin species was utilised as an amino acid sequence template for the *E. coli*-produced gliadins (Paper III), synthetic peptides (Paper III) and computational models (Papers II-IV) (Kasarda & D'Ovidio 1999; Senger et al. 2005). The sequence has been proven in actual gliadin samples (Kasarda & D'Ovidio 1999), and was used in this thesis work as a descriptive model for other gliadins.

Extracted fractions and crude gluten powder were used to describe glutenins (Paper IV). Only monomeric glutenins were used in the simulations (Paper II), with the sequence of a wheat LMW-GS that has been proven to exist (Colot et al. 1989; Hammond-Kosack et al. 1993).

5.2 Treatments

Wet laboratory experimental work on gluten proteins was conducted in Paper III to describe their aggregational and oligomerisation properties and in Paper IV to assess cross-linking in the context of enhanced liquid swelling. To determine the type of cross-links and clustering, potential hydrogen bonds or disulphide bonds in the samples in Papers III and IV, they were systematically tested with polar solvents, detergents, reduction and chaotropic agents, according to a protocol developed by Rasheed et al. 2014. A teabag test was used as described in previous swelling experiments by Capezza et al. 2020b to test protein absorption potential in Paper IV. This test compares the dry weight of a sample with its weight after being submerged in liquid for different periods of time.

5.3 Use of analytical methods

Several analytical instruments were used to characterise the gluten proteins, including different spectroscopy methods and high-performance liquid chromatography (HPLC). Information on size was obtained using size-exclusion HPLC (SE-HPLC) (Papers III & IV), where the particles are separated by a column (Hong et al. 2012), or dynamic light scattering (Paper III), where particle size is estimated based on how the sample scatters light (Stetefeld et al. 2016). Secondary structure prediction data in Paper III were retrieved using Fourier-transform infrared (FTIR) spectroscopy, which measures how a sample absorbs red light (Byler & Susi 1986), and circular dichroism, which measures absorption of circularly polarised light by a sample material (Greenfield 2006). Paper III and IV were also using reverse-phase HPLC (RP-HPLC), where a sample fractionates depending on its polarity in a column (Snyder et al. 2004).

6. Results and Discussion

6.1 The field of plant protein modelling

The review revealed that only a minor proportion (9%) of all structures deposited in the RCSB protein databank are related to plant proteins (Paper I), indicating that the field is still relatively new compared with structural protein research for other phylogenetic kingdoms. The structural data from individual protein deposits are frequently used as templates for various homology, docking and dynamic modelling studies. Most of the plant protein structures found in the protein databank are related in some way to one or more of the compounds in the following list (Paper I):

- Photosynthetic machinery
- Ribosomes
- Enzymes
- Stress/defence proteins
- Allergens
- Sweet-tasting proteins

There were no deposits of gluten proteins in the databank, except for minor peptides, probably due to the challenges in obtaining pure samples of gluten proteins and depicting them accurately. These limitations in terms of accessible structural data are likely a bottleneck in modelling of gluten proteins, leading to only a few pioneering studies being performed to date.

6.2 Characteristics of α-gliadin and LMW-GS based on modelling

With Monte-Carlo algorithms, it was possible to predict the size, folding, shape and secondary structures of the gluten proteins α-gliadin and LMW-GS. The results revealed noticeable similarities for all characteristics of α-gliadin and LMW-GS, whether with or without disulphide bonds (Paper II).

According to the simulations on folding and R_g data, the two proteins are of similar size, with an extended and mobile tail, a compact core and a ring-like motif (resembling a cup handle). The average shape of both proteins is elliptical and the dominant secondary structures are random coils and turns. Further, coarse-grained models of α-gliadin showed that the protein resembles a tadpole, with a compact core and a mobile tail (Paper III). Compared to previous modelling attempts and SAXS studies (Kasarda et al. 1968; Shewry et al. 1997; Thomson et al. 1999; Sato et al. 2015), the α-gliadin models gave similar average shape, size (R_g value 35-41 Å) and secondary structure. Most validation data for LMW-GS were lacking even after careful literature searches, most likely due to difficulties in obtaining purified monomeric protein samples. However, the size and polarities of LMW-GS should be similar to those of gliadins according to RP-HPLC and SEC-HPLC (Paper IV) (DuPont et al. 2004), and further similarities can be expected due to their comparable amino acid sequences.

6.3 The mechanism for internal cross-linking

The conserved internal disulphide bond motifs in gliadins and LMW-GS were shown by the models to be regulated by hydrophobic interactions (Paper II). In their inherent order, during protein synthesis CYS are joined together by the hydrophobic attraction of their peptide sections and form disulphide bonds that create the conserved motif (Paper II). Folding assistance administered by the cell, in the form of foldases and chaperons (Li et al. 1993; DuPont et al. 1998; Vitale & Ceriotti 2004; Vitale & Boston 2008), is likely not necessary for formation of these internal disulphide bonds, but might have other functions. Analysis of samples produced *ex situ* in *E. coli* systems confirmed that α-gliadins can form monomers without folding assistance from the wheat cell (Paper III).

6.4 Oligomer assembly

The peptides sections of α-gliadin have different mechanisms by which they assemble into oligomers. Those most prone to aggregating were found to be

strongly hydrophobic (Paper III), as shown in Figure 6. The hydrophobic peptide sections were shown to flocculate with identical counterparts and aggregate with disulphide bonds (Paper III).

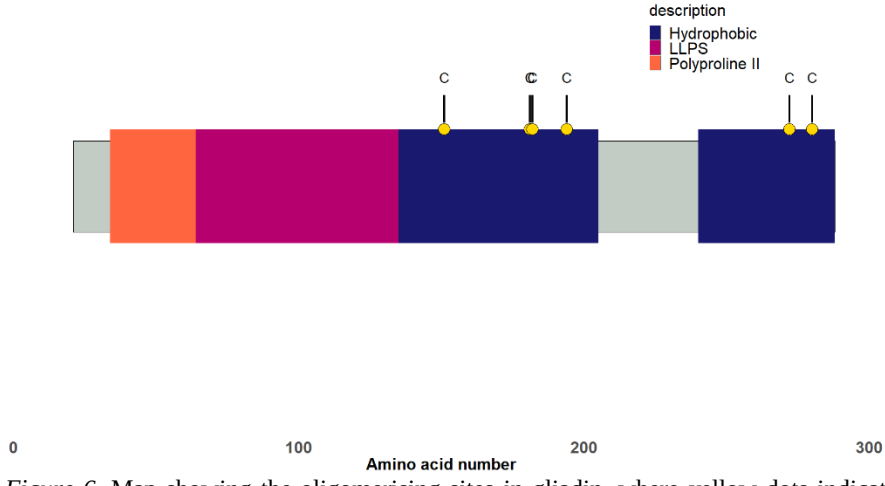


Figure 6. Map showing the oligomerising sites in gliadin, where yellow dots indicate cysteines. LLPS: liquid-liquid phase separation.

Among the polar peptides, it was predicted that the repetitive and poly-Q sequences would be prone to condense into a liquid through liquid-liquid phase separation (LLPS; Paper III) (Figure 6). The aromatic residues in these two sequences are likely to form a weak but attractive force between their π electrons that becomes stronger at higher concentrations, leading to LLPS (Gabryelczyk et al. 2019). Several previous studies have suggested that LLPS formation may be one mechanism by which the gliadins form protein bodies (Sahli et al. 2019; Banc et al. 2021), potentially explaining how ω-gliadins are present in these protein bodies despite lacking major hydrophobic peptide sections.

The peptide closest to the N-terminal was found to display an ability to form visual aggregates in water that are likely to be polyproline II structures (Paper III) (Figure 6). This specific peptide is known to form assembling β-sheets with increasing peptide concentration (Herrera et al. 2014; 2015). When dried, an increase in β-sheet structure was found for α-gliadins, but it is uncertain whether this is linked to the polyproline II structures (Paper III).

A tadpole model composed of a hydrophobic core and a mobile polar tail, as shown in Figure 7A, could be used for describing the oligomerisation of gliadins. The gliadins are attracted into a larger flocculant by the

hydrophobic cores, potentially forming a super core (Figure 7B**),** which is possibly cross-linked by disulphide bonds if the CYS are not already reduced. Interaction occurs between the flocculants at the gliadin tails (Figure 7C), which entangle and form polyproline II and β-sheet structures or condense the proteins into LLPS droplets at higher concentrations.

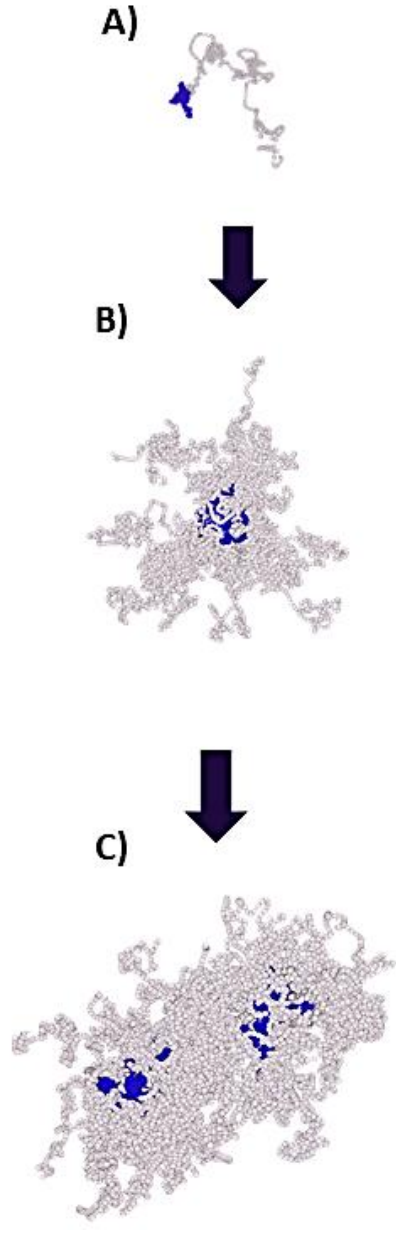


Figure 7. A) Gliadin displayed as a simplified tadpole model, where the blue beads are hydrophobic amino acids and the white beads are polar amino acids. B) Gliadins assemble first by their hydrophobic cores into a larger oligomer. C) Interactions with oligomers for the gliadin tails, which could induce entanglement, polyproline II structures, β-sheets or condensation into LLPS droplets.

6.5 Polymerisation by disulphide bonds

The results obtained in Paper II indicated that LMW-GS form inter-disulphide bonds with surrounding glutenins during or directly after protein synthesis, most likely due to the assembling properties of the peptide sections surrounding the participating CYS. During protein synthesis, the protein's first CYS interacts with that on surrounding LMW-GS through its placement at the end of the N-terminal, which functions as a tail. This tail, is likely similar to the tail of gliadins which entangle and attract other tails by polyproline II, β-sheet and LLPS formation (Papers II & III). The LMW-GS models in Paper II indicated that the first inter-disulphide bond occurs during or soon after protein synthesis, explaining why the literature describes the polymerisation process as rapid (Lombardi et al. 2012). The second inter-disulphide bond involves CYS 210, which is slower in cross-linking owing to its placement in a more rigid part of the protein (Paper II), and on occasions can reportedly form disulphide bonds with HMW-GS (Köhler et al. 1993; Keck et al. 1995; Orsi et al. 2001; D'Ovidio & Masci 2004; Lombardi et al. 2012). There is a preference for the LMW-GS to form disulphide bonds between corresponding cysteines, and the N-terminal cysteine is believed to be strongly involved in polymerisation with other LMW-GS (Köhler et al. 1993; Keck et al. 1995; Orsi et al. 2001; D'Ovidio & Masci 2004). Summarising the polymerisation of LMW-GS, it is indicated to be a step-wise process with a tail-to-tail direction between the two first monomers forming a dimer which builds a chain in a head-to-head direction (Weegels 1996, Weegels 1994).

6.6 Role of cross-linking in absorption by gluten proteins

Both gliadins and glutenins were found to form lysinoalanine (LAL) and lanthionine (LAN) cross-links at elevated pH (Paper IV). These forms of cross-link occur by chemical reactions between CYS, lysine and serine amino acids (Haraguchi et al. 1980; Denoël et al. 2018). For gliadins, increased liquid absorption was found to be correlated with increased LAL and LAN cross-linking. Generally, an increased amount of cross-linking is unfavourable for absorption, but gliadins are monomeric and, without the

cross-links, would most likely have poor absorption properties. In addition, studies on proteins suggest that increased cross-linking can lead to increased adsorption (Capezza et al. 2019b, 2020a). To fully understand the role of cross-links in the adsorption potential of gluten proteins, further measurements are needed on the ability of these proteins to resist being dissolved when submerged in a liquid.

7. Conclusions

A molecular model of gluten proteins (gliadins and LMW-GS) in the wheat seed cell was developed and used to show how these proteins form internal disulphide bonds, a process that is regulated by the inherent order of CYS and the hydrophobicity of their surrounding peptide sections. Further molecular modelling of the gliadins revealed their ability to assemble into oligomers through hydrophobic interactions, polyproline II structures, β-sheet formation and LLPS. It was found that LMW-GS form inter-disulphide bonds via CYS at polar peptide sections, where one CYS is located on a mobile end section of the protein and is rapidly polymerised. When exposed to alkaline environments, LAL and LAN cross-links were shown to be formed in both gliadins and LMW-GS, a development which was positively correlated with increased liquid absorption for the gliadins.

8. Future outlook

The Monte Carlo-based molecular models developed in this thesis provided valuable descriptions of the gluten proteins. Inclusion in the models of simulation modules describing the dynamic behaviours of these proteins could lead to further insights into *e.g*. their flexibility, strength, viscosity and reaction when altering the type or position of cross-links. Further, description of the dynamics could enable more detailed analysis of the oligomerisation process for gluten proteins. To describe the dynamic features of the proteins, molecular dynamic (MD) simulations in which Newton's law of motion is solved for all atoms would be needed (Hollingsworth & Dror 2018). Some early pioneering works have described the gliadin glutenin proteins by MD simulations, providing valuable information on their characteristics (Herrera et al. 2018; Mioduszewski & Cieplak 2021). However, in attempts with MD methods performed in this thesis (see Figure 8), it proved challenging to gather sufficient samples that adequately described protein behaviours. This indicates that MD simulations of the gluten proteins/peptides will require a vast amount of computational power, which would be difficult to achieve. However, this issue of a need for great computation power will hopefully be overcome in the near future with the ongoing developments in computer technology.

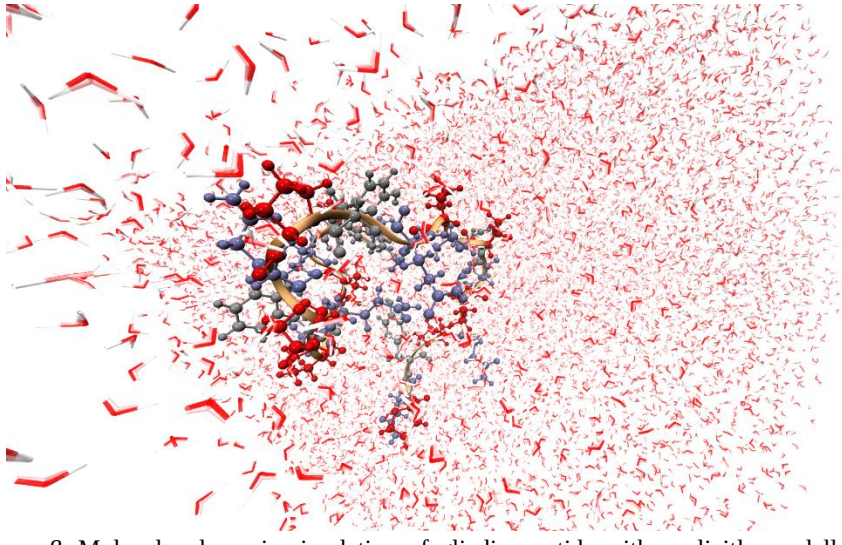


Figure 8. Molecular dynamic simulation of gliadin peptide with explicitly modelled water molecules.

References

Aibara, S. (1995). Crystallization of wheat γ-gliadin under a microgravity environment using space station MIR. *Journal of Crystal Growth*, 155, 247–253. https://doi.org/10.1016/0022-0248(95)00245-6

Altschuler, Y. & Galili, G. (1994). Role of conserved cysteines of a wheat gliadin in its transport and assembly into protein bodies in Xenopus oocytes. *Journal of Biological Chemistry*, 269 (9), 6677–6682

Andrade, D.F., Newson, W.R., Bernardinelli, O.D., Rasheed, F., Cobo, M.F., Plivelic, T.S., Ribeiro deAzevedo, E. & Kuktaite, R. (2018). An insight into molecular motions and phase composition of gliadin/glutenin glycerol blends studied by 13C solid-state and 1H time-domain NMR. *Journal of Polymer Science Part B: Polymer Physics*, 56 (9), 739–750. https://doi.org/10.1002/polb.24586

Anjum, F.M., Khan, M.R., Din, A., Saeed, M., Pasha, I. & Arshad, M.U. (2007). Wheat Gluten: High Molecular Weight Glutenin Subunits—Structure, Genetics, and Relation to Dough Elasticity. *Journal of Food Science*, 72 (3), R56–R63. https://doi.org/10.1111/j.1750-3841.2007.00292.x

Arai, M. (2018). Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophysical Reviews*, 10 (2), 163–181. https://doi.org/10.1007/s12551-017-0346-7

Arcalis, E., Ibl, V., Hilscher, J., Rademacher, T., Avesani, L., Morandini, F., Bortesi, L., Pezzotti, M., Vitale, A., Pum, D., De Meyer, T., Depicker, A. & Stoger, E. (2019). Russell-Like Bodies in Plant Seeds Share Common Features With Prolamin Bodies and Occur Upon Recombinant Protein Production. *Frontiers in Plant Science*, 10. https://doi.org/10.3389/fpls.2019.00777

Arcalis, E., Marcel, S., Altmann, F., Kolarich, D., Drakakaki, G., Fischer, R., Christou, P. & Stoger, E. (2004). Unexpected Deposition Patterns of Recombinant Proteins in Post-Endoplasmic Reticulum Compartments of Wheat Endosperm. *Plant Physiology*, 136 (3), 3457–3466. https://doi.org/10.1104/pp.104.050153

Arêas, E.P.G. & Cassiano, M.M. (2001). Folding interpenetration in a gliadin model: the role of the characteristic octapeptide motif. *Biophysical Chemistry*, 90 (2), 135–146. https://doi.org/10.1016/S0301-4622(01)00138-7

Banc, A., Navailles, L., Leng, J. & Renard, D. (2021). Dense Phases of γ-Gliadins in Confined Geometries. *Colloids and Interfaces*, 5 (4), 51. https://doi.org/10.3390/colloids5040051

Barak, S., Mudgil, D. & Khatkar, B.S. (2014). Influence of Gliadin and Glutenin Fractions on Rheological, Pasting, and Textural Properties of Dough. *International Journal of Food Properties*, 17 (7), 1428–1438. https://doi.org/10.1080/10942912.2012.717154

Bartels, D., Thompson, R.D. & Rothstein, S. (1985). Synthesis of a wheat storage protein subunit in Escherichia coli using novel expression vectors. *Gene*, 35 (1), 159–167. https://doi.org/10.1016/0378-1119(85)90168-4

Belton, P.S. (1999). Mini Review: On the Elasticity of Wheat Gluten. *Journal of Cereal Science*, 29 (2), 103–107. https://doi.org/10.1006/jcrs.1998.0227

Blomfeldt, T.O.J., Kuktaite, R., Plivelic, T.S., Rasheed, F., Johansson, E. & Hedenqvist, M.S. (2012). Novel freeze-dried foams from glutenin- and gliadin-rich fractions. *RSC Advances*, 2 (16), 6617–6627. https://doi.org/10.1039/C2RA20946B

Byler, D.M. & Susi, H. (1986). Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers*, 25 (3), 469–487. https://doi.org/10.1002/bip.360250307

Capezza, A.J., Cui, Y., Numata, K., Lundman, M., Newson, W.R., Olsson, R.T., Johansson, E. & Hedenqvist, M.S. (2020a). High Capacity Functionalized Protein Superabsorbents from an Agricultural Co-Product: A Cradle-to-Cradle Approach. *Advanced Sustainable Systems*, 4 (9), 2000110. https://doi.org/10.1002/adsu.202000110

Capezza, A.J., Lundman, M., Olsson, R.T., Newson, W.R., Hedenqvist, M.S. & Johansson, E. (2020b). Carboxylated Wheat Gluten Proteins: A Green Solution for Production of Sustainable Superabsorbent Materials. *Biomacromolecules*, 21 (5), 1709–1719. https://doi.org/10.1021/acs.biomac.9b01646

Capezza, A.J., Newson, W.R., Olsson, R.T., Hedenqvist, M.S. & Johansson, E. (2019a). Advances in the Use of Protein-Based Materials: Toward Sustainable Naturally Sourced Absorbent Materials. *ACS Sustainable Chemistry & Engineering*, 7 (5), 4532–4547. https://doi.org/10.1021/acssuschemeng.8b05400

Capezza, A.J., Wu, Q., Newson, W.R., Olsson, R.T., Espuche, E., Johansson, E. & Hedenqvist, M.S. (2019b). Superabsorbent and Fully Biobased Protein Foams with a Natural Cross-Linker and Cellulose Nanofibers. *ACS Omega*, 4 (19), 18257–18267. https://doi.org/10.1021/acsomega.9b02271

Cazalis, R., Aussenac, T., Rhazi, L., Marin, A. & Gibrat, J.-F. (2003). Homology modeling and molecular dynamics simulations of the N-terminal domain of wheat high molecular weight glutenin subunit 10. *Protein Science*, 12 (1), 34–43. https://doi.org/10.1110/ps.0229803

Ceresino, E.B., Kuktaite, R., Hedenqvist, M.S., Sato, H.H. & Johansson, E. (2020). Processing conditions and transglutaminase sources to "drive" the wheat gluten dough quality. *Innovative Food Science & Emerging Technologies*, 65, 102439. https://doi.org/10.1016/j.ifset.2020.102439

Colot, V., Bartels, D., Thompson, R. & Flavell, R. (1989). Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. *MGG Molecular & General Genetics*, 216 (1), 81–90. https://doi.org/10.1007/BF00332234

Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X. & Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*, 8 (1), 65. https://doi.org/10.1186/1471-2105-8-65

Cragnell, C., Durand, D., Cabane, B. & Skepö, M. (2016). Coarse-grained modeling of the intrinsically disordered protein Histatin 5 in solution: Monte Carlo simulations in combination with SAXS. *Proteins: Structure, Function, and Bioinformatics*, 84 (6), 777–791. https://doi.org/10.1002/prot.25025

Cragnell, C., Rieloff, E. & Skepö, M. (2018). Utilizing Coarse-Grained Modeling and Monte Carlo Simulations to Evaluate the Conformational Ensemble of Intrinsically Disordered Proteins and Regions. *Journal of Molecular Biology*, 430 (16), 2478–2492. https://doi.org/10.1016/j.jmb.2018.03.006

Cragnell, C., Staby, L., Lenton, S., Kragelund, B.B. & Skepö, M. (2019). Dynamical Oligomerisation of Histidine Rich Intrinsically Disordered Proteins Is Regulated through Zinc-Histidine Interactions. *Biomolecules*, 9 (5), 168. https://doi.org/10.3390/biom9050168

Das, R.K. & Pappu, R.V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences*, 110 (33), 13392–13397. https://doi.org/10.1073/pnas.1304749110

Das, R.K., Ruff, K.M. & Pappu, R.V. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins.

Current Opinion in Structural Biology, 32, 102–112. https://doi.org/10.1016/j.sbi.2015.03.008

De Re, V., Caggiari, L., Tabuso, M. & Cannizzaro, R. (2013). The versatile role of gliadin peptides in celiac disease. *Clinical Biochemistry*, 46 (6), 552–560. https://doi.org/10.1016/j.clinbiochem.2012.10.038

Denoël, T., Lemaire, C. & Luxen, A. (2018). Progress in Lanthionine and Protected Lanthionine Synthesis. *Chemistry – A European Journal*, 24 (58), 15421–15441. https://doi.org/10.1002/chem.201801115

Dill, K.A. & Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nature Structural & Molecular Biology*, 4 (1), 10–19. https://doi.org/10.1038/nsb0197-10

D'Ovidio, R. & Masci, S. (2004). The low-molecular-weight glutenin subunits of wheat gluten. *Journal of Cereal Science*, 39 (3), 321–339. https://doi.org/10.1016/j.jcs.2003.12.002

Dowd, C. & Bekes, F. (2002). Large-Scale Expression and Purification of High-Molecular-Weight Glutenin Subunits. *Protein Expression and Purification*, 25 (1), 97–104. https://doi.org/10.1006/prep.2001.1614

DuPont, F.M., Hurkman, W.J., Tanaka, C.K. & Chan, R. (1998). BiP, HSP70, NDK and PDI in wheat endosperm. I. Accumulation of mRNA and protein during grain development. *Physiologia Plantarum*, 103 (1), 70–79. https://doi.org/10.1034/j.1399-3054.1998.1030109.x

DuPont, F.M., Vensel, W., Encarnacao, T., Chan, R. & Kasarda, D.D. (2004). Similarities of omega gliadins from Triticum urartu to those encoded on chromosome 1A of hexaploid wheat and evidence for their post-translational processing. *Theoretical and Applied Genetics*, 108 (7), 1299–1308. https://doi.org/10.1007/s00122-003-1565-9

Fagerberg, E., Lenton, S. & Skepö, M. (2019). Evaluating Models of Varying Complexity of Crowded Intrinsically Disordered Protein Solutions Against SAXS. *Journal of Chemical Theory and Computation*, 15 (12), 6968–6983. https://doi.org/10.1021/acs.jctc.9b00723

Fagerberg, E., Månsson, L.K., Lenton, S. & Skepö, M. (2020). The Effects of Chain Length on the Structural Properties of Intrinsically Disordered Proteins in Concentrated Solutions. *The Journal of Physical Chemistry B*, 124 (52), 11843–11853. https://doi.org/10.1021/acs.jpcb.0c09635

FAOSTAT (n.d.). http://www.fao.org/faostat/en/#compare [2020-09-06]

Fido, R.J., Békés, F., Gras, P.W. & Tatham, A.S. (1997). Effects of α-, β-, γ- and ω-Gliadins on the Dough Mixing Properties of Wheat Flour.

Journal of Cereal Science, 26 (3), 271–277. https://doi.org/10.1006/jcrs.1997.0138

Flory, P.J. (1949). The Configuration of Real Polymer Chains. *The Journal of Chemical Physics*, 17 (3), 303–310. https://doi.org/10.1063/1.1747243

Francin-Allami, M., Bouder, A. & Popineau, Y. (2013). Comparative study of wheat low-molecular-weight glutenin and α-gliadin trafficking in tobacco cells. *Plant Cell Reports*, 32 (1), 89–101. https://doi.org/10.1007/s00299-012-1343-8

Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23 (4), 566–579. https://doi.org/10.1002/prot.340230412

Gabryelczyk, B., Cai, H., Shi, X., Sun, Y., Swinkels, P.J.M., Salentinig, S., Pervushin, K. & Miserez, A. (2019). Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides. *Nature Communications*, 10 (1), 5465. https://doi.org/10.1038/s41467-019-13469-8

Galili, G. (1989). Heterologous expression of a wheat high molecular weight glutenin gene in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 86 (20), 7756–7760

Greene, F.C. (1981). In Vitro Synthesis of Wheat (Triticum aestivum L.) Storage Proteins. *Plant Physiology*, 68 (3), 778–783. https://doi.org/10.1104/pp.68.3.778

Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1 (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202

Groot, N.S., Aviles, F.X., Vendrell, J. & Ventura, S. (2006). Mutagenesis of the central hydrophobic cluster in Aβ42 Alzheimer's peptide. *The FEBS Journal*, 273 (3), 658–668. https://doi.org/10.1111/j.1742-4658.2005.05102.x

Groot, N.S., Pallarés, I., Avilés, F.X., Vendrell, J. & Ventura, S. (2005). Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Structural Biology*, 5 (1), 18. https://doi.org/10.1186/1472-6807-5-18

Guy, R.C.E. (1995). Cereal processing: The baking of bread, cakes and pastries, and pasta production. In: Beckett, S.T. (ed.) *Physico-Chemical Aspects of Food Processing*. Boston, MA: Springer US, 258–275. https://doi.org/10.1007/978-1-4613-1227-7_12

Hammond-Kosack, M.C., Holdsworth, M.J. & Bevan, M.W. (1993). In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *The EMBO Journal*, 12 (2), 545–554

Haraguchi, T., Abe, K., Arai, S., Homma, S. & Fujimaki, M. (1980). Lysinoalanine Formation in Gluten: A Conformational Effect. *Agricultural and Biological Chemistry*, 44 (8), 1951–1952. https://doi.org/10.1080/00021369.1980.10864244

Henriques, J., Arleth, L., Lindorff-Larsen, K. & Skepö, M. (2018). On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *Journal of Molecular Biology*, 430 (16), 2521–2539. https://doi.org/10.1016/j.jmb.2018.03.002

Henriques, J., Cragnell, C. & Skepö, M. (2015). Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *Journal of Chemical Theory and Computation*, 11 (7), 3420–3431. https://doi.org/10.1021/ct501178z

Henriques, J. & Skepö, M. (2016). Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the TIP4P-D Water Model and the Representativeness of Protein Disorder Models. *Journal of Chemical Theory and Computation*, 12 (7), 3407–3415. https://doi.org/10.1021/acs.jctc.6b00429

Henrissat, B., Popineau, Y. & Kader, J.C. (1988). Hydrophobic-cluster analysis of plant protein sequences. A domain homology between storage and lipid-transfer proteins. *Biochemical Journal*, 255 (3), 901–905. https://doi.org/10.1042/bj2550901

Herrera, M.G., Vazquez, D.S., Sreij, R., Drechsler, M., Hertle, Y., Hellweg, T. & Dodero, V.I. (2018). Insights into gliadin supramolecular organization at digestive pH 3.0. *Colloids and Surfaces B: Biointerfaces*, 165, 363–370. https://doi.org/10.1016/j.colsurfb.2018.02.053

Herrera, G. M., A. Benedini, L., Lonez, C., L. Schilardi, P., Hellweg, T., Ruysschaert, J.-M. & I. Dodero, V. (2015). Self-assembly of 33-mer gliadin peptide oligomers. *Soft Matter*, 11 (44), 8648–8660. https://doi.org/10.1039/C5SM01619C

Herrera, M.G., Zamarreño, F., Costabel, M., Ritacco, H., Hütten, A., Sewald, N. & Dodero, V.I. (2014). Circular dichroism and electron microscopy studies in vitro of 33-mer gliadin peptide revealed secondary structure transition and supramolecular organization. *Biopolymers*, 101 (1), 96–106. https://doi.org/10.1002/bip.22288

Hollingsworth, S.A. & Dror, R.O. (2018). Molecular dynamics simulation for all. *Neuron*, 99 (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Hong, P., Koza, S. & Bouvier, E.S.P. (2012). A Review Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and Their Aggregates. *Journal of Liquid Chromatography & Related Technologies*, 35 (20), 2923–2950. https://doi.org/10.1080/10826076.2012.743724

Hussain, A., Larsson, H., Kuktaite, R., Prieto-Linde, M.L. & Johansson, E. (2012). Towards the understanding of bread-making quality in organically grown wheat: Dough mixing behaviour, protein polymerisation and structural properties. *Journal of Cereal Science*, 56 (3), 659–666. https://doi.org/10.1016/j.jcs.2012.07.016

Hussain, A., Larsson, H., Kuktaite, R., Prieto-Linde, M.L. & Johansson, E. (2013). Amount and Size Distribution of Monomeric and Polymeric Proteins in the Grain of Organically Produced Wheat. *Cereal Chemistry*, 90 (1), 80–86. https://doi.org/10.1094/CCHEM-05-12-0064-R

Irbäck, A., Mitternacht, S. & Mohanty, S. (2009). An effective all-atom potential for proteins. *PMC Biophysics*, 2 (1), 2. https://doi.org/10.1186/1757-5036-2-2

Irbäck, A. & Mohanty, S. (2006). PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *Journal of Computational Chemistry*, 27 (13), 1548–1555. https://doi.org/10.1002/jcc.20452

Jephthah, S., Staby, L., Kragelund, B.B. & Skepö, M. (2019). Temperature Dependence of Intrinsically Disordered Proteins in Simulations: What are We Missing? *Journal of Chemical Theory and Computation*, 15 (4), 2672–2683. https://doi.org/10.1021/acs.jctc.8b01281

Johansson, E., Malik, A.H., Hussain, A., Rasheed, F., Newson, W.R., Plivelic, T., Hedenqvist, M.S., Gällstedt, M. & Kuktaite, R. (2013). Wheat Gluten Polymer Structures: The Impact of Genotype, Environment, and Processing on Their Functionality in Various Applications. *Cereal Chemistry*, 90 (4), 367–376. https://doi.org/10.1094/CCHEM-08-12-0105-FI

José-Estanyol, M., Gomis-Rüth, F.X. & Puigdomènech, P. (2004). The eight-cysteine motif, a versatile structure in plant proteins. *Plant Physiology and Biochemistry*, 42 (5), 355–365. https://doi.org/10.1016/j.plaphy.2004.03.009

Jurij, R. & Per, L. (2015). MOLSIM: A modular molecular simulation software. *Journal of Computational Chemistry*, 36 (16), 1259–1274. https://doi.org/10.1002/jcc.23919

Karimi-Varzaneh, H.A. & Müller-Plathe, F. (2012). Coarse-Grained Modeling for Macromolecular Chemistry. In: Kirchner, B. & Vrabec, J. (eds.) *Multiscale Molecular Methods in Applied Chemistry*. Berlin, Heidelberg: Springer, 295–321. https://doi.org/10.1007/128_2010_122

Kasarda, D.D., Bernardin, J.E. & Gaffield, William. (1968). Circular dichroism and optical rotatory dispersion of .alpha.-gliadin. *Biochemistry*, 7 (11), 3950–3957. https://doi.org/10.1021/bi00851a023

Kasarda, D.D. & D'Ovidio, R. (1999). Deduced Amino Acid Sequence of an α-Gliadin Gene from Spelt Wheat (Spelta) Includes Sequences Active in Celiac Disease. *Cereal Chemistry*, 76 (4), 548–551. https://doi.org/10.1094/CCHEM.1999.76.4.548

Keck, B., Köhler, P. & Wieser, H. (1995). Disulphide bonds in wheat gluten: cystine peptides derived from gluten proteins following peptic and thermolytic digestion. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 200 (6), 432–439. https://doi.org/10.1007/BF01193253

Kent, S. (2017). Chemical protein synthesis: Inventing synthetic methods to decipher how proteins work. *Bioorganic & Medicinal Chemistry*, 25 (18), 4926–4937. https://doi.org/10.1016/j.bmc.2017.06.020

Kent, H. S.B. (2009). Total chemical synthesis of proteins. *Chemical Society Reviews*, 38 (2), 338–351. https://doi.org/10.1039/B700141J

Kim, W.T., Franceschi, V.R., Krishnan, H.B. & Okita, T.W. (1988). Formation of wheat protein bodies: Involvement of the Golgi apparatus in gliadin transport. *Planta*, 176 (2), 173–182. https://doi.org/10.1007/BF00392442

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983). Optimization by Simulated Annealing. *Science*, 220 (4598), 671–680. https://doi.org/10.1126/science.220.4598.671

Krishnan, H.B. & Coe, E.H. (2001). Seed Storage Proteins. In: Brenner, S. & Miller, J.H. (eds.) *Encyclopedia of Genetics*. New York: Academic Press, 1782–1787. https://doi.org/10.1006/rwgn.2001.1714

Kuktaite, R., Ravel, C. (2020). Wheat gluten protein structure and function: Is there anything new under the sun. *Wheat Quality for Improving Processing and Human Health, 1st ed.; Igrejas, G., Ikeda, T., Guzman, C., Eds*, 9-20

Kuktaite, R., Newson, W.R., Rasheed, F., Plivelic, T.S., Hedenqvist, M.S., Gällstedt, M. & Johansson, E. (2016). Monitoring Nanostructure Dynamics and Polymerization in Glycerol Plasticized Wheat Gliadin and Glutenin Films: Relation to Mechanical Properties. *ACS Sustainable Chemistry & Engineering*, 4 (6), 2998–3007. https://doi.org/10.1021/acssuschemeng.5b01667

Kuktaite, R., Plivelic, T.S., Cerenius, Y., Hedenqvist, M.S., Gällstedt, M., Marttila, S., Ignell, R., Popineau, Y., Tranquet, O., Shewry, P.R. & Johansson, E. (2011). Structure and Morphology of Wheat Gluten Films: From Polymeric Protein Aggregates toward Superstructure Arrangements. *Biomacromolecules*, 12 (5), 1438–1448. https://doi.org/10.1021/bm200009h

Kuktaite, R., Larsson, H., Johansson, E. (2004). Variation in protein composition of wheat flour and its relationship to dough mixing behavior. Journal of Cereal Science, 40 (1), 31-39. https://doi.org/10.1016/j.jcs.2004.04.007

Köhler, P., Belitz, H.-D. & Wieser, H. (1993). Disulphide bonds in wheat gluten: further cystine peptides from high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin and from ?-gliadins. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 196 (3), 239–247. https://doi.org/10.1007/BF01202740

Lagrain, B., Brijs, K. & Delcour, J.A. (2008). Reaction Kinetics of Gliadin−Glutenin Cross-Linking in Model Systems and in Bread Making. *Journal of Agricultural and Food Chemistry*, 56 (22), 10660–10666. https://doi.org/10.1021/jf801894r

Levanony, H., Rubin, R., Altschuler, Y. & Galili, G. (1992). Evidence for a novel route of wheat storage proteins to vacuoles. *The Journal of Cell Biology*, 119 (5), 1117–1128

Li, X., Wu, Y., Zhang, D.Z., Gillikin, J.W., Boston, R.S., Franceschi, V.R. & Okita, T.W. (1993). Rice prolamine protein body biogenesis: a BiP-mediated process. *Science*, 262 (5136), 1054–1056. https://doi.org/10.1126/science.8235623

Lombardi, A., Marshall, R.S., Castellazzi, C.L. & Ceriotti, A. (2012). Redox regulation of glutenin subunit assembly in the plant endoplasmic reticulum. *The Plant Journal*, 72 (6), 1015–1026. https://doi.org/10.1111/tpj.12020

Loussert, C., Popineau, Y. & Mangavel, C. (2008). Protein bodies ontogeny and localization of prolamin components in the developing endosperm of wheat caryopses. *Journal of Cereal Science*, 47 (3), 445–456. https://doi.org/10.1016/j.jcs.2007.05.012

Masci, S., D'Ovidio, R., Lafiandra, D. & Kasarda, D.D. (1998). Characterization of a Low-Molecular-Weight Glutenin Subunit Gene from Bread Wheat and the Corresponding Protein That Represents a Major Subunit of the Glutenin Polymer. *Plant Physiology*, 118 (4), 1147–1158

Mioduszewski, Ł. & Cieplak, M. (2021). Viscoelastic properties of wheat gluten in a molecular dynamics study. *PLoS computational biology*, 17 (3), e1008840. https://doi.org/10.1371/journal.pcbi.1008840

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21 (6), 1087–1092. https://doi.org/10.1063/1.1699114

Mueller, L.K., Baumruck, A.C., Zhdanova, H. & Tietze, A.A. (2020). Challenges and Perspectives in Chemical Synthesis of Highly Hydrophobic Peptides. *Frontiers in Bioengineering and Biotechnology*, 8, 162. https://doi.org/10.3389/fbioe.2020.00162

Muneer, F., Andersson, M., Koch, K., Menzel, C., Hedenqvist, M.S., Gällstedt, M., Plivelic, T.S. & Kuktaite, R. (2015). Nanostructural Morphology of Plasticized Wheat Gluten and Modified Potato Starch Composites: Relationship to Mechanical and Barrier Properties. *Biomacromolecules*, 16 (3), 695–705. https://doi.org/10.1021/bm5017496

Norrander, J.M., Vieira, J., Rubenstein, I. & Messing, J. (1985). Manipulation and expression of the maize zein storage proteins in Escherichia coli. *Journal of Biotechnology*, 2 (3), 157–175. https://doi.org/10.1016/0168-1656(85)90036-7

Okamoto, Y. (1998). *Protein Folding Problem as Studied by New Simulation Algorithms*.

Orsi, A., Sparvoli, F. & Ceriotti, A. (2001). Role of Individual Disulfide Bonds in the Structural Maturation of a Low Molecular Weight Glutenin Subunit. *Journal of Biological Chemistry*, 276 (34), 32322–32329. https://doi.org/10.1074/jbc.M103833200

Ortolan, F. & Steel, C.J. (2017). Protein Characteristics that Affect the Quality of Vital Wheat Gluten to be Used in Baking: A Review. *Comprehensive Reviews in Food Science and Food Safety*, 16 (3), 369–381. https://doi.org/10.1111/1541-4337.12259

Osborne, T.B. (1908). Our Present Knowledge of Plant Proteins. *Science*, 28 (718), 417–427

Osborne, T.B., Mendel, L.B. & Wakeman, W. the cooperation of E.L.F. and A.J. (1919). The Nutritive Value of the Wheat Kernel and Its Milling Products. *Journal of Biological Chemistry*, 37 (4), 557–601

Paradís-Bas, M., Tulla-Puche, J. & Albericio, F. (2016). The road to the synthesis of "difficult peptides." *Chemical Society Reviews*, 45 (3), 631–654. https://doi.org/10.1039/C5CS00680E

Parchment, O., Shewry, P.R., Tatham, A.S. & Osguthorpe, D.J. (2001). Molecular Modeling of Unusual Spiral Structure in Elastomeric Wheat Seed Protein. *Cereal Chemistry*, 78 (6), 658–662. https://doi.org/10.1094/CCHEM.2001.78.6.658

Pazos, F., Pietrosemoli, N., García-Martín, J. & Solano, R. (2013). Protein intrinsic disorder in plants. *Frontiers in Plant Science*, 4, 363. https://doi.org/10.3389/fpls.2013.00363

Rasheed, F., Newson, W.R., Plivelic, T.S., Kuktaite, R., Hedenqvist, M.S., Gällstedt, M. & Johansson, E. (2015). Macromolecular changes and nano-structural arrangements in gliadin and glutenin films upon chemical modification: Relation to functionality. *International Journal of Biological Macromolecules*, 79, 151–159. https://doi.org/10.1016/j.ijbiomac.2015.04.033

Rasheed, F., Plivelic, T.S., Kuktaite, R., Hedenqvist, M.S. & Johansson, E. (2018). Unraveling the Structural Puzzle of the Giant Glutenin Polymer—An Interplay between Protein Polymerization, Nanomorphology, and Functional Properties in Bioplastic Films. *ACS Omega*, 3 (5), 5584–5592. https://doi.org/10.1021/acsomega.7b02081

Rasheed, F., R. Newson, W., S. Plivelic, T., Kuktaite, R., S. Hedenqvist, M., Gällstedt, M. & Johansson, E. (2014). Structural architecture and solubility of native and modified gliadin and glutenin proteins: non-crystalline molecular and atomic organization. *RSC Advances*, 4 (4), 2051–2060. https://doi.org/10.1039/C3RA45522J

Reeck, G.R. & Hedgcoth, C. (1985). Amino acid sequence alignment of cereal storage proteins. *FEBS Letters*, 180 (2), 291–294. https://doi.org/10.1016/0014-5793(85)81088-7

Rieloff, E. & Skepö, M. (2021). The Effect of Multisite Phosphorylation on the Conformational Properties of Intrinsically Disordered Proteins. *International Journal of Molecular Sciences*, 22 (20), 11058. https://doi.org/10.3390/ijms222011058

Rieloff, E., Tully, M.D. & Skepö, M. (2019). Assessing the Intricate Balance of Intermolecular Interactions upon Self-Association of Intrinsically Disordered Proteins. *Journal of Molecular Biology*, 431 (3), 511–523. https://doi.org/10.1016/j.jmb.2018.11.027

Roberts, S., Dzuricky, M. & Chilkoti, A. (2015). Elastin-like polypeptides as models of intrinsically disordered proteins. *FEBS Letters*, 589

(19PartA), 2477–2486. https://doi.org/10.1016/j.febslet.2015.08.029

Rombouts, I., Lagrain, B., Delcour, J.A., Türe, H., Hedenqvist, M.S., Johansson, E. & Kuktaite, R. (2013). Crosslinks in wheat gluten films with hexagonal close-packed protein structures. *Industrial Crops and Products*, 51, 229–235. https://doi.org/10.1016/j.indcrop.2013.08.077

Rubin, R., Levanony, H. & Galili, G. (1992). Evidence for the Presence of Two Different Types of Protein Bodies in Wheat Endosperm. *Plant Physiology*, 99 (2), 718–724

Sahli, L., Renard, D., Solé-Jamault, V., Giuliani, A. & Boire, A. (2019). Role of protein conformation and weak interactions on γ-gliadin liquid-liquid phase separation. *Scientific Reports*, 9 (1), 1–13. https://doi.org/10.1038/s41598-019-49745-2

Sato, N., Matsumiya, A., Higashino, Y., Funaki, S., Kitao, Y., Oba, Y., Inoue, R., Arisaka, F., Sugiyama, M. & Urade, R. (2015). Molecular Assembly of Wheat Gliadins into Nanostructures: A Small-Angle X-ray Scattering Study of Gliadins in Distilled Water over a Wide Concentration Range. *Journal of Agricultural and Food Chemistry*, 63 (39), 8715–8721. https://doi.org/10.1021/acs.jafc.5b02902

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A. & Sali, A. (2013). Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophysical Journal*, 105 (4), 962–974. https://doi.org/10.1016/j.bpj.2013.07.020

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A. & Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Research*, 44 (W1), W424–W429. https://doi.org/10.1093/nar/gkw389

Scollo, F. & La Rosa, C. (2020). Amyloidogenic Intrinsically Disordered Proteins: New Insights into Their Self-Assembly and Their Interaction with Membranes. *Life*, 10 (8), 144. https://doi.org/10.3390/life10080144

Seilmeier, W., Belitz, H.-D. & Wieser, H. (1991). Separation and quantitative determination of high-molecular-weight subunits of glutenin from different wheat varieties and genetic variants of the variety Sicco. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 192 (2), 124–129. https://doi.org/10.1007/BF01202625

Senger, S., Maurano, F., Mazzeo, M.F., Gaita, M., Fierro, O., David, C.S., Troncone, R., Auricchio, S., Siciliano, R.A. & Rossi, M. (2005). Identification of Immunodominant Epitopes of α-Gliadin in HLA-

DQ8 Transgenic Mice following Oral Immunization. *The Journal of Immunology*, 175 (12), 8087–8095. https://doi.org/10.4049/jimmunol.175.12.8087

Shewry, P.R. (2009). Wheat. *Journal of Experimental Botany*, 60 (6), 1537–1553

Shewry, P.R. & Halford, N.G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. *Journal of Experimental Botany*, 53 (370), 947–958

Shewry, P.R., Miflin, B.J. & Kasarda, D.D. (1984). The Structural and Evolutionary Relationships of the Prolamin Storage Proteins of Barley, Rye and Wheat. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 304 (1120), 297–308

Shewry, P.R., Miles, M.J., Thomson, N.H. & Tatham, A.S. (1997). Scanning Probe Microscopes—Applications in Cereal Science. *Cereal Chemistry*, 74 (3), 193–199. https://doi.org/10.1094/CCHEM.1997.74.3.193

Shewry, P.R., Napier, J.A. & Tatham, A.S. (1995). Seed storage proteins: structures and biosynthesis. *The Plant Cell*, 7 (7), 945–956

Shewry, P.R. & Tatham, A.S. (1990). The prolamin storage proteins of cereal seeds: structure and evolution. *Biochemical Journal*, 267 (1), 1–12

Shewry, P.R., Tatham, A.S., Forde, J., Kreis, M. & Miflin, B.J. (1986). The classification and nomenclature of wheat gluten proteins: A reassessment. *Journal of Cereal Science*, 4 (2), 97–106. https://doi.org/10.1016/S0733-5210(86)80012-1

Snyder, L.R., Dolan, J.W. & Carr, P.W. (2004). The hydrophobic-subtraction model of reversed-phase column selectivity. *Journal of Chromatography A*, 1060 (1), 77–116. https://doi.org/10.1016/j.chroma.2004.08.121

Souza, C. de, Soares, R., Balbuena, T.S. & Arruda, P. (2015). Structure, Organization, and Expression of the Alpha Prolamin Multigenic Family Bring New Insights into the Evolutionary Relationships among Grasses. *The Plant Genome*, 8 (1). https://doi.org/10.3835/plantgenome2014.06.0027

Stetefeld, J., McKenna, S.A. & Patel, T.R. (2016). Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews*, 8 (4), 409–427. https://doi.org/10.1007/s12551-016-0218-6

Sun, T., Li, Q., Xu, Y., Zhang, Z., Lai, L. & Pei, J. (2019). Prediction of liquid-liquid phase separation proteins using machine learning. *bioRxiv*, 842336. https://doi.org/10.1101/842336

Sun, X., Rikkerink, E.H.A., Jones, W.T. & Uversky, V.N. (2013). Multifarious Roles of Intrinsic Disorder in Proteins Illustrate Its Broad Impact on Plant Biology. *The Plant Cell*, 25 (1), 38–55. https://doi.org/10.1105/tpc.112.106062

Tan-Wilson, A.L. & Wilson, K.A. (2012). Mobilization of seed protein reserves. *Physiologia Plantarum*, 145 (1), 140–153. https://doi.org/10.1111/j.1399-3054.2011.01535.x

Tatham, A.S. & Shewry, P.R. (2000). Elastomeric proteins: biological roles, structures and mechanisms. *Trends in Biochemical Sciences*, 25 (11), 567–571. https://doi.org/10.1016/S0968-0004(00)01670-4

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A.K., Daughdrill, G.W. & Uversky, V.N. (2013). The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins*, 1 (1). https://doi.org/10.4161/idp.24360

Thomson, N.H., Miles, M.J., Popineau, Y., Harries, J., Shewry, P. & Tatham, A.S. (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. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1430 (2), 359–366. https://doi.org/10.1016/S0167-4838(99)00019-9

Tosi, P. (2012). Trafficking and deposition of prolamins in wheat. *Journal of Cereal Science*, 56 (1), 81–90. https://doi.org/10.1016/j.jcs.2012.02.004

Tosi, P., Gritsch, C.S., He, J. & Shewry, P.R. (2011). Distribution of gluten proteins in bread wheat (Triticum aestivum) grain. *Annals of Botany*, 108 (1), 23–35. https://doi.org/10.1093/aob/mcr098

Tosi, P., Parker, M., Gritsch, C.S., Carzaniga, R., Martin, B. & Shewry, P.R. (2009). Trafficking of storage proteins in developing grain of wheat. *Journal of Experimental Botany*, 60 (3), 979–991. https://doi.org/10.1093/jxb/ern346

Urade, R., Sato, N. & Sugiyama, M. (2017). Gliadins from wheat grain: an overview, from primary structure to nanostructures of aggregates. *Biophysical Reviews*, 10 (2), 435–443. https://doi.org/10.1007/s12551-017-0367-2

Uversky, V.N. (2015). The intrinsic disorder alphabet. III. Dual personality of serine. *Intrinsically Disordered Proteins*, 3 (1), e1027032. https://doi.org/10.1080/21690707.2015.1027032

Uversky, V.N. (2016). Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *Journal of Biological Chemistry*, 291 (13), 6681–6688. https://doi.org/10.1074/jbc.R115.685859

Uversky, V.N., Oldfield, C.J. & Dunker, A.K. (2008). Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annual Review of Biophysics*, 37 (1), 215–246. https://doi.org/10.1146/annurev.biophys.37.032807.125924

Vitale, A. & Boston, R.S. (2008). Endoplasmic Reticulum Quality Control and the Unfolded Protein Response: Insights from Plants. *Traffic*, 9 (10), 1581–1588. https://doi.org/10.1111/j.1600-0854.2008.00780.x

Vitale, A. & Ceriotti, A. (2004). Protein Quality Control Mechanisms and Protein Storage in the Endoplasmic Reticulum. A Conflict of Interests? *Plant Physiology*, 136 (3), 3420–3426. https://doi.org/10.1104/pp.104.050351

Wang, J.J., Liu, G., Huang, Y.-B., Zeng, Q.-H., Hou, Y., Li, L., Ou, S., Zhang, M. & Hu, S.-Q. (2017). Dissecting the Disulfide Linkage of the N-Terminal Domain of HMW 1Dx5 and Its Contributions to Dough Functionality. *Journal of Agricultural and Food Chemistry*, 65 (30), 6264–6273. https://doi.org/10.1021/acs.jafc.7b02449

Weegels, P.L., van de Pijpekamp, A.M., Graveland, A., Hamer, R.J., Schofield, J.D. (1996a). Depolymerisation an Re-polymerisation of Wheat Glutenin During Dough Processing. I. Relationships between Glutenin Macropolymer Content and Quality Parameters. https://doi.org/10.1006/jcrs.1996.0010

Weegels, P.L., Hamer, R.J., Schofield, J.D. (1996b). Functional Properties of Wheat Glutenin. *Journal of Cereal Science*, 23 (1), 1-17. https://doi.org/10.1006/jcrs.1996.0001

Weegels, P.L. (1996). Theoretical considerations on the directionality in polymerization mechanism and polymer type. *Macromolecular Theory and Simulations*, 5, 299-303. https://doi.org/10.1002/mats.1996.040050210

Weegels, P.L. (1994). Depolymerisation and re-polymerisation of wheat glutenin during dough processing and effects of low M_r. wheat proteins [Doctoral dissertation, King's College London (University of London)]. ETHOS

Wieser, H. (2007). Chemistry of gluten proteins. *Food Microbiology*, 24 (2), 115–119. https://doi.org/10.1016/j.fm.2006.07.004

Wohl, S., Jakubowski, M. & Zheng, W. (2021). Salt-Dependent Conformational Changes of Intrinsically Disordered Proteins. *The Journal of Physical Chemistry Letters*, 12 (28), 6684–6691. https://doi.org/10.1021/acs.jpclett.1c01607

Wrigley, C.W. (2016). Wheat: An Overview of the Grain That Provides 'Our Daily Bread.' In: Wrigley, C., Corke, H., Seetharaman, K., & Faubion, J. (eds.) *Encyclopedia of Food Grains (Second Edition)*.

Oxford: Academic Press, 105–116. https://doi.org/10.1016/B978-0-12-394437-5.00020-6

Wu, Q., Yu, S., Kollert, M., Mtimet, M., Roth, S.V., Gedde, U.W., Johansson, E., Olsson, R.T. & Hedenqvist, M.S. (2016). Highly Absorbing Antimicrobial Biofoams Based on Wheat Gluten and Its Biohybrids. *ACS Sustainable Chemistry & Engineering*, 4 (4), 2395–2404. https://doi.org/10.1021/acssuschemeng.6b00099

Xu, J.-H. & Messing, J. (2009). Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. *Theoretical and Applied Genetics*, 119 (8), 1397. https://doi.org/10.1007/s00122-009-1143-x

Yaşar, F., Çelik, S. & Köksel, H. (2003). Molecular modeling of various peptide sequences of gliadins and low-molecular-weight glutenin subunits. *Food / Nahrung*, 47 (4), 238–242. https://doi.org/10.1002/food.200390056

Popular science summary

For a seed to germinate and become a new plant, it needs readily available nutrients. In wheat, the nutrients available to the embryo are stored in the form of protein bodies packed with seed storage proteins. The most common storage proteins in the wheat seed are called gluten proteins. These proteins form large polymeric networks in the seed and could possibly be a future substitute for petroleum-based plastic materials. These proteins are also elastic, a property which is important in baking, together with their ability to form polymers. The gluten proteins give elasticity and viscosity to the dough, so that it can expand during fermentation and form a fluffy bread when heated.

There are two main types of gluten proteins, gliadins and glutenins. These differ mainly in their ability to form cross-linking molecular bonds between cysteine amino acids. The gliadins, in their native state, form only internal cross-links, while the glutenins form both internal bonds and cross-links with surrounding glutenin proteins. How these cross-linking mechanisms work was previously unknown and is confusing, as several of the gliadin and glutenin proteins share a similar composition of amino acids. The aim of this thesis was to review previous plant protein modelling work and then to identify the mechanism by which gliadins and glutenins cross-link and assemble into oligomers and determine their absorptive properties.

Studying gluten proteins is challenging because they are entangled and difficult to separate from each other when extracted from the seed. Thus, they are difficult to image with high-resolution methods. They are also disordered and constantly changing shape. Only a limited number of previous studies have attempted to model gluten proteins, and indeed plant protein modelling in general is a relatively young field. However, using a combination of computer simulations and laboratory experiments on gluten proteins, the

work in this thesis revealed their cross-linking properties and how these properties affect the expansion in liquids. It was found that the gluten proteins contain hydrophobic parts that contract the protein chain in such a way that cysteines come close to each other and form internal cross-links. However, the glutenins contained cysteines that did not cross-link internally, due to specific locations.

The internal cross-linking was found to form quickly, taking place even before protein is completely produced in the cell. When the proteins are produced, they then assemble into larger gluten protein clusters, both through their hydrophobic sections and also through other mechanisms. These mechanisms are related to the protein's ability to form liquid protein droplets, specific protein structures and their entanglement. It is likely that, through these aggregating forces, the glutenins come close enough to cysteines in other glutenins to form cross-links that bind them together in a polymer.

Tests were also performed on how well cross-linked gluten proteins absorb liquids, as they could potentially be used as a replacement for petroleum-based super-absorbents in hygiene products such as sanitary pads or diapers. To ensure formation of cross-links in the proteins, they were treated with heat and a strong base, which formed other types of cross-links called lysinoalanines and lanthionines. The treatment increased the cross-linking between the gliadins and seemed to increase their absorption capacity to several times their own weight.

This thesis showed how gluten proteins are cross-linked and assembled into clusters/aggregates and determined their absorption potential. The new insights that these findings provide into the seed biology of storage proteins can be valuable in developing agronomic practises or plant-breeding methods aimed at producing high-quality flour. The results may also be valuable for future research into the development of non-petroleum-based materials and absorbents.

Populärvetenskaplig sammanfattning

Ett vetefrö är packat med näringsämnen för att embryot ska kunna gro och bli en ny planta. Bland dessa näringsämnen finns ett antal frölagringsproteiner som kallas för glutenproteiner och som bildar stora polymera nätverk i fröet. Dessa proteiner är viktiga vid bakning, då de ger elasticitet och viskositet till degen så att den kan expandera under jäsning och bilda ett fluffigt bröd vid upphettning.

Det finns främst två typer av glutenproteiner; "gliadiner" och "gluteniner". Dessa skiljer sig huvudsakligen åt i sin förmåga att bilda tvärbindande molekylbindningar mellan cystein- aminosyror. Båda glutenproteinerna bildar interna tvärbindningar, medan gluteninerna också bildar tvärbindningar med omgivande gluteninproteiner. Hur dessa tvärbindningsmekanismer fungerar har tidigare varit okänt och är intrikat, eftersom flera av gliadin- och gluteninproteinerna delar en liknande sammansättning av aminosyror. Denna avhandling avser att reda ut mekanismen för hur gliadinerna och gluteninerna tvärbinder, samlas ihop till oligomerer och absorberar vätska. Avhandlingen gör även en litterär genomgång av växtproteinmodellering.

Att studera glutenproteiner är utmanande, eftersom de är intrasslade och svåra att separera från varandra när de extraheras från fröet. Därmed är de också svåra att avbilda med hög upplösning. De är dessutom oordnade och ändrar konstant form. Därtill finns det enbart en begränsad mängd forskning som modellerat proteinerna, och växtproteinmodellering är i allmänhet ett relativt ungt område. Med en kombination av datorsimuleringar och laboratorieexperiment av glutenproteiner kunde vi dock beskriva deras tvärbindningsegenskaper och hur dessa egenskaper påverkar expansionen i vätskor. Vi fann att de undersökta glutenproteinerna innehåller hydrofoba delar som drar ihop proteinkedjan på sådana sätt att cysteiner kommer nära

varandra och bildar interna tvärbindningar. I gluteninerna fanns dock cysteiner som inte tvärband internt på grund av specifika placeringar.

Den interna tvärbindningen bildas snabbt och sker redan innan proteinet är helt producerat i cellen. När proteinerna sedan är producerade samlas de till större glutenproteininsamlingar, både genom sina hydrofoba sektioner, men även genom andra mekanismer. Dessa mekanismer är relaterade till förmågor att bilda flytande proteindroppar, specifika proteinstrukturer och att de trasslar in sig i varandra. Genom dessa mekanismer kommer gluteninerna tillräckligt nära cysteiner i andra gluteniner och bildar då tvärbindningar som länkar ihop dem till en polymer.

Vi testade också hur väl tvärbundna glutenproteiner absorberar vätskor, eftersom de potentiellt skulle kunna användas som en ersättning för petroleumbaserade superabsorbenter i hygienprodukter som t.ex. bindor eller blöjor. För att säkerställa att tvärbindningarna bildades i proteinerna behandlades de därför med värme och en stark bas, vilket bildade en annan typ av tvärbindningar som kallas lysinoalaniner och lantioniner. Behandlingen ökade tvärbindningarna mellan gliadinerna och verkade öka deras absorbtionsförmåga flera gånger sin egen vikt.

Denna avhandling beskriver hur glutenproteiner tvärbinds, hur de sätts samman till kluster/aggregat samt deras absorptionspotential. Dessa resultat ger nya insikter i fröbiologin hos lagringsproteiner som kan vara värdefulla för förädlings och odlingsmetodermetoder riktade för att producera mjöl av hög kvalitet. Dessutom kan resultaten vara värdefulla för framtida forskning om att utveckla icke-petroleumbaserade material och absorbenter.

Acknowledgements

The work described in this thesis would not have been possible without all the help and support I have gratefully received during my studies. Some of the key people are mentioned below with a few words.

Eva Johansson, my primary supervisor, was a great source of inspiration, knowledge and support. I am grateful for your supervision, which introduced and provided me with a further interest in science and gluten proteins.

Mikael Hedenqvist, my co-supervisor, was always just a phone call away for discussion and support regarding challenging scientific issues.

Faiza Rasheed, my co-supervisor, was helpful and supportive throughout my studies.

Marie Skepö, my co-supervisor, allowed me to come and study simulation techniques and work in her laboratory. Your kind input and ideas have been more than appreciated and I always found it interesting to discuss science with you.

Antonio Capezza, thank you for being my informal supervisor when working with super-absorbents. I highly value our scientific venture in combining our two fields of studies, where you were always available for passionate discussions with me.

Anna-Lovisa Nynäs, Emilia Berndtsson, Ida Lager, Magnus Carlsson, Per Snell, **Simon Jeppson**, **Yuzhou Lan,** thanks for all the fun discussions, motivational talks and readiness to help out when needed.

Fredrik Reslow, who provided highly valued social support, it was always fun to go to work because of you.

SACO at SLU, thanks for tireless trade union work, with a particular acknowledgment to **Andrus Kangro** for his mentorship and role modelling in advocating workers' rights.

Anders Ekholm, **Maria Luisa Prieto-Linde** and **Samuel Lenton,** thanks for helping me to become acquainted with the lab environment, especially **Anders** who like a mentor helped me with everything. Naturally, **William Newson** has to be mentioned and acknowledged too, for encouraging a jazzy and curious attitude in the lab.

Thanks to everyone who spent time at the **PhD council** with me in Alnarp, for all the fun we had - it was necessary to keep up my motivation during challenging moments in my studies

My thanks to my **family**, who provided love and support at all times. You always showed a sweet and appreciated interest in my studies and arranged parties for my stepwise success.

Last but not least, thanks to my beloved wife **Susanna Markgren,** who has taken care of me with love during all these years.

More formally, I would like to acknowledge Vetenskapsrådet (VR), Trees and Crops for the Future (TC4F) and the High Performance Computing Center North (HPC2N), which kindly provided funding and resources for this work.

I

molecules

MDPI

Review

Modeling to Understand Plant Protein Structure-Function Relationships—Implications for Seed Storage Proteins

Faiza Rasheed [1,2,†], Joel Markgren [1,†], Mikael Hedenqvist [2] and Eva Johansson [1,*]

[1] Department of Plant Breeding, The Swedish University of Agricultural Sciences, Box 101, SE-230 53 Alnarp, Sweden; Faiza.rasheed@slu.se (F.R.); Joel.markgren@slu.se (J.M.)
[2] School of Chemical Science and Engineering, Fibre and Polymer Technology, KTH Royal Institute of Technology, SE–100 44 Stockholm, Sweden; mikaelhe@kth.se
* Correspondence: Eva.johansson@slu.se
† These authors contributed equally to this work.

Received: 2 January 2020; Accepted: 14 February 2020; Published: 17 February 2020

Abstract: Proteins are among the most important molecules on Earth. Their structure and aggregation behavior are key to their functionality in living organisms and in protein-rich products. Innovations, such as increased computer size and power, together with novel simulation tools have improved our understanding of protein structure-function relationships. This review focuses on various proteins present in plants and modeling tools that can be applied to better understand protein structures and their relationship to functionality, with particular emphasis on plant storage proteins. Modeling of plant proteins is increasing, but less than 9% of deposits in the Research Collaboratory for Structural Bioinformatics Protein Data Bank come from plant proteins. Although, similar tools are applied as in other proteins, modeling of plant proteins is lagging behind and innovative methods are rarely used. Molecular dynamics and molecular docking are commonly used to evaluate differences in forms or mutants, and the impact on functionality. Modeling tools have also been used to describe the photosynthetic machinery and its electron transfer reactions. Storage proteins, especially in large and intrinsically disordered prolamins and glutelins, have been significantly less well-described using modeling. These proteins aggregate during processing and form large polymers that correlate with functionality. The resulting structure-function relationships are important for processed storage proteins, so modeling and simulation studies, using up-to-date models, algorithms, and computer tools are essential for obtaining a better understanding of these relationships.

Keywords: albumin; globulin; glutelin; monte carlo simulation; molecular dynamics simulation; prolamin

1. Introduction—Plant Proteins: Types, Characteristics, and Presence

1.1. The Variety of Plant Proteins and their Functions

Proteins are biological macromolecules, that are responsible for a majority of the biological roles (sometimes together with other biomolecules) in the cell. Thus, proteins can be considered biomolecular devices with natural structural and functional properties that are often challenging to recreate in the laboratory. Some classical examples of proteins with exceptional functions are spider silk proteins, combining very high strength with excellent elasticity [1], and bacterial flagella, which are microscopic propellers [2]. Proteins in the form of enzymes catalyze reactions at rapid rates with great selectivity.

Like proteins in general, plant proteins play various enzymatic, structural and functional roles (photosynthesis, biosynthesis, transport, immunity, etc). They also act as storage mediums to meet the growth and nutritional demands of developing seedlings. Proteins perform these functions in their

composition and specific structural forms, e.g., through folding, which can range from compact and well-ordered to unfolded and intrinsically disordered.

1.2. Classifications of Plant Proteins—Relationships to Seed Storage Proteins

The first attempts to classify plant proteins were based on the extractability and solubility of these proteins, with the first systematic study performed on seed storage proteins by T.B. Osborne, applying the classification scheme suggested by the American Committee on Protein Nomenclature [3]. This nomenclature basically classifies proteins into three types; simple, conjugated, and derived. Proteins in all plant tissues are classified as simple proteins, which are further divided into four types [4] (Table 1). These four types of plant proteins are mainly associated with seed storage proteins and are known as albumins, globulins, prolamins and glutelins, separated based on Osborne fractionation with water, salt, alcohol, and alkali, respectively (Table 1). Later attempts at more complex classifications of plant proteins have been made, based, e.g., on their chemical structure, mechanism/s of actions, biological function or within-plant location. Despite these attempts to more modern classification systems, Osborne classification is still the most widely used system, particularly used for protein extraction and purification procedures [5]. In practice, however, the Osborne classification has only been used on seed storage proteins, while classification of other plant proteins is generally more complex and sometimes unclear.

Table 1. Types, characteristics (based on Osborne fractionation [3,4]), and presence of plant proteins.

Protein Type	Solubility	Characteristics	Examples in Plants	References (examples)
Albumins	Water	Globular, coagulable by heat	2S-type; e.g., Leucine, Legumalin, Phaselin, Ricin	[4–6]
Globulins	Salt solution	Globular, higher molecular weight than albumins	7S vicilin-type (peas, soy etc.) 11S legumin-type (brassicas, oat, rice)	[4,7]
Prolamins	Alcohol/water mixtures (e.g., 70% ethanol)	Intra-molecular disulfide bonds, High proportion of proline and glutamine, repetitive motifs in central domains	Gliadins, Zein, Hordein, Secalin	[4,7–10]
Glutelins	Alkaline solutions	Inter-molecular disulfide bonds, High proportion of proline and glutamine, repetitive motifs in central domains	Glutenins in wheat	[4,7–11]

1.3. Characterization and Presence of Osborne Classified Proteins

Albumins are defined as water-soluble, globular proteins that are coagulable by heat (Table 1). The most well-known albumins are serum albumin, the major protein in human blood, and egg white. In plants, albumin is present as a 2S albumin storage protein in seeds, e.g., as leucine in barley, wheat, and rye, as legumelin in pea, soybean and cowpea, as phaselin in kidney bean, and as ricin in castor bean [6] (Table 1). Many proteins in the green plant tissues, including ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO), an enzyme catalyzing the first step of carbon fixation and the most abundant protein on Earth, are not defined as albumins, despite being water-soluble. RuBisCO is water-soluble and coagulable by heat [12], which should by the Osborne definition, make it an albumin protein. Similarly, the majority of the enzymatic proteins in plants are water-soluble and coagulable by heat, but apart from being characterized as enzymes, they have not been further defined into a protein type.

Globulins are also globular proteins, which have a higher molecular weight than albumins, and are soluble in dilute salt solution, but insoluble in water (Table 1). The most well-known globulin is a major human blood protein (serum gamma globulin) [13]. In plants, globulins are present as storage proteins in both, dicots and monocots, making them the most common group of storage proteins [7]. Based on sedimentation coefficient, the plant storage globulins are basically divided into two groups, 7S vicilin-type, which has been found and extensively studied in pea, soy bean etc., and 11S legumin-type, which is, e.g., the major storage protein in most legumes and in dicots such as brassica, oat, and rice [7] (Table 1). Some plant leaf proteins are clearly not directly water-soluble and may be classified as globulins, in some characterizations of leaf proteins [14]. However, the reason that these leaf proteins are not water-soluble might be because they are cell wall-bound, interact with pectin, or are hydrophobic [15], and thereby, obviously not easy to categorize as globulins or albumins.

The additional two protein types, prolamins and glutelins, are found in particular as storage proteins in seeds of the grass family (Triticeae), where they are the dominant proteins, comprising up to 85% of total protein [7] (Table 1). The prolamins found in wheat are called gliadins, while the nomenclature of the prolamins in other cereals is based on their Latin names; zein in maize, hordein in barley, secalin in rye etc. [8] (Table 1). The most commonly found glutelin is that found in wheat (glutenin), although glutelins are also present in barley and rye [11] (Table 1). Prolamins and glutelins have several similarities, including a high proportion of proline and glutamine, and a high proportion of repetitive motifs or sequences, with non-repetitive domains in their N- and C-terminals [8]. Although, these proteins differ in molar mass, the major difference between the two protein types is the formation of intra- and inter-molecular disulfide bonds in the native state of prolamins and glutelins, respectively [9], which explains the differences in their extractability.

1.4. Seed Storage Proteins—Types and Characteristics

From the previous sections, it is clear that seed storage proteins are present in the seed in various types, including albumin, globulin, prolamin and glutelin [6–8]. These proteins may have different characteristics in the plant cell and also during processing and for various applications, where the storage proteins contribute functionality. However, to date, the seed storage proteins have mainly been characterized based on their chemical performance during fractionation, sedimentation, etc. [4,6–8]

2. Functionality of Plant Proteins—In the Plant and for Food and Industrial Applications

2.1. Function of Plant Proteins in the Plant

Plant proteins meet the needs of the emerging seedling in terms of nutrition and growth, through their enzymatic, structural, functional and storage functions [16]. Plants contain a number of specific types of proteins not found in other living organisms, and these have certain functions (Table 2). For example, most plants have some kind of storage organ (seeds, tubers etc.) for reproduction, where different nutrient sources are stored so that the new plant will have the resources to grow during the coming season. Proteins, carbohydrates, and oils are different types of nutrient sources accumulated in plant storage organs. Such proteins are normally described as storage proteins (Table 2). Their primary function is to be broken down into amino acids, to form the necessary building blocks for emerging proteins in the next-generation growing plant [17]. The plant cell contains a number of organelles, including chloroplast, which are responsible for photosynthesis (Table 2). Plants also have a specific protein, the enzyme RuBisCO (Table 2), for catalyzing the transfer of solar energy to chemical energy that can be used by the plant, through CO_2 fixation [18].

Table 2. Examples of plant proteins with their unique functions in plants, and in food and non-food applications

Utilization	Protein Type (Examples)	Occurrence (Examples)	Function (Examples)	References (Examples)
Living plants	Storage proteins	Seeds, tubers	Growth and nutrition to seedlings and plantlets	[16,17]
	RuBisCo	Chloroplast	Photosynthesis	
	Plasma membrane proteins e.g., surface proteins, globular proteins	Cell membrane, Protein channels	Transport, structural support, ion regulation	[16,18]
Food products	Vicilin	Pea	Heat-induced gelation emulsifying properties	[19–22]
Bio-based materials	Glutenins	Wheat	Cohesive matrix, gas barrier, strength	[10,23–34]
	Gliadins	Wheat	Cohesive matrix, gas barrier, flexibility	

2.2. Functions of Plant Proteins in Food

In food products, proteins are often the main functional component, especially in processed food products with a high protein content [19]. In such products, proteins contribute to nutrition, but also to food quality, texture, aroma, flavor, feeling of satiety, and ease of processing [20] (Table 2). Most plant proteins have characteristics that make them interesting in food processing although their specific characteristics may differ. For example, the pea protein vicilin has been shown to have better heat-induced gelation and emulsifying properties than the pea legumins, which instead have been found to be more nutritious for humans [21]. The performance of various plant proteins for food applications is highly dependent on their structural features and their ability to form specific three-dimensional (3D) configurations/conformations, as well as their ability to cross-link [10,22].

2.3. Functions of Plant Proteins in Industrial Applications

Plant proteins also have properties that make them interesting for use as materials (Table 2). A range of plant proteins, including mainly storage proteins from e.g., wheat, soybean, potato and oilseed crops, have interesting properties for applications such as packaging, fire resistant and absorbent materials [23–30] (Table 2). The ability of proteins to cross-link and aggregate is important for good protein-based material properties [10,31–33].

2.4. Impact of Seed Storage Proteins on Functionality

The descriptions above clearly show the primary importance of the storage proteins in plants as a source of amino acids to be used as building blocks for the developing young seedling at emergence [17]. In food and industrial applications, seed storage proteins play a central role for functionality [19–30], making it important to understand how the functionality is influenced and can be fine-tuned [35–37].

3. Modeling—State of the Art for Plant Proteins

3.1. Basis for Modeling to Study Protein Structures

As described above, the three-dimensional (3D)-conformation of native proteins, together with their ability to crosslink, reshape, and form specific structural features, contribute substantially to the functional properties of specific proteins in the plant and in plant-products [31,34,38]. The methods for determining the structure of proteins can essentially be divided into three types, i.e., those based on different types of microscopy techniques, those based on scattering/diffraction and spectroscopy (e.g., X-ray and nuclear magnetic resonance (NMR) techniques), and those based on modeling [39] (Table 3).

In this review, we focus on the opportunities and drawbacks of using modeling to determine plant protein structures, with specific emphasis on plant storage proteins.

3.2. Template-Based and Ab Initio Modeling

The methods applied to model plant protein structures are similar to those used for any type of protein (Table 3). Template-based modeling is the most simplified method for protein modeling. Template-based models are built on comparisons of amino acid sequences from proteins with known protein structures (often identified by crystallography) and the assumption that similar sequences result in the same protein structures [40]. However, many proteins do not share sequence similarities with other proteins already present in data-bases and with known structures. For such proteins (showing no relationship with known proteins), modeling from sequence information alone, i.e., ab initio protein structure prediction, is the only option. In ab initio predictions, thermodynamic principles are applied to the case of protein conformation, through a search for the overall energy minimum [40].

3.3. Monte Carlo, Molecular Dynamics and Machine-Learning Methods

Two more advanced/modern methods, Monte Carlo (MC) and Molecular Dynamics (MD) have been developed during recent years for protein structure predictions and simulations. Both methods resemble the ab initio approach as modeling is based on sequence information alone. (Table 3). Both methods can be applied on different scales with different degrees of detail; from all-atom and united atom to coarse-grain models [41]. The advantage with MC computations is that it is faster and easier to perform than MD computation, due to the fact that it is free from the restriction of solving Newton's equation of motion [42] (Table 3). However, this also affects the results, since no "dynamic" information is gathered from the MC run [42,43]. Thus, in several protein simulations, MC has been combined with MD [44]. Furthermore, modern machine-learning techniques, such as the algorithm AlphaFold, have brought great advances in prediction of protein structure from sequence information [45]. In the most recent Critical Assessment of Protein Structure Prediction (CASP) experiments, the AlphaFold 3D models of proteins were placed first in the Free Modeling category, in terms of accuracy [46]. The AlphaFold algorithm uses artificial neural networks to build a protein-specific fragment library [47]. However, algorithms, such as AlphaFold have not yet been applied to predict the structure of seed storage proteins, probably because these algorithms generally predict protein structures, based on folding, while some of the most useful storage proteins are intrinsically disordered.

Table 3. Methods used to determine protein structures and structure-function relationships.

Method Type	Method (Examples)	Used for/Applications (Examples)	References (Examples)
Microscopy	Transmission Electron Microscopy (TEM)	Three-dimensional (3D) structure of proteins from 2D particle images	[39,48]
	Cryo Electron Microscopy (Cryo-EM)	3D structure of biomacro-molecules in native state	[39,49]
	Tomography	High-resolution 3D images	[50]
	Imaging	Images of individual proteins by low-electron holography	[50,51]
Scattering/ Diffraction, Spectroscopy	Nuclear magnetic resonance Spectroscopy (NMR)	Chemical shifts reflecting conformations of proteins	[52]
	Small-angle X-ray scattering	Shape, conformation and assembly of proteins	[53]
	Wide-angel X-ray scattering	Characterization of structural models, similarities, and changes in atomic packing	[54]
	Fourier Transform Infrared Spectroscopy (FT-IR)	Protein conformation through peak fitting of amide bands	[39,55]
	X-ray crystallography	Atomic resolution of 3D protein structures	[56]
Modeling-Simulation	Template-based	Modeling based on homology	[34]
	Ab initio	Modeling based only on sequence information	[40,41]
	Monte Carlo (MC)	Statistical method evaluating moves of a protein	[42,43]
	Molecular Dynamics (MD)	Solving Newton's equation of motion	[43]
	Machine aearning (AlphaFold)	Artificial neural network	[45–47]

3.4. Modeling Plant Proteins with Specific Emphasis on Seed Storage Proteins

The plant protein modeling research field is comparatively small and relatively new, with a limited number of plant-related deposits in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). A search for plant deposits currently results in 2300 hits, with one quarter from Arabidopsis. Sorting, based on the most common organisms, shows 1382 hits on Arabidopsis and 15000 hits on others (others = plant based organisms but also additional organisms) in an overall total of more than 160 000 total deposits (the majority coming from humans and microorganisms) [57]. One third of the deposits in the current plant search are from 2015 and later. Only 47 deposits are available for seed storage proteins, the majority being globulins and the rest albumins. These deposits indicate that plant protein modeling has focused on specific areas. This is also reflected in by the information in the RCSB PDB, where a high number of deposits are available for example, in proteins associated with the photosynthetic machinery, proteins associated with ribosomes, enzymes, stress and defense, and for allergens and sweet-tasting proteins [57]. Plant protein modeling, within these different areas, is summarized in Sections 4.1–4.6, in order to identify opportunities of relevance for modeling seed storage proteins.

4. Main Plant Protein Modeling Areas and Impact for Modeling of Seed Storage Proteins

Modeling approaches for six of the most common plant proteins in the RCSB PDF are described below. Seed storage proteins are not among these, although knowledge derived from other areas of research might be of relevance for modeling seed storage proteins.

4.1. Photosynthetic Machinery

The photosynthetic machinery is responsible for the transfer of solar energy to chemical energy through CO_2 fixation [18]. It is, thus, one of the most important and specific traits of plants, and has long attracted the interest of the scientific community. MD simulations, complemented with quantum mechanical descriptions, were applied already in the early 1990s to describe how electron transfer is controlled by protein motion in photosynthetic reaction centers [58]. Molecular dynamics simulations have since contributed to the understanding that protein movement in the reaction centers is key to the kinetics of the primary electron transfer reaction [59]. To understand the full mechanisms of the photosynthetic machinery, a range of both, analytical and simulation methods have been adopted, as recently reviewed by Blumberger [60,61]. The tools for assessing protein movements in reaction centers might be useful for investigating seed storage protein movements during processing and dynamics/energy transfers for these movements. The structure of RuBisCO, and of divergent, mutant, and hybrid forms of this enzyme, has been characterized through X-ray crystallography [62]. Molecular dynamics simulations have also been used to explain variations in the functionality of RuBisCO mutants, where the use of structural checkpoints has been found to enable fine-tuning of the dynamics of the enzyme [63]. Similarly, dynamics of seed storage proteins already characterized by experimental methods such as X-ray crystallography can be understood by computer-based simulation methods.

4.2. Proteins Associated with Ribosomes

Cryo-EM studies, combined with X-ray crystallography, have been applied to determine the structures of ribosomes, producing static models for the various states of ribosomes. Recently, computational studies involving simulations have been successful in shedding light on structural fluctuations and transitions among the different ribosomal configurations [64]. Advances in MD simulations, including large-scale MD, are one explanation for the successful simulation of large macromolecular complexes, such as ribosomes [65]. The size of some seed storage proteins creates challenges in their modeling, resulting in similarities with the modeling of ribosomes. Thus, the use of large-scale MD might be an alternative for modeling seed storage proteins.

4.3. Enzymes

Mathematical modeling approaches have long been a useful tool for investigating the complexity of metabolic networks and their enzymatic regulation, while more recent models have contributed greatly to the growing field of systems biology [66]. Most commonly, enzyme-kinetic models have been applied to examine enzymatic regulations [66], although more recently, MD techniques and molecular docking simulations have been used for similar purposes [67]. The enzyme-kinetic modeling of metabolic pathways differs substantially from the modeling of structures and functions in seed storage proteins.

4.4. Stress and Defense

Regarding stress proteins and defense mechanisms, simulations have been used to gain a better understanding of the mechanisms involved [68–71]. Again, recent uses of MD simulations has improved our understanding of these proteins. Both homology modeling and MD simulations have been used to examine the background to herbicide resistance in plants [72]. Homology modeling, molecular docking, and MD simulations have been used to assess differences in protein conformations contributing to resistance compared with susceptible reactions in plants to different diseases [73]. The development of the highest relevance for modeling seed storage proteins is methodology describing differences in protein conformation related to different functions and functionality.

4.5. Allergens

Allergens are often proteins [74] and plants contain a variety of allergens towards which sensitive humans display allergic reactions [75]. Plant proteins responsible for allergic reactions have been structurally modeled using crystallography, X-ray scattering and NMR, as well as docking simulations of protein models with similar sequence [76,77]. Further, the 3D crystal structure of various plant-derived allergy proteins has been determined and MD simulations have been used to detect molecular conformations of the proteins involved in allergic reactions [78–80]. Several plant-based allergens are also seed storage proteins in plants, so modeling carried out on allergens might be directly transferable to research on seed storage proteins and structure-function relationships.

4.6. Sweet-Tasting Proteins

Sweet-tasting proteins are specific plant-based proteins of great interest as they have characteristics making them hundreds to thousand times sweeter in taste than sugar [81,82]. Several studies have focused on investigating this property, using techniques, such as crystallization of the proteins and determination of their structure with NMR and X-ray crystallography [81,83]. Comparative/homology modeling and molecular docking techniques have been used to predict 3D structures of dimer and tetramer forms of some sweet proteins, while the effects of pH on protein conformation have been evaluated using MD simulations [82]. MD simulations have also been used as a tool in structure-guided protein engineering for designing improved low-calorie plant-based sweeteners for pharmaceutical and food applications [84]. The methods applied to understand structure-function behavior and those used for structure-guided engineering might be of interest for seed storage protein research and applications.

5. Understanding Structure-Function Relationships of Seed Storage Proteins

Plant storage proteins are probably the second most abundant protein group in plants (after RuBisCO). In all plants containing storage organs, such as grains and seeds, the function of the storage proteins is to store amino acids necessary for growth and development of the emerging seedling [85]. Few investigations with modeling tools have been performed on these proteins, despite the abundance of plant storage proteins, their importance as a source of nutrition for the emerging young plantlet [85], and their impact on the functionality of products from plant grains [10,22,23,86–89]. However, bioinformatics techniques are increasingly being employed for classification of different plant proteins, with neural networks displaying accuracy of 95.3% in classifying rice proteins into different

classes (albumins, globulins, prolamins, glutelins) [90]. Additionally, machine-learning algorithms have been successfully used to classify seed storage proteins from rice, wheat, maize, castor bean, and thale cress into their classes [91]. Structurally, the different classes of seed storage proteins can be divided into two types, albumins/globulins and prolamins/glutelins (Figure 1). Albumins and globulins are generally highly structured and thereby able to crystallize, and their folding can be simulated using a range of methods, including machine learning and ab initio modeling [92–98] (Table 4). Most prolamins and glutelins are instead intrinsically disordered [99], and thereby, pose more challenges in modeling. They would require MC- and MD-based algorithms for modeling their structures, although small-angle scattering methods combined with infrared spectroscopy (IR) and high-performance liquid chromatography (HPLC) have also been applied to examine structural changes during processing [100–104].

Table 4. Structural characteristics of some modeled seed storage proteins.

Protein Type		Structure Prediction Method	Experimental Form	Structure Characteristics	References (Examples)
Globulin	Canavalin	X-ray diffraction	Native	Compact crystal structure with salt bridges and hydrophobic clusters	[93–96]
	Vicilin	Homology modeling	Amino acid sequence	Trimer	[97]
	Cruciferin	Homology modeling	Amino acid sequence	Hexamer via inter-protomer (IE) disulfide bonds between two trimers	[98]
Prolamin	Gliadins	Dynamic light scattering, cryo-transmission electron microscopy, small-angle X-ray scattering, MD simulations	Gliadin solution at pH 3.0	Dimers of 5.72 nm, aggregated clusters of 30 nm, oligomers of 68 and 103 nm	[103,105]
Glutelin	Glutenins	Small-angle-X-ray scattering	Native/extracted from wheat seed	Intrinsically disordered structure	[34,104]

5.1. Albumins

Crystal structures have been identified for a few albumin storage proteins using NMR and X-ray crystallography [92], and two of these protein structures have been deposited in the RCSB PDB [57]. Limited information is available on structural modeling using simulation tools for grain storage albumin proteins.

5.2. Globulins

Practically all deposits in the RCSB PDB on plant globulins are based on X-ray diffraction. The fact that X-ray diffraction-based models are available, enables simulation-based verification of the protein structures and simplifies further computer-based modeling. Globulins from legume seeds were the first storage protein to be crystallized and evaluated with X-ray diffraction [93–96]. The crystal structure was found to be compact with salt bridges and hydrophobic clusters, resulting in layers of packed molecules forming aggregates [94]. Since then, homology modeling has been used to model other globulin proteins, e.g., vicilin in cocoa, based on crystal structures of legume globulins such as jack bean canavalin and French bean phaseolin [97] (Table 4). These studies indicate that hydrophobic amino acids are buried inside the protein molecule at trimer formation, while histidine residues are found at the interfaces towards other globulins [97]. Later studies used homology modeling of *Arabidopsis thaliana* to identify the structure of the oilseed storage protein cruciferin [98] (Table 4). The impact of structure on the function of cruciferin has been evaluated through the use of different isoforms of the protein [98]. Recent studies using 3D molecular models and computational simulations have demonstrated ability of vicilin-like proteins from leguminous plants to bind to chitin or chitinous structures through three chitin-binding sites at each vicilin

trimer [106]. Examples of structural features of the storage protein 7S globulin in soybean and 11S globulin in pea, obtained through simulation based on amino acid sequences [39,107,108] of single subunits and their polymers, are shown in Figure 1.

5.3. Prolamins

Prolamins and glutelins have been less studied by modeling and simulation than the globulins and most other types of proteins. The major reason for the lack of models on prolamin proteins is that their structure is intrinsically disordered [99]. Solubility studies, using various solvents, have shown that the prolamins are monomeric in their natural stage [4,10] (Figure 1). However, during processing, the prolamins have been shown to form disulfide bonds with other seed storage proteins, thereby, contributing to the formation of polymers [10,109]. Early studies evaluating the structural features of the prolamins in wheat (the gliadins), using Fourier transform infrared spectroscopy (FT-IR), detected equal amounts of α-helix, linear β-structure, β-turns, and unordered structure [110]. The structures of the gliadins have been evaluated at pH 3.0 with dynamic light scattering, cryo-transmission electron microscopy, and small-angle X-ray scattering, followed by ab initio prediction and MD simulations [105] (Table 4). These studies indicate the presence of dimers with a hydrodynamic radius of 5.72 nm, aggregated clusters of 30 nm, and oligomers of 68 and 103 nm [105]. Importance of the concentration of gliadins in distilled water for their state of aggregation has also been reported [103]. At 0.5 wt-% gliadin, repulsion of the gliadin assemblies can be observed, resulting in the protein mainly being present in its monomeric form, with only limited amounts of dimers and oligomers, while at 15 wt-% gliadin, a gel-like hydrated solid is formed as a result of formation of aggregates [103]. Under heat and pressure, the gliadins have also been shown to form hexagonal structures with a 65 Å lattice parameter [27,33,38]. Ab initio modeling has been used to investigate the reasons for formation of hexagonal structures by the gliadins, and has partly ascribed it to glycerol acting as a chemical chaperon aiding in the packing of the protein molecules [38]. The prolamins in maize (zeins) have been modeled with MD simulations to a greater extent than the wheat gliadins. A structural feature reported for the zeins is an α-helix with four amino acid residues per turn and a hydrophobic face inside the helix, formed by the non-polar residues with the carotenoid lutein, helping to stabilize the structure [111]. During solvent evaporation, the zeins are able to self-assemble into different protein shapes, including rods, spheres, and films of different sizes, partly due to so-called head-to-tail binding of the proteins [112]. Early studies on wheat suggested head-to-tail binding [113,114] of the gliadins while aggregating, which has also been shown in confirmed studies [103,105].

5.4. Glutelins

The glutelins are even larger and more complicated molecules than the prolamins [104] (Table 4), and wheat glutelins (glutenins) are known to form the largest polymers in nature [115] (Figure 1). This is one reason for the limited modeling of these proteins. The vast majority of previous studies have focused on examining how these large polymers are formed and the background to their formation [10,31]. Depolymerization and re-polymerization of the gluten polymer are known to occur during processing operations, such as dough mixing [116]. Therefore, theoretical models of polymerization mechanisms have been one way forward in using modeling tools to predict the glutenin polymer structure [117]. Through these models, the directionality of formation of the polymers (head-to-tail, head-to-head, tail-to-tail or tail-to-head) can be evaluated. In a recent study using HPLC and small-angle X-ray scattering, a highly intrinsic and disordered structure in the native glutenin protein of wheat was reported, while the structures of unknown types were formed during processing of the proteins into films [34]. Molecular modeling, homology modeling and MD simulations have also been used on the *N*-terminal domain of glutenins to examine the polymerization of the proteins into giant oligomers through disulfide bond formation [118,119].

5.5. Future Opportunities for Modeling Cereal Seed Storage Proteins

The main reasons for the lack of studies, using modeling tools and MC/MD simulations to investigate the prolamin and glutelin proteins, are that the (i) proteins are large and among the largest

proteins in nature [120]; (ii) proteins in their native form are intrinsically disordered and not soluble in water [99]; (iii) structures of the proteins change with dilution [99] or processing [31], i.e., into low levels of structures in dilute condition [103], hierarchical structures in concentrated regimes [99], and hexagonal [27,31] and other types of structures [34] developing under certain processing conditions; and iv) reported structures are always connected to aggregation of the proteins into polymers [10]. Thus, modeling these proteins is extremely difficult and requires a large amount of computer power. However, the storage proteins of cereals are highly useful and important: (i) as a storage resource for the plant to regenerate itself [7]; and (ii) for human [7] and animal [121] food and feed purposes; and are potentially useful in (iii) non-food applications, such as replacement of petroleum-based plastics [10,122]. A clear structure-function relationship has been demonstrated between the formation of protein structures during processing and the functionality of products produced from cereal storage proteins [123,124]. Therefore, a better understanding of the structural features of the plant storage proteins, especially for the under-researched prolamins and glutelins, and opportunities to fine-tune their structural performance would contribute significantly to their utility in various applications. Furthermore, such an understanding helps reveal the physiological and evolutionary reasons for variations in storage proteins among plants. Novel modeling tools, together with faster/more powerful computers and computer cluster tools to empower the simulation algorithms, are of utmost importance in such work, and will open up opportunities for identifying even the largest and most complex protein structures. Novel simulation tools and newly-developed algorithms will be available for researchers and others in the near future, together with stronger computers and computer clusters, which will pave the way for modeling, even the largest and most intrinsically disordered proteins, such as seed storage glutelin proteins. In a future perspective, research on gliadins and glutenins might also benefit from a similar approach to the approaches used for elastomeric proteins, utilizing low complexity (not coarse-graine) models and/or work with peptide fragments to reveal protein behavior [125]. Such modeling will create novel opportunities for assessing biological features in plants and for fine-tuning the properties of foods and materials produced from these proteins.

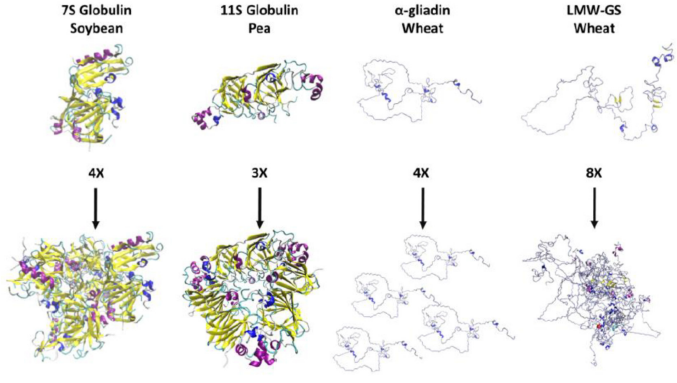


Figure 1. Single subunits and polymers obtained from these, of (left to right): The storage proteins 7S globulin in soybean, 11S globulin in pea, a-gliadin in wheat, and low molecular weight glutenin subunits (LMW-GS) in wheat. Amino acid sequences and simulation tools applied to determine the structures are described in Yoshizawa et al. [107] and Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (3AUP) [39] for 7S globulin, and in Tandang-Silvas et al. [108] and RSCB PDB (3KSC) [39] for 11S globulin. Amino acids sequences for a-gliadin and LMW-GS can be found at Uniprot accession numbers Q9ZP09 and P10386 and simulation was carried out using an all-atom model with Monte Carlo algorithms in the program Profasi [126]. All models were visualized using Visual Molecular Dynamics [127].

6. Conclusions

A combination of innovations within computing technology, increased speed, and power of computers, and novel modeling/simulation tools will increase the opportunities that determine protein structures and reveal structure-function relationships. Until recently, modeling and simulation tools have been rarely used for the evaluation of plant protein structures and structure-function relationships. Thus, less than 9% of the current deposits in the Research Collaboratory for Structural Bioinformatics Protein Data Bank are plant-related and one-third of these deposits are less than five years old. For the two most abundant plant protein types, RuBisCO and the storage proteins, modeling has been used to different extents. For RuBisCO and the photosynthetic machinery, modeling has been used rather frequently to describe the process and structural impacts of changes in the protein. Structural features of the storage proteins have been described to a lesser extent by modeling, especially for the prolamin and glutelin. The present status of modeling as regards structural features of the storage proteins is exemplified in Figure 1 for globulin (soybean, pea), prolamin (wheat α-gliadin), and glutelin (wheat glutenin) proteins, respectively.

Author Contributions: Idea for the project and this paper E.J. and M.H.; conceptualization of the present paper was done jointly by F.R., J.M. and E.J.; planning and definition of subchapters was done jointly by F.R., J.M., and E.J.; preparation of tables was done by F.R; preparation and simulation for Figure 1 was done by J.M.; writing of separate subchapters, F.R, J.M., and E.J. writing—original draft preparation, E.J.; writing—review and editing, M.H.; visualization, E.J.; supervision, E.J., M.H., F.R.; project administration, E.J.; funding acquisition, E.J., M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swedish Research Council and The Strategic Research Programme Trees and Crops for the Future (TC4F), grant number 2013-05991.

Acknowledgments: The authors want to thank Anders Ihrbäck for providing the PROFASI program for free. English language was carefully checked by William R Newson.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Altman, G.H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R.L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D.L. Silk-based biomaterials. *Biomaterials.* **2003**, *24*, 401–416. [CrossRef]
2. Meister, M.; Lowe, G.; Berg, H.C. The proton flux through the bacterial flagellar motor. *Cell* **1987**, *49*, 643–650. [CrossRef]
3. Committee on Protein Nomenclature. Joint recommendations of the physiological and biochemical committees on protein nomenclature. *J. Biol. Chem.* **1908**, *4*, 48.
4. Osborne, T.B. *The Vegetable Proteins*, 2nd ed.; Longmans, Green and Co.: London, UK, 1924; pp. 1–154.
5. Boulter, D.; Derbyshire, E. The general properties, classification and distribution of plant proteins. In *Plant. Proteins: Easter School in Agricultural Science*; Norton, G., Ed.; Butterworth-Heinemann: Oxford, UK, 2013.
6. Shewry, P.R.; Pandya, M.J. *The 2S Albumin Storage Proteins*; Springer Science and Business Media LLC: Berlin, Germany, 1999; pp. 563–586.
7. Shewry, P.R.; Napier, J.A.; Tatham, A.S. Seed storage proteins: Structures and biosynthesis. *Plant Cell* **1995**, *7*, 945–956.
8. Shewry, P.R.; Tatham, A.S. The prolamin storage proteins of cereal seeds: structure and evolution. *Biochem. J.* **1990**, *267*, 1–12. [CrossRef]
9. Belitz, H.-D.; Kieffer, R.; Seilmeier, W.; Wieser, H. Structure and function of gluten proteins. *Cereal Chem.* **1986**, *63*, 336–341.
10. Johansson, E.; Malik, A.H.; Hussain, A.; Rasheed, F.; Newson, W.R.; Plivelic, T.; Hedenqvist, M.S.; Gällstedt, M.; Kuktaite, R. Wheat Gluten Polymer Structures: The Impact of Genotype, Environment, and Processing on Their Functionality in Various Applications. *Cereal Chem. J.* **2013**, *90*, 367–376. [CrossRef]
11. Shang, H.-Y.; Wei, Y.-M.; Long, H.; Yan, Z.-H.; Zheng, Y.-L. Identification of LMW glutenin-like genes from Secale sylvestre host. *Генетика* **2005**, *41*, 1372–1380. [CrossRef]
12. Hojilla-Evangelista, M.P.; Selling, G.W.; Hatfield, R.D.; Digman, M. Extraction, composition, and functional properties of dried alfalfa (Medicago sativaL.) leaf protein. *J. Sci. Food Agric.* **2016**, *97*, 882–888. [CrossRef]

13. Proteins of the Blood Serum. Encyclopaedia Britannica. Available online: https://www.britannica.com/science/protein/Proteins-of-the-blood-serum (accessed on 21 July 2019).
14. Rodriguez, P.; Pérez, E.; Romel, G.; Dufor, D. Characterization of the proteins fractions extracted from leaves of *Amaranthus dubius* (*Amaranthus* spp.). *Afr. J. Food Sci.* **2011**, *5*, 417–424.
15. Teng, Z.; Wang, Q. Extraction, identification and characterization of the water-insoluble proteins from tobacco biomass. *J. Sci. Food Agric.* **2012**, *92*, 1368–1374. [CrossRef] [PubMed]
16. Day, P.R. The biology of plant proteins. *Crit. Rev. Food Sci. Nutr.* **1996**, *36*, 39–47. [CrossRef] [PubMed]
17. Krishnan, H.; Coe, E. Seed Storage Proteins. In *Encyclopedia of Genetics*; Elsevier BV: Amsterdam, The Netherlands, 2001; pp. 1782–1787.
18. Tabita, F.R.; Hanson, T.; Li, H.; Satagopan, S.; Singh, J.; Chan, S. Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiol. Mol. Boil. Rev.* **2007**, *71*, 576–599. [CrossRef] [PubMed]
19. Zayas, J.F. *Functionality of Proteins in Food*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 1–367.
20. Day, L. Proteins from land plants—Potential resources for human nutrition and food security. *Trends Food Sci. Technol.* **2013**, *32*, 25–42. [CrossRef]
21. Barac, M.B.; Pesic, M.; Stanojevic, S.; Kostic, A.; Cabrilo, S. Techno-functional properties of pea (Triticum sativum) protein isolates: A review. *Acta Period. Technol.* **2015**, *46*, 1–18. [CrossRef]
22. Muneer, F.; Johansson, E.; Hedenqvist, M.S.; Plivelic, T.S.; Markedal, K.E.; Petersen, I.L.; Sørensen, J.C.; Kuktaite, R. The impact of newly produced protein and dietary fiber rich fractions of yellow pea (Pisum sativum L.) on the structure and mechanical properties of pasta-like sheets. *Food Res. Int.* **2018**, *106*, 607–618. [CrossRef]
23. Gällstedt, M.; Mattozzi, A.; Johansson, E.; Hedenqvist, M.S. Transport and Tensile Properties of Compression-Molded Wheat Gluten Films. *Biomacromolecules* **2004**, *5*, 2020–2028. [CrossRef]
24. Tummala, P.; Liu, W.; Drzal, L.T.; Mohanty, A.K.; Misra, M. Influence of Plasticizers on Thermal and Mechanical Properties and Morphology of Soy-Based Bioplastics. *Ind. Eng. Chem. Res.* **2006**, *45*, 7491–7496. [CrossRef]
25. Newson, W.R.; Kuktaite, R.; Hedenqvist, M.S.; Gällstedt, M.; Johansson, E. Oilseed Meal Based Plastics from Plasticized, Hot Pressed Crambe abyssinica and Brassica carinata Residuals. *J. Am. Oil Chem. Soc.* **2013**, *90*, 1229–1237. [CrossRef]
26. Newson, W.R.; Rasheed, F.; Kuktaite, R.; Hedenqvist, M.S.; Gällstedt, M.; Plivelic, T.S.; Johansson, E. Commercial potato protein concentrate as a novel source for thermoformed bio-based plastic films with unusual polymerisation and tensile properties. *RSC Adv.* **2015**, *5*, 32217–32226. [CrossRef]
27. Rasheed, F.; Kuktaite, R.; Hedenqvist, M.S.; Gällstedt, M.; Plivelic, T.; Johansson, E. The use of the plant as a "green factory" to produce high strength gluten-based plastics. *Green Chem.* **2016**, *18*, 2782–2792. [CrossRef]
28. Wu, Q.; Andersson, R.L.; Holgate, T.; Johansson, E.; Gedde, U.; Olsson, R.T.; Hedenqvist, M.S. Highly porous flame-retardant and sustainable biofoams based on wheat gluten and in situ polymerized silica. *J. Mater. Chem. A* **2014**, *2*, 20996–21009. [CrossRef]
29. Wu, Q.; Yu, S.; Kollert, M.; Mtimet, M.; Roth, S.V.; Gedde, U.; Johansson, E.; Olsson, R.T.; Hedenqvist, M.S. Highly Absorbing Antimicrobial Biofoams Based on Wheat Gluten and Its Biohybrids. *ACS Sustain. Chem. Eng.* **2016**, *4*, 2395–2404. [CrossRef]
30. Capezza, A.J.; Newson, W.R.; Olsson, R.T.; Hedenqvist, M.S.; Johansson, E. Advances in the Use of Protein-Based Materials: Toward Sustainable Naturally Sourced Absorbent Materials. *ACS Sustain. Chem. Eng.* **2019**, *7*, 4532–4547. [CrossRef]
31. Kuktaite, R.; Plivelic, T.S.; Cerenius, Y.; Hedenqvist, M.S.; Gällstedt, M.; Marttila, S.; Ignell, R.; Popineau, Y.; Tranquet, O.; Shewry, P.R.; et al. Structure and Morphology of Wheat Gluten Films: From Polymeric Protein Aggregates toward Superstructure Arrangements. *Biomacromolecules* **2011**, *12*, 1438–1448. [CrossRef]
32. Rasheed, F.; Hedenqvist, M.S.; Kuktaite, R.; Plivelic, T.S.; Gällstedt, M.; Johansson, E. Mild gluten separation–A non-destructive approach to fine tune structure and mechanical behaviour of wheat gluten films. *Ind. Crop. Prod.* **2015**, *73*, 90–98. [CrossRef]
33. Rasheed, F.; Newson, W.R.; Plivelic, T.S.; Kuktaite, R.; Hedenqvist, M.S.; Gällstedt, M.; Johansson, E. Macromolecular changes and nano-structural arrangements in gliadin and glutenin films upon chemical modification. *Int. J. Boil. Macromol.* **2015**, *79*, 151–159. [CrossRef]

34. Rasheed, F.; Plivelic, T.S.; Kuktaite, R.; Hedenqvist, M.S.; Johansson, E. Unraveling the Structural Puzzle of the Giant Glutenin Polymer—An Interplay between Protein Polymerization, Nanomorphology, and Functional Properties in Bioplastic Films. *ACS Omega* **2018**, *3*, 5584–5592. [CrossRef]
35. Muneer, F.; Johansson, E.; Hedenqvist, M.S.; Plivelic, T.S.; Kuktaite, R. Impact of pH Modification on Protein Polymerization and Structure–Function Relationships in Potato Protein and Wheat Gluten Composites. *Int. J. Mol. Sci.* **2018**, *20*, 58. [CrossRef]
36. Capezza, A.J.; Wu, Q.; Newson, W.R.; Olsson, R.T.; Espuche, E.; Johansson, E.; Hedenqvist, M.S. Superabsorbent and Fully Biobased Protein Foams with a Natural Cross-Linker and Cellulose Nanofibers. *ACS Omega* **2019**, *4*, 18257–18267. [CrossRef]
37. Capezza, A.J.; Glad, D.; Özeren, H.D.; Newson, W.R.; Olsson, R.T.; Johansson, E.; Hedenqvist, M.S. Novel sustainable super absorbents: A one-pot method for functionalization of side-stream potato proteins. *ACS Sust. Chem. Eng.* **2019**, *7*, 17845–17854. [CrossRef]
38. Rasheed, F.; Newson, W.R.; Plivelic, T.S.; Kuktaite, R.; Hedenqvist, M.S.; Gällstedt, M.; Johansson, E. Structural architecture and solubility of native and modified gliadin and glutenin proteins: non-crystalline molecular and atomic organization. *RSC Adv.* **2014**, *4*, 2051–2060. [CrossRef]
39. RCSB PDF-101. Metods for Determining Atomic Structures. Available online: https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure (accessed on 12 February 2020).
40. Jothi, A. Principles, challenges and advances in ab initio protein structure prediction. *Protein Pept. Lett.* **2012**, *19*, 1194–1204. [CrossRef] [PubMed]
41. Lee, J.; Wu, S.; Zhang, Y. Ab Initio protein structure prediction. In *From Protein Structure to Function with Bioinformatics*; Rigden, D.J., Ed.; Springer: Dordrecht, The Netherlands, 2009.
42. Ulmschneider, J.P.; Ulmschneider, M.B.; Di Nola, A. Monte Carlo vs Molecular Dynamics for All-Atom Polypeptide Folding Simulations. *J. Phys. Chem. B* **2006**, *110*, 16733–16742. [CrossRef]
43. Earl, D.J.; Deem, M.W. Monte Carlo simulations. In *Molecular Modelling of Protein*; Kukol, A., Ed.; Humana Press, Springer: Oviedo, Spain, 2008.
44. Bürgi, R.; Kollman, P.A.; Van Gunsteren, W.F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct. Funct. Bioinform.* **2002**, *47*, 469–480. [CrossRef]
45. Al Quraishi, M. Protein-structure prediction gets real. *Nature* **2019**, *577*, 627. [CrossRef]
46. AlQuraishi, M. AlphaFold at CASP13. *Bioinform.* **2019**, *35*, 4862–4865. [CrossRef]
47. Wei, G.-W. Protein structure prediction beyond AlphaFold. *Nat. Mach. Intell.* **2019**, *1*, 336–337. [CrossRef]
48. Sato, C.; Mio, K.; Kawata, M.; Ogura, T. 3D structure determination of protein using TEM single particle analysis. *Microscopy* **2014**, *63*, i9–i10. [CrossRef]
49. Murata, K.; Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim. et Biophys. Acta (BBA) Gen. Subj.* **2018**, *1862*, 324–334. [CrossRef]
50. Lidke, D.S.; Lidke, K.A. Advances in high-resolution imaging–techniques for three-dimensional imaging of cellular structures. *J. Cell Sci.* **2012**, *125*, 2571–2580. [CrossRef] [PubMed]
51. Longchamp, J.-N.; Rauschenbach, S.; Abb, S.; Escher, C.; Latychevskaia, T.; Kern, K.; Fink, H.-W. Imaging proteins at the single-molecule level. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 1474–1479. [CrossRef] [PubMed]
52. Cavalli, A.; Salvatella, X.; Dobson, C.M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9615–9620. [CrossRef]
53. Kikhney, A.G.; Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577. [CrossRef] [PubMed]
54. Makowski, L. Characterization of proteins with wide-angle X-ray solution scattering (WAXS). *J. Struct. Funct. Genom.* **2010**, *11*, 9–19. [CrossRef]
55. Wei, W.; Hu, W.; Zhang, X.-Y.; Zhang, F.-P.; Sun, S.-Q.; Liu, Y.; Xu, F. Analysis of protein structure changes and quality regulation of surimi during gelation based on infrared spectroscopy and microscopic imaging. *Sci. Rep.* **2018**, *8*, 5566. [CrossRef]
56. Parker, M.W. Protein Structure from X-Ray Diffraction. *J. Boil. Phys.* **2003**, *29*, 341–362. [CrossRef] [PubMed]
57. RCSB Protein Data Bank. Available online: https://www.rcsb.org/pdb/home/sitemap.do (accessed on 20 July 2019).

58. Schulten, K.; Tesch, M. Coupling of protein motion to electron transfer: Molecular dynamics and stochastic quantum mechanics study of photosynthetic reaction centers. *Chem. Phys.* **1991**, *158*, 421–446. [CrossRef]
59. Wang, H.; Lin, S.; Allen, J.; Williams, J.C.; Blankert, S.; Laser, C.; Woodbury, N.W. Protein Dynamics Control the Kinetics of Initial Electron Transfer in Photosynthesis. *Science* **2007**, *316*, 747–750. [CrossRef] [PubMed]
60. Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chem. Rev.* **2015**, *115*, 11191–11238. [CrossRef]
61. Blumberger, J. Electron transfer and transport through multi-heme proteins: recent progress and future directions. *Curr. Opin. Chem. Boil.* **2018**, *47*, 24–31. [CrossRef] [PubMed]
62. Spreitzer, R.J.; Salvucci, M.E. RuBisCO: Structure, regulatory interactions, and possibilities for a better enzyme. *Ann. Rev. Plant. Biol.* **2002**, *53*, 449–475. [CrossRef] [PubMed]
63. Van Lun, M.; van der Spoel, D.; Andersson, I. Subunig interface dynamics in hexadecameric Ru BisCO. *J. Mol. Biol.* **2011**, *411*, 1083–1098. [CrossRef] [PubMed]
64. Sanbonmatsu, K.Y. Computational studies of molecular machines: the ribosome. *Curr. Opin. Struct. Boil.* **2012**, *22*, 168–174. [CrossRef] [PubMed]
65. Perilla, J.R.; Goh, B.C.; Cassidy, C.K.; Liu, B.; Bernardi, R.C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Boil.* **2015**, *31*, 64–74. [CrossRef]
66. Schallau, K.; Junker, B.H. Simulating Plant Metabolic Pathways with Enzyme-Kinetic Models1. *Plant Physiol.* **2010**, *152*, 1763–1771. [CrossRef]
67. Awasthi, M.; Jaiswal, N.; Singh, S.; Pandey, V.; Dwivedi, U.N. Molecular docking and dynamics simulation analyses unraveling the differential enzymatic catalysis by plant and fungal laccases with respect to lignin biosynthesis and degradation. *J. Biomol. Struct. Dyn.* **2014**, *33*, 1–15. [CrossRef]
68. Rambal, S. The differential role of mechanisms for drought resistance in a Mediterranean evergreen shrub: a simulation approach. *Plant Cell Environ.* **1993**, *16*, 35–44. [CrossRef]
69. Stone, J.M.; Heard, J.E.; Asai, T.; Ausubel, F. Simulation of Fungal-Mediated Cell Death by Fumonisin B1 and Selection of Fumonisin B1–Resistant (fbr) Arabidopsis Mutants. *Plant Cell* **2000**, *12*, 1811–1822.
70. Letort, V.; Mahe, P.; Cournède, P.H.; de Reffye, P.; Courtois, B. Plant growth models: Simulation of quantitative trait loci detection for model parameters and application to potential yield optimization. *Ann. Bot.* **2008**, *101*, 1243–1254. [CrossRef]
71. Lannou, C.; Vallavieille-Pope, C.; Biass, C.; Goyeau, H. The Efficacy of Mixtures of Susceptible and Resistant Hosts to Two Wheat Rusts of Different Lesion Size: Controlled Condition Experiments and Computerized Simulations. *J. Phytopathol.* **1994**, *140*, 227–237. [CrossRef]
72. Zhu, X.-L.; Zhang, L.; Chen, Q.; Wan, J.; Yang, G.-F. Interactions of Aryloxyphenoxypropionic Acids with Sensitive and Resistant Acetyl-Coenzyme A Carboxylase by Homology Modeling and Molecular Dynamic Simulations. *J. Chem. Inf. Model.* **2006**, *46*, 1819–1826. [CrossRef]
73. Dehury, B.; Sahu, M.; Sarma, K.; Sahu, J.; Sen, P.; Modi, M.; Sharma, G.D.; Choudhury, M.D.; Barooah, M. Molecular Phylogeny, Homology Modeling, and Molecular Dynamics Simulation of Race-Specific Bacterial Blight Disease Resistance Protein (xa5) of Rice: A Comparative Agriproteomics Approach. *OMICS: A J. Integr. Boil.* **2013**, *17*, 423–438. [CrossRef] [PubMed]
74. Mari, A. When does a protein become an allergen? Searching for a dynamic definition based on most advanced technology tools. *Clin. Exp. Allergy* **2008**, *38*, 1089–1094. [CrossRef] [PubMed]
75. Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, *113*, 821–830. [CrossRef]
76. Jimenez-Lopez, J.C.; Kotchoni, S.O.; Rodriguez-García, M.I.; Alche, J.D.D. Structure and functional features of olive pollen pectin methylesterase using homology modeling and molecular docking methods. *J. Mol. Model.* **2012**, *18*, 4965–4984. [CrossRef]
77. Pedraza, M.; Becerril-Luján, B.; Agundis, C.; Dominguez-Ramirez, L.; Pereyra, A.; Riaño-Umbarila, L.; Rodriguez-Romero, A. Analysis of B-cell epitopes from the allergen Hev b 6.02 revealed by using blocking antibodies. *Mol. Immunol.* **2009**, *46*, 668–676. [CrossRef]
78. Pasquato, N.; Berni, R.; Folli, C.; Folloni, S.; Cianci, M.; Pantano, S.; Helliwell, J.R.; Zanotti, G. Crystal Structure of Peach Pru p 3, the Prototypic Member of the Family of Plant Non-specific Lipid Transfer Protein Pan-allergens. *J. Mol. Boil.* **2006**, *356*, 684–694. [CrossRef]

79. Rundqvist, L.; Tengel, T.; Zdunek, J.; Björn, E.; Schleucher, J.; Alcocer, M.; Larsson, G. Solution Structure, Copper Binding and Backbone Dynamics of Recombinant Ber e 1–The Major Allergen from Brazil Nut. *PLoS ONE* **2012**, *7*, e46435. [CrossRef]
80. Grutsch, S.; Fuchs, J.E.; Freier, R.; Kofler, S.; Bibi, M.; Asam, C.; Wallner, M.; Ferreira, F.; Brandstetter, H.; Liedl, K.R.; et al. Ligand Binding Modulates the Structural Dynamics and Compactness of the Major Birch Pollen Allergen. *Biophys. J.* **2014**, *107*, 2972–2981. [CrossRef]
81. Hung, L.-W.; Kohmura, M.; Ariyoshi, Y.; Kim, S.-H. Structural differences in D and L-monellin in the crystals of racemix mixture. *J. Mol. Biol.* **1999**, *285*, 311–321. [CrossRef] [PubMed]
82. Paladino, A.; Costantini, S.; Colonna, G.; Facchiano, A. Molecular modelling of miraculin: Structural analyses and functional hypotheses. *Biochem. Biophys. Res. Commun.* **2008**, *367*, 26–32. [CrossRef] [PubMed]
83. Assadi-Porter, F.; Tonelli, M.; Radek, J.T.; Cornilescu, C.C.; Markley, J.L. How sweet is it: Detailed molecular and functional studies on Brazzein, a sweet protein and its analogs. *ACS Symp. Ser.* **2008**, *979*, 560–572.
84. Rega, M.F.; Di Monaco, R.; Leone, S.; Donnarumma, F.; Spadaccini, R.; Cavella, S.; Picone, D. Design of sweet protein based sweeteners: Hints from structure–function relationships. *Food Chem.* **2015**, *173*, 1179–1186. [CrossRef]
85. Staswick, P.E. Storage proteins of vegetative plant tissues. Ann. Rev. Plant. Physiol Plant. *Mol. Biol.* **1994**, *45*, 303–322.
86. Das, O.; Rasheed, F.; Kim, N.K.; Johansson, E.; Capezza, A.J.; Kalamkarov, A.L.; Hedenqvist, M.S. The development of fire and microbe resistant sustainable gluten plastics. *J. Clean. Prod.* **2019**, *222*, 163–173. [CrossRef]
87. Ceresino, E.B.; Kuktaite, R.; Sato, H.H.; Hedenqvist, M.S.; Johansson, E. Impact of gluten separation process and transglutaminase source on gluten based dough properties. *Food Hydrocoll.* **2019**, *87*, 661–669. [CrossRef]
88. Alander, B.; Capezza, A.; Wu, Q.; Johansson, E.; Olsson, R.; Hedenqvist, M.S. A facile way of making inexpensive rigid and soft protein biofoams with rapid liquid absorption. *Ind. Crop. Prod.* **2018**, *119*, 41–48. [CrossRef]
89. Wu, Q.; Rabu, J.; Goulin, K.; Sainlaud, C.; Chen, F.; Johansson, E.; Olsson, R.; Hedenqvist, M. Flexible strength-improved and crack-resistant biocomposites based on plasticised wheat gluten reinforced with a flax-fibre-weave. *Compos. Part A: Appl. Sci. Manuf.* **2017**, *94*, 61–69. [CrossRef]
90. Marla, S.; Bharatiya, D.; Bala, M.; Singh, V.; Kumar, A. Classification of rice seed storage proteins using neural netoworks. *J. Plant Biochem. Biophy.* **2010**, *19*, 123–126.
91. Radhika, V.; Sree Hari Rao, V. Computational approaches for the classification of seed storage proteins. *J. Food Sci. Technol.* **2015**, *52*, 4246–4255. [CrossRef] [PubMed]
92. McCoy, A.J.; A Kortt, A. The 1.8 Å crystal structure of winged bean albumin 1, the major albumin from Psophocarpus tetragonolobus (L.) DC 1 1Edited by K. Nagai. *J. Mol. Boil.* **1997**, *269*, 881–891. [CrossRef] [PubMed]
93. Hennig, M.; Schlesier, B.; Dauter, Z.; Pfeffer, S.; Betzel, C.; Hohne, W.E.; Wilson, K.S. A tim barrel protein without enzymatic-activity—Crystal-structure of Narbonin at 1.8-Angstrom resolution. *Febs. Lett.* **1992**, *306*, 80–84. [CrossRef]
94. Hennig, M.; Schlesier, B. Crystallization of seed globulins from legumes. *Acta Crystallogr. Sect. D Boil. Crystallogr.* **1994**, *50*, 627–631. [CrossRef]
95. Ko, T.P.; Ng, J.D.; McPherson, A. The three-dimensional structure of canavalin from jack bean (Canavalia ensiformis). *Plant Physiol.* **1993**, *101*, 729–744. [CrossRef]
96. Lawrence, M.C.; Izard, T.; Beuchat, M.; Blagrove, R.J.; Colman, P.M. Structure of phaseoilin at 2.2 Å resolution: Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *J. Mol. Biol.* **1994**, *238*, 748–776. [CrossRef]
97. Warwicker, J.; Connor, J. A model for vicilin solubility at mild acidic pH, based on homology modelling and electrostatics calculations. *Protein Eng. Des. Sel.* **1995**, *8*, 1243–1251. [CrossRef]
98. Withana-Gamage, T.S.; Hegedus, D.D.; Qiu, X.; Wanasundara, J. In Silico Homology Modeling To Predict Functional Properties of Cruciferin. *J. Agric. Food Chem.* **2011**, *59*, 12925–12938. [CrossRef]
99. Dahesh, M.; Banc, A.; Duri, A.; Morel, M.-H.; Ramos, L. Polymeric Assembly of Gluten Proteins in an Aqueous Ethanol Solvent. *J. Phys. Chem. B* **2014**, *118*, 11065–11076. [CrossRef]
100. Ullsten, N.H.; Gällstedt, M.; Johansson, E.; Gräslund, A.; Hedenqvist, M.S. Enlarged processing window of plasticised wheat gluten using salisylic acid. *Biomacromolecules* **2006**, *7*, 771–776. [CrossRef]

101. Olabarrieta, I.; Cho, S.-W.; Gällstedt, M.; Sarasua, J.-R.; Johansson, E.; Hedenqvist, M.S. Aging Properties of Films of Plasticized Vital Wheat Gluten Cast from Acidic and Basic Solutions. *Biomacromolecules* **2006**, *7*, 1657–1664. [CrossRef] [PubMed]
102. Ullsten, N.H.; Gällstedt, M.; Cho, S.W.; Spencer, G.; Johansson, E.; Hedenqvist, M.S. Properties of extruded vital wheat gluten sheets with sodium hydroxide and salisylic acid. *Biomacromolecules* **2009**, *10*, 479–488. [CrossRef] [PubMed]
103. Sato, N.; Matsumiya, A.; Higashino, Y.; Funaki, S.; Kitao, Y.; Oba, Y.; Inoue, R.; Arisaka, F.; Sugiyama, M.; Urade, R. 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.* **2015**, *63*, 8715–8721. [CrossRef] [PubMed]
104. Shewry, P.R.; Casey, R. *Seed Proteins*; Springer Science and Business Media LLC: Berlin, Germany, 1999; pp. 1–10.
105. Herrera, M.; Vazquez, D.; Geisler, R.; Drechsler, M.; Hertle, Y.; Hellweg, T.; Dodero, V.I. Insights into gliadin supramolecular organization at digestive pH 3.0. *Colloids Surfaces B: Biointerfaces* **2018**, *165*, 363–370. [CrossRef]
106. Rocha, A.J.; Sousa, B.L.; Girão, M.S.; Barroso-Neto, I.L.; Monteiro-Júnior, J.E.; Oliveira, J.T.; Nagano, C.S.; Carneiro, R.F.; Monteiro-Moreira, A.C.; Rocha, B.A.M.; et al. Cloning of cDNA sequences encoding cowpea (Vigna unguiculata) vicilins: Computational simulations suggest a binding mode of cowpea vicilins to chitin oligomers. *Int. J. Boil. Macromol.* **2018**, *117*, 565–573. [CrossRef]
107. Yoshizawa, T.; Shimizu, T.; Yamabe, M.; Taichi, M.; Nishiuchi, Y.; Shichijo, N.; Unzai, S.; Hirano, H.; Sato, M.; Hashimoto, H. Crystal structure of basic 7S globulin, a xyloglucan-specific endo-β-1,4-glucanase inhibitor protein-like protein from soybean lacking inhibitory activity against endo-β-glucanase. *FEBS J.* **2011**, *278*, 1944–1954. [CrossRef]
108. Tandang-Silvas, M.R.G.; Fukuda, T.; Fukuda, C.; Prak, K.; Cabanos, C.; Kimura, A.; Itoh, T.; Mikami, B.; Utsumi, S.; Maruyama, N. Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochim. et Biophys. Acta (BBA) Proteins Proteom.* **2010**, *1804*, 1432–1442. [CrossRef]
109. Kuktaite, R.; Larsson, H.; Johansson, E. Variation in protein composition of wheat flour and its relationship to dough mixing behavior. *J. Cereal Sci.* **2004**, *40*, 31–39. [CrossRef]
110. Purcell, J.M.; Kasarda, D.D.; Wu, C.-S.C. Secondary structures of wheat α- and ω-gliadin proteins: Fourier transform infrared spectroscopy. *J. Cereal Sci.* **1988**, *7*, 21–32. [CrossRef]
111. Momany, F.A.; Sessa, D.J.; Lawton, J.W.; Selling, G.W.; Hamaker, S.A.H.; Willett, J.L. Structural Characterization of α-Zein. *J. Agric. Food Chem.* **2006**, *54*, 543–547. [CrossRef]
112. Wang, Y.; Padua, G.W. Formation of Zein Microphases in Ethanol−Water. *Langmuir* **2010**, *26*, 12897–12901. [CrossRef] [PubMed]
113. Wieser, H. Chemistry of gluten proteins. *Food Microbiol.* **2007**, *24*, 115–119. [CrossRef] [PubMed]
114. Lindsay, M.P.; Skerritt, J.H. The glutenin macropolymer of wheat flour doughts: Structure-function perspectives. *Trends Food Sci Technol.* **1999**, *10*, 247–253. [CrossRef]
115. Wrigley, C.W. Giant proteins with flour power. *Nature* **1996**, *381*, 738–739. [CrossRef] [PubMed]
116. Weegels, P.; Hamer, R.; Schofield, J. Depolymerisation and Re-polymerisation of Wheat Glutenin During Dough Processing. II. Changes in Composition. *J. Cereal Sci.* **1997**, *25*, 155–163. [CrossRef]
117. Weegels, P.L. Theoretical considerations on the directionality in polymerization mechanism and the polymer type. Macromol. *Theory Simul.* **1996**, *5*, 299–303. [CrossRef]
118. Cazalis, R.; Aussenac, T.; Rhazi, L.; Marin, A.; Gibrat, J.-F. Homology modelling and molecular dynamics simulations of the N-terminal domain of wheat high molecular weight glutenin subumit 10. *Prot. Sci.* **2003**, *12*, 34–43. [CrossRef]
119. Köhler, P.; Keck-Gassenmeier, B.; Wieser, H.; Kasarda, N.D. 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. J.* **1997**, *74*, 154–158. [CrossRef]
120. Altpeter, F.; Vasil, V.; Srivastava, V.; Vasil, I. Integration and expression of the high-molecular-weight glutenin subunit 1Ax1 gene into wheat. *Nat. Biotechnol.* **1996**, *14*, 1155–1159. [CrossRef]
121. Shewry, P.R.; Halford, N.G. Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* **2002**, *53*, 947–958. [CrossRef]

122. Gällstedt, M.; Pettersson, H.; Johansson, T.; Newson, W.R.; Johansson, E.; Hedenqvist, M.S. Film Extrusion of Crambe abyssinica/Wheat Gluten Blends. *J. Vis. Exp.* **2017**, e54770. [CrossRef] [PubMed]
123. Johansson, E.; Prieto-Linde, M.L.; Jönsson, J. Effects of wheat cultivar and nitrogen application on storage protein composition and bread-making quality. *Cereal Chem.* **2001**, *78*, 19–25. [CrossRef]
124. Kuktaite, R.; Newson, W.R.; Rasheed, F.; Plivelic, M.S.; Hedenqvist, M.S.; Gällstedt, M.; Johansson, E. Monitoring nano-structure dynamics and polymerization in glycerol plasticized wheat gliadin and glutenin films: Relation to mechanical properties. *ACS Sust. Chem. Eng.* **2016**, *4*, 2998–3007. [CrossRef]
125. Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like polypeptides as models of intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2477–2486. [CrossRef]
126. Irbäck, A.; Mohanty, S. PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *J. Comput. Chem.* **2006**, *27*, 1548–1555. [CrossRef]
127. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

II

biomolecules

MDPI

Article

Glutenin and Gliadin, a Piece in the Puzzle of their Structural Properties in the Cell Described through Monte Carlo Simulations

Joel Markgren [1], Mikael Hedenqvist [2], Faiza Rasheed [2], Marie Skepö [3] and Eva Johansson [1,*]

1 Department of Plant Breeding, Swedish University of Agricultural Sciences, P.O. Box 101, SE-230 53 Alnarp, Sweden; joel.markgren@slu.se

2 Department of Fibre and Polymer Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden; mikaelhe@kth.se (M.H.); faizar@kth.se (F.R.)

3 Theoretical Chemistry, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden; marie.skepo@teokem.lu.se

* Correspondence: eva.johansson@slu.se; Tel.: +46-40-41-55-62

Received: 10 June 2020; Accepted: 21 July 2020; Published: 23 July 2020

Abstract: Gluten protein crosslinking is a predetermined process where specific intra- and intermolecular disulfide bonds differ depending on the protein and cysteine motif. In this article, all-atom Monte Carlo simulations were used to understand the formation of disulfide bonds in gliadins and low molecular weight glutenin subunits (LMW-GS). The two intrinsically disordered proteins appeared to contain mostly turns and loops and showed "self-avoiding walk" behavior in water. Cysteine residues involved in intramolecular disulfide bonds were located next to hydrophobic peptide sections in the primary sequence. Hydrophobicity of neighboring peptide sections, synthesis chronology, and amino acid chain flexibility were identified as important factors in securing the specificity of intramolecular disulfide bonds formed directly after synthesis. The two LMW-GS cysteine residues that form intermolecular disulfide bonds were positioned next to peptide sections of lower hydrophobicity, and these cysteine residues are more exposed to the cytosolic conditions, which influence the crosslinking behavior. In addition, coarse-grained Monte Carlo simulations revealed that the protein folding is independent of ionic strength. The potential molecular behavior associated with disulfide bonds, as reported here, increases the biological understanding of seed storage protein function and provides opportunities to tailor their functional properties for different applications.

Keywords: modeling; intrinsically disordered proteins; gluten; disulfide bonds; cysteine; prolamin; Monte Carlo

1. Introduction

The gluten protein complex is highly abundant in the wheat seed [1] and consists of storage proteins, which are known to form the largest protein networks in nature [2]. Similar types of proteins are found in other cereals and grasses (Poaceae) and are all genetically closely related and show high resemblance [3,4]. These proteins often share a high proline and glutamine content, amino acids lined in repetitive motifs, and cysteines (CYS) that form disulfide bonds that, in several cases, crosslink with other storage proteins [5]. Gluten proteins are divided into two groups: gliadins and glutenins [5,6]. Based on the amino acid composition, the gliadins are then further subdivided into three major classes, α-/β-, γ-, and ω-gliadins, whereas the glutenins are divided into low molecular weight (LMW) and high molecular weight (HMW) glutenin subunits (GS). Among these proteins, the gliadins and the LMW-GS have a similar evolutionary background; they share several features in their primary structure.

Both types of protein contain regions of repetitive sequences, mainly of hepta- and dodecapeptides repeat motifs rich in proline and glutamine residues [7]. Even though the gliadins and the LMW-GS share similarities in their amino acid composition, their behavior of forming polymers differs in their native state, i.e., in the wheat grain. Gliadins are known as monomeric in their native state, with only intramolecular disulfide bonds (except for ω-gliadins due to the lack of CYS), while LMW-GS are known to form polymers by intermolecular disulfide bonds with other glutenin proteins, forming large polymers [8].

In the cell, both gliadins and glutenins are synthesized in the ribosomes of the rough endoplasmic reticulum (ER) and transported to ER lumen [9]. After this, the proteins form disulfide bonds between CYS, and the glutenins polymerize into polymers [10]. The formation of disulfide bonds is regulated by the cytosolic and redox conditions in the cell [10]. Furthermore, the level of humidity in the wheat grain is important for the amount and size distribution of the polymers [11]. The ER has been shown to administrate the folding of the proteins, while smaller proteins, such as chaperons and foldases, assist in the folding process [12–14]. The proteins are stored at different locations in the plant seed cell in the form of protein bodies (PB) [15], where the transport to the destined locations is co-translational [16,17]. Two trafficking routes have been described for proteins leaving the ER: one is via the Golgi apparatus with deposition in PBs of vacuolar origin, and one is in the ER lumen for later fusion into the vacuole [9,18,19]. The reason for the different pathways is not yet fully understood; the same type of protein has been found to use both pathways, and all types of gluten proteins have been found in the different types of PBs [9,15]. However, segregation among the gluten proteins being deposited in various types of PBs has also been reported, with the actual stage of grain development being suggested as one determinant for the transportation pathway [1,18].

The folding and assembling of gliadins and glutenins into monomers and polymers seems to be a predetermined process dictating the formation of the disulfide bonds within the glutenin polymers and gliadin monomers. Among the gliadins (except for the ω-gliadins), this behavior is observed in the specific intramolecular disulfide bonds that are consistently formed between certain conserved CYS [5,20–22]. Similarly, the LMW-GS contains several CYS that form specific intramolecular disulfide bonds, being homologous to those of the gliadins [20,22,23]. Additionally, LMW-GS and HMW-GS form intermolecular disulfide bonds, which leads to glutenin polymers. These are formed by the CYS that are unique to LMW-GS and HMW-GS and absent in the gliadins [5,22,23]. Therefore, the consistency of the formation of the disulfide bonds indicates the importance of structural features in the proteins for disulfide binding opportunities.

To the authors' knowledge, there are no studies available today that explain the reasons for the differences in crosslinking behavior or in-vivo structural properties of gliadins and LMW-GS. The complexity of gluten proteins constantly hampers cellular studies, where issues like the synthesis of multiple types of gluten proteins, rapid polymerization of several types of the proteins, and difficulties in purification [10,23] are major obstacles. Computer simulations that consider the physical and chemical properties of the amino acids in the protein could here be a valuable tool to describe the structures of the proteins [24]. Both gliadins and LMW-GS contain high contents of proline and glutamine [6], which are known to destabilize the protein structure as well as hinder specific conformations [25]. Proteins that lack specific 3D structural conformations in their native state are classified as intrinsically disordered proteins (IDP) [26], and in previous studies, gluten proteins have also been classified as IDPs [27]. In previous work on IDPs from sources other than gluten, computer simulations have been used on both the atomistic and coarse-grained levels to describe structural properties. Most commonly, the simulated results have been compared and validated with experimental results from various methods, including small-angle X-ray scattering (SAXS) [28–32]. For gluten proteins in particular, only a limited number of studies have focused on the use of computational tools to understand their structural behavior [33–37]. At present, no computational simulation studies are available describing the structural differences between gliadins and LMW-GS in their native state.

Here, atomistic and coarse-grained Monte Carlo simulations were used to characterize two wheat storage protein types, gliadins and LMW-GS. The aim was to understand the differences and similarities in structure-dependent behavior between the two proteins, explaining the reason for the formation of disulfide bonds through intra- and intermolecular interactions. Furthermore, a model for the synthesis of the proteins and their folding is proposed.

2. Materials and Methods

2.1. Protein Amino Acid Sequences Used for Simulations

In the present study, two proteins, an α-gliadin and an LMW-GS, were selected for Monte Carlo simulations, to understand their structural behavior. The amino acid sequences, as well as their charge distribution, are shown in supplementary A, as collected from the UniProt database (https://www.uniprot.org/) under accession numbers Q9ZP09 and P10386, respectively. They contain 268 (α-gliadin) and 288 (LMW-GS) amino acids with a positive net charge of one (α-gliadin) and seven (LMW-GS). Studies have previously been conducted on these two specific proteins, characterizing their existence and primary structures experimentally [38–40]. The origins of the two sequences are spelt wheat and spring wheat [38–40], resembling typical α-gliadin and M-type LMW-GS primary sequences according to the literature [7,41]. Here, the sequences of the proteins are considered as representative sequences for gliadins and LMW-GS, respectively, due to the close relationship and similarities with most of the gliadins and LMW-GS [6]. In the simulations, the *N*-terminal sections of 20 and 23 amino acids of both proteins were omitted since these amino acids are known to code for a transport peptide that is removed during synthesis [34,42,43].

2.2. Models

Two different types of models, one all-atom and one coarse-grained, in combination with Monte Carlo simulations, were applied to model the protein structures, conformational ensembles, and sensitivity to salinity.

The all-atom models were used to estimate the average folding of the proteins, with and without disulfide bonds, using the force field "Lund FF08" with the software "PROFASI" [44,45]. In principle, the all-atom simulations followed the same principles as described by Fagerberg et al. [46], where the interaction potential includes electrostatic interactions between adjacent peptide units, a contribution from the excluded volume, two types of hydrogen bonds, i.e., backbone–backbone bonds and charged side-chain–backbone bonds, as well as effective hydrophobic interactions. All the interactions were assumed to be pairwise additive. To model intramolecular disulfide bonds, the distance between the involved CYS—more precisely, between the sulfur atoms of the thiol groups— was restrained to a mean distance of 2 Å based on previous findings [47,48]. The restraints were present throughout the entire simulations for all intramolecular disulfide bonds. The modeled disulfide bonds were applied between the CYS pairs listed in Table 1, which have previously been experimentally described [20,21,23].

Table 1. Disulfide bond pairs in α-gliadin and low molecular weight glutenin subunits (LMW-GS), where the cysteines (CYS) number refers to the CYS amino acid position.

α-gliadin	LMW-GS
CYS 128–CYS 158	CYS 2—Interdisulfide bond
CYS 159–CYS 249	CYS 127–CYS 162
CYS 171–CYS 257	CYS 135–CYS 155
	CYS 163–CYS 260
	CYS 210—Interdisulfide bond

The coarse-grained models were used to evaluate how electrostatic interactions affect the conformational ensemble of the proteins. Here, each amino acid corresponds to one bead, which can

be either charged or neutral depending on the primary sequence. For more information about the model, see supplementary B or [28,49].

The all-atom models treat the water, implicitly utilizing hydrogen bonding, excluded volume, and side-chain potential components to describe it, with cutoffs of 4.3–4.5 Å, which correspond to salinity in the range of 450–500 mM. In the coarse-grained model, the salt and water are treated implicitly and regulated by the Debye screening length and the dielectric constant, respectively, utilizing an extended Debye–Hückel potential. Four different salt concentrations corresponding to 10, 80, 500, or 1000 mM were used to evaluate how the conformational ensemble of the proteins is affected by the charged amino acids. Counter-ions were added explicitly to the simulation box to obtain an electroneutral system.

2.3. Monte Carlo Parameters

The all-atom systems were studied utilizing Monte Carlo simulations with the Metropolis algorithm. The simulations were performed in the canonical ensemble, meaning that a constant number of particles (N), volume (V), and temperature (T) were used. The model protein was located in a cubic box of length 1500 Å, where periodic boundary conditions (PBC) were applied in all directions. The force field utilizes cutoffs of 4.3 Å for its excluded volume term and 4.5 Å for its hydrogen bond and side-chain components. Five different displacements were allowed: (i) rotation of the whole chain, (ii) translation of the whole chain, (iii) pivot rotation of single backbone angles, (iv) rotation of side-chain angles, and (v) biased Gaussian steps, which is a local move described by Favrin et al. [50]. For these simulations, an annealing procedure was used to accelerate the simulations and to avoid the simulations becoming trapped in local energy minima [51]. Here, ten different temperatures in the range 300–400 K were applied, and a total of 1000 configurations were produced at each temperature at each cycle. For more information, see supplementary C. The protein model was located with a random start conformation in the box, and an initial simulation run composing of 200 annealing cycles resulting in $2{\cdot}\times 10^5$ configurations at 300 K was performed for equilibration purposes. The following production run involved 800 annealing cycles, corresponding to $8 \times{\cdot}10^5$ configurations at 300 K. Every 10th of the produced configuration was saved for further analysis, to secure the availability of the variation in the data, simultaneously considering data storage capacity (the exact amount of saved configurations for each temperature and model system is presented in supplementary C).

The equilibrium properties of the coarse-grained Monte Carlo simulations were evaluated using the Metropolis algorithm. For this purpose, the integrated Monte Carlo/molecular dynamics/Brownian dynamics simulation package Molsim [52] was used. The simulations were performed in the canonical ensemble. Further simulation details are described in supplementary B.

2.4. Analysis

2.4.1. Hydrophobicity

A hydropathy index graph for the amino acids in the protein was produced with the ExPASy tool "Prot Scale" [53] to describe the hydropathy of the amino acids in the peptide chain. The tool was set to produce a value for every individual amino acid residue in the protein sequence, utilizing the Kyte and Doolittle scale with a window size of 15. The window size was used to illustrate patterns for more extensive parts of the proteins. The chosen scale is based on the free energy of the amino acids when they are transferred between the vapor, ethanol, and water phases [54]. Most hydropathy scales provide a similar estimation of the non-polar amino acid properties but can vary greatly when estimating the polar ones, depending on which solvents the amino acids are transferred between [55]. To broaden the analysis of the proteins, the Rose et al. scale [56] was also used, which estimates an amino acid's tendency to be buried in hydrophobic cores or to be exposed to solvent at protein surfaces.

2.4.2. Radius of Gyration

The R_g value, also referred to as the mean square distance of the center of mass, for the individual obtained configurations from the all-atom and coarse-grained systems were calculated by built-in modules in the simulation packages PROFASI and MOLSIM.

2.4.3. Polymer Scaling Laws, Shape Factor, and Small Angle X-Ray Scattering (SAXS)

The Flory power law was used to describe the proteins' behavior in the solvent [57]:

$$R_g \alpha\ R_0 N^{\nu} \tag{1}$$

where the R_g is considered proportional to a constant (R_0) and the number of amino acids (raised to the power of ν). R_0 is a prefactor value that equals 2 Å, as determined from experiments on other proteins [58,59]. A ν close to 1/3 indicates a collapsed protein structure due to a poor solvent, 1/2 indicates theta conditions, and 3/5 indicates good solvent conditions (an expanded "self-avoiding" chain) [59,60].

To further describe the shape, the ratio of the mean-square of the end-to-end distance (R_{ee}) and mean-square R_g was used as a shape factor (R_{shape}).

$$\frac{\left\langle R_{ee}^2 \right\rangle}{\left\langle R_g^2 \right\rangle} = R_{shape} \tag{2}$$

A R_{shape} equal to 1 represents a globular shape, 6 a self-avoiding random walk (SARW) shape, and 12 an extended rod shape [28]. Here, the distance between the terminal α carbons was considered as the R_{ee}.

The spherical average and flexibility of the two proteins were described by a wavelength function derived from the SAXS data, depicted in the form of Kratky plots [61]. The SAXS calculations were in turn performed on the all-atom models using a Debye formula in the foXS software [62,63].

2.4.4. Contact Mapping

Contact maps were generated in the all-atom simulations to illustrate the distance between the amino acids. Here, the built-in function module "generate native contact list" in the PROFASI software was used. The module was set to register distances closer than 5 Å between all non-hydrogen atoms in the amino acid side-chains, which resulted in distinguishable patterns of the protein contacts. The contact map does not register distances between adjacent amino acids.

2.4.5. Secondary Structure Estimation

The software STRIDE [64], which estimates secondary structure based on patterns of hydrogen bonding and backbone angles, was used to estimate the secondary structure for each conformation of the all-atom simulations.

2.4.6. Plotting, Graphics, and Statistical Calculations

All plotting and statistical calculations were made with the statistical programming language R version 3.5.1, "Feather Spray" [65]. The packages used for plotting purposes were as follows: ggplot2, scales, egg, gridExtra, grid, plyr, dplyr, doParallel, ggraph, ggpubr, and igraph.

The 3D rendering graphics were made with the molecular visualization software VMD, with tachyon in-memory rendering settings [66].

3. Results

3.1. Shape Classification Based on Charges

At physiological pH, the α-gliadin carries seven positive and six negative charges and the LMW-GS nine positive and two negative charges, resulting in corresponding net charges of +1 and +7. This would suggest that the proteins are classified as polar tracts with globular and tadpole shapes according to the Das–Pappu classification (Figure 1) [67,68]. However, the prediction is uncertain due to a high content of proline and hydrophobic amino acids which, in the Das–Pappu classification, have an unknown impact on protein structure.

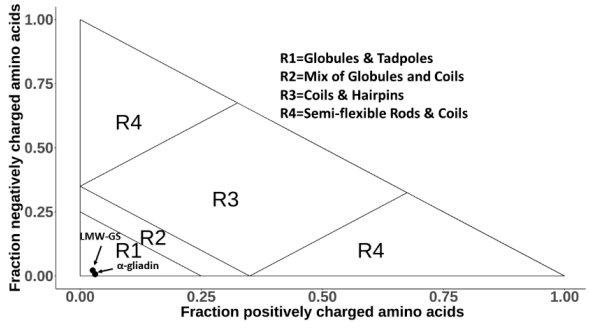


Figure 1. Das–Pappu plot, where intrinsically disordered proteins (IDPs) are categorized into four different structure categories depending on their fractions of charged amino acids [67,68]. Both the α-gliadin and the LMW-GS are positioned in the R1 category, meaning that their shape is supposed to be either globular or tadpole-like.

3.2. Hydrophobicity of Gluten Proteins

Differences in hydropathy were observed along the amino acid chain for both α-gliadin (Figure 2a) and LMW-GS (Figure 2b) after the primary sequences were evaluated against the Kyte and Doolittle scale with the Prot Scale tool. In general, the proteins were similar from a hydropathy perspective, although some differences were visible between the two proteins (compare Figure 2a with Figure 2b). The segments containing the first 100 residues of amino acids were similar for the α-gliadin and LMW-GS, being rather hydrophilic, although with some parts of lower hydrophilicity and also a larger variation in hydrophilicity in the α-gliadin (Figure 2a) as compared to the LMW-GS (Figure 2b). A short, highly hydrophilic segment was found in both proteins directly after amino acid residue 100. Furthermore, for both proteins, the segments from amino acid residue 125 to approximately amino acid residue 180 were found to be more hydrophobic, but again with more hydrophilic parts in α-gliadin compared to LMW-GS. In the next segment, from approximately amino acid residue 180 to 250, both proteins were found to be rather hydrophilic, with a slightly higher amount of hydrophilicity for the LMW-GS than α-gliadin. Then, from amino acid residue 250 to the end of the protein, both proteins were again found to be relatively hydrophobic. The accumulated hydropathy index for α-gliadin was −0.98 (variation of 1.08), whereas the corresponding number for LMW-GS was −0.71 (variation of 0.75). The Rose et al. [56] scale provided a similar depiction of the proteins (supplementary D).

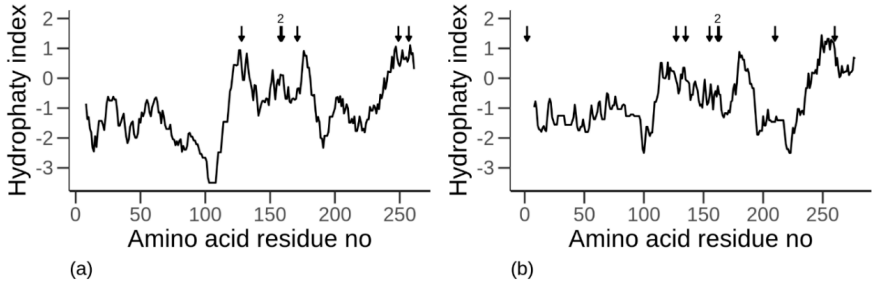


Figure 2. Hydropathy index (hydrophobicity for each amino acid with a window size of 15 amino acids) for each amino acid residue of (**a**) α-gliadin and (**b**) low molecular weight glutenin, indicating hydrophobic (+) and hydrophilic (−) regions in the protein. Cysteine amino acids are marked with an arrow, and the number 2 indicates two vicinal cysteines marked with two arrows at that position.

The positions of the CYS in each of the proteins (Figure 2) indicate that all CYS, including the vicinal ones, are located in relatively hydrophobic regions (−0.5 to +1) for the α-gliadin (Figure 2a). This is contrary to the position of two of the CYS (1 and 7 (from left to right) of the CYS in LMW-GS, Figure 2b) in the LMW-GS that are located in relatively hydrophilic regions (−1 to −1.5). The remaining CYS (2–6 and 8) in the LMW-GS were located in the hydrophobic regions of the protein, similarly to the CYS in the α-gliadin (Figure 2b)

3.3. Protein Solubility and Salt Effects

The probability distribution of R_g (Figure 3) for the all-atom simulations was skewed for both proteins without disulfide bonds, with a mean of 47.8 Å and 47.9 Å and with a variance of 118 Å and 117 Å for α-gliadin and LMW-GS, respectively, indicating an open structure. A less skewed distribution, with mean values of 44.0 Å and 43.3 Å and a variance of 55 Å and 36.5 Å, respectively, was found for α-gliadin and LMW-GS with intramolecular disulfide bonds, indicating a more compact structure. Intramolecular disulfide bonds in the proteins also contributed to larger peaks and a close to normal probability distribution function in terms of R_g, the latter especially for the LMW-GS protein.

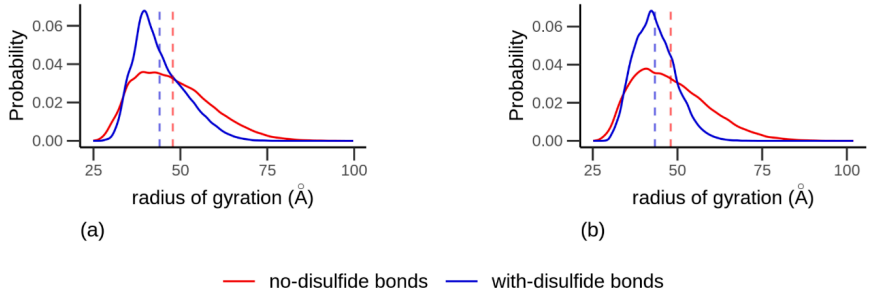


Figure 3. The probability distribution of the radius of gyration for (**a**) α-gliadin and (**b**) low molecular weight glutenin from all-atom simulations, where the red curve indicates a protein without intramolecular disulfide bonds, and the blue curve indicates a protein with intramolecular disulfide bonds according to experimental data [20,21,23]. The broken line displays the mean of the curve.

The R_g for the coarse-grained models did not alter with variations in salinity and provided a root square mean (rms) of 63 Å for α-gliadin and 58 Å for LMW-GS, results which are compared with the all-atom model results in supplementary B. The small change in R_g due to altered Debye screening length and the small energy contributions indicate that the proteins are composed of a relatively low

amount of charged amino acids (net charge +1 and +7), shown in the charge distribution map in supplementary A, with low impact on the proteins' conformation.

Flory values (ν) of 0.55 and 0.57 were obtained with and without intramolecular disulfide bonds, respectively, for both the all-atom α-gliadin and LMW-GS models when applying the Flory scaling law; thus, the implicitly modeled water can be considered as a theta solvent for both proteins, since ν is between 1/2 and 3/5.

3.4. Protein Shape and Internal Structural Arrangement

The Kratky plots of the α-gliadin (Figure 4a) and the LMW-GS (Figure 4b) from the all-atom simulations visualized the similarity in the folding of the two proteins. Both proteins showed a broad maximum at qR_g of 1 to 12, indicating the presence of slightly denser or more folded regions in the proteins. For $12 < qR_g < 15$, the curves were relatively horizontal, indicating the presence of unfolded regions in both proteins. Thereafter, a slight increase is shown, indicating that both proteins also contain extended regions. When proteins were modeled with the proposed disulfide bonds [20–23], the Kratky plots resulted in a maximum/shoulder occurring at lower qR_g values, indicating both proteins to be denser or more folded in comparison to the case without disulfide bonds.

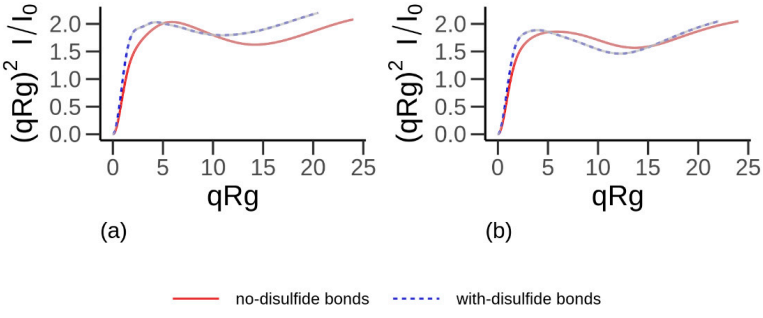


Figure 4. Dimensionless Kratky plot (visualizing shape of proteins) of α-gliadin (**a**) and low molecular weight glutenin (**b**). The solid red curve corresponds to protein without intramolecular disulfide bonds, and the dashed blue curve corresponds to proteins with intramolecular disulfide bonds. Data obtained from all-atom simulations. q is the scattering vector [61] (also referred to as the momentum transfer), I is intensity, and I_0 is the intensity of the incoming ray.

The calculations of the estimated shape of α-gliadin and LMW-GS yielded the following averaged R_{shape} values: when disulfide bonded, 2.97 and 2.18, respectively, and when not disulfide bonded, 4.9 and 5.1, respectively. The distribution of the R_{shape} values can be found in supplementary E. Thus, these calculations indicated a shape in-between globular and SARW, with a more globular shape when disulfide bonds are present.

Ring and tail like motifs could often be observed when visually inspecting snapshots from all-atom simulations of both proteins when these contained disulfide bonds (Figure 5a,b). The ring in the present snapshots is located between CYS 171 (being from amino acid residue 171 of the protein) to CYS 249 in α-gliadin and LMW-GS between CYS 163 to CYS 260. The tail in the present snapshots is observed from the *N*-terminal to the vicinity of amino acid 120 in both proteins, yielding a tadpole-like structure. The proteins had random chain structures when similar visual inspections were performed on the coarse-grained simulated structures (Figure 5c).

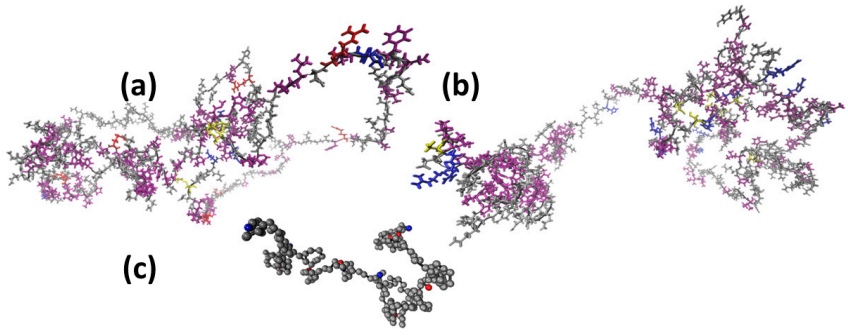


Figure 5. Simulation snapshots of all-atom (**a**) α-gliadin, (**b**) LMW-GS, and coarse-grained (**c**) α-gliadin. The grey colored groups represent neutral amino acids/beads, red indicates positively charged, blue negatively charged, purple are amino acids with hydrophobic potentials, and yellow are cysteine amino acids.

Additionally, the contact maps for the α-gliadin (Figure 6a,b) and the LMW-GS (Figure 6c,d), determined with (Figure 6b,d) and without (Figure 6a,c) disulfide bonds, indicated high similarity in the protein structure between the two evaluated proteins. An increased contact among amino acids for the next-nearest neighbors is visible for both proteins, independently of whether they are disulfide bonded or not; see black diagonal (Figure 6a–d). Furthermore, both proteins, with or without disulfide bonds, show a higher probability in the *N*-terminal and the repetitive region (residue 1 to 100), i.e., enhanced contacts. Differences among the evaluated proteins are visible in the C-terminal region of the proteins, where the α-gliadin shows an increased probability of contact from residue 200 and onward, whereas the LMW-GS shows contacts of amino acids primarily from residue 250 and onward. For disulfide-bonded proteins, only a low frequency of contacts is found in the C-terminal end for both proteins. Instead, distinct areas of contacts between amino acids increase in the vicinities of the CYS involved in the disulfide bonds.

Determination of the frequency of contacts between various CYS for the proteins without disulfide-bonds indicated that all residues are occasionally in contact with other CYS (not necessarily all other CYS) (Figure 7) during the all-atom simulation. For α-gliadin, CYS 159 showed an elevated occurrence with three CYS (CYS 128, CYS 249, and CYS 257). The other CYS in α-gliadin showed different tendencies for contacts. For the LMW glutenin, CYS 2 and CYS 210 showed contacts with all other cysteine residues in the proteins, although, for CYS 210, the instances of contact were generally low, in contrast to CYS 2, which had a high number of contacts with CYS 127, Cys 135, and CYS 260. A relatively high frequency of contacts was also visible between CYS 127 and CYS 162, as well as between CYS 135 and CYS 155. Since the software does not register contacts for adjacent amino acids, there are no data on the frequency of contacts for the vicinal CYS (CYS 158 and 159 in α-gliadin or CYS 162 and 163 in LMW-GS).

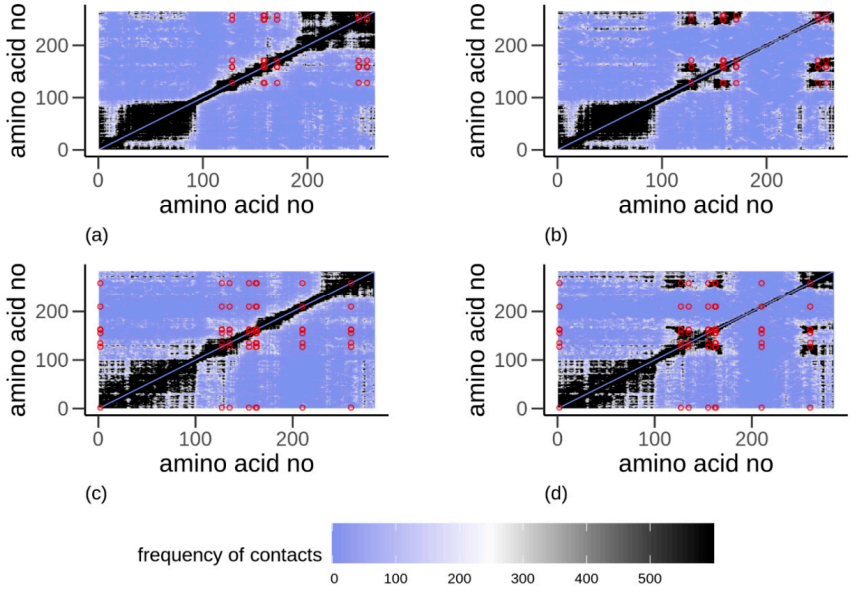


Figure 6. Contact maps for (**a**,**b**) α-gliadin and (**c**,**d**) low molecular weight glutenin, (**a**,**c**) without disulfide bonds and (**b**,**d**) with disulfide bonds obtained from all-atom simulations. Dark regions indicate a high amount of contacts among amino acids. A high degree of contact among amino acid residues positioned two steps away is seen as a black diagonal. Dark areas annotated by circles show a high degree of contact between disulfide bonded cysteine residues.

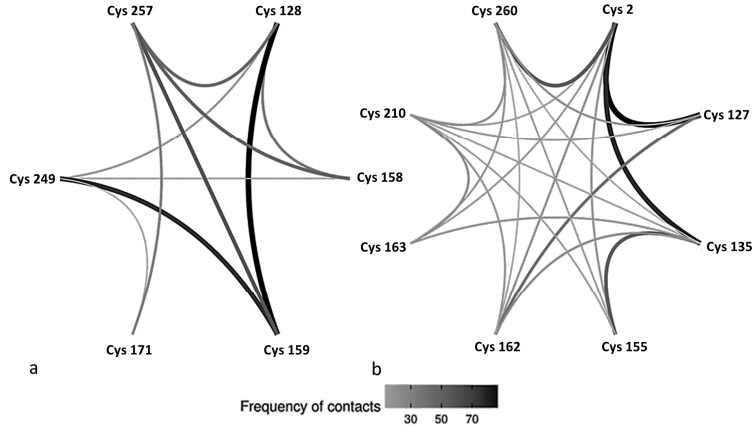


Figure 7. Contact map between cysteine residues in (**a**) α-gliadin without disulfide bonds and (**b**) low molecular weight glutenin without disulfide bonds. The darker color of the line corresponds to a higher frequency of contacts. The results are derived from all-atom simulations.

3.5. Secondary Structure

The α-gliadin (Figure 8a,b) and the LMW-GS (Figure 8c,d) showed a similar secondary structure propensity in the absence of intramolecular disulfide bonds (dark curve), with a relatively uniform distribution of α-helices and β-sheets/strands, whereas when intramolecular disulfide bonds are

included, there is an increased tendency for β-sheets/strands for the amino acids involved in these specific bonds. In general, the proteins showed a large change in secondary structure propensity when disulfide bonds were introduced (compare dark-grey curves with light-grey curves in Figure 8), resulting in a diminishing of α-helices (from around amino acid residue 125 and onwards). In contrast, for amino acid residues 1–125, no such effects were visible.

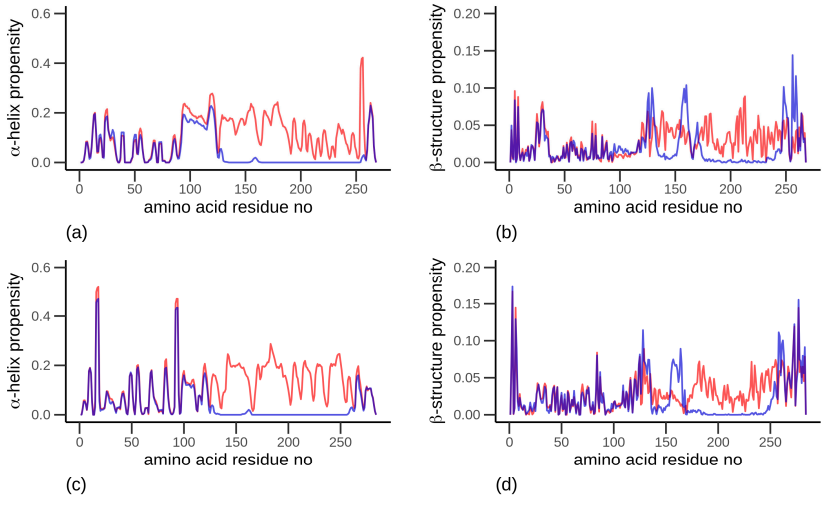


Figure 8. Propensity of secondary structure for (**a**,**c**) α-helix and (**b**,**d**) β-sheets/strands, for (**a**,**b**) α-gliadin and (**c**,**d**) low molecular weight glutenin. The red curves indicate proteins without intramolecular disulfide bonds, and the blue curves indicate protein with intramolecular disulfide bonds. Observe the differences in scales, which were used to more clearly visualize the β-structure propensity. The data is obtained from all-atom simulations.

4. Discussion

The present study clearly demonstrates the striking similarities between two of the main types of wheat grain storage proteins, gliadins and LMW-GS, concerning secondary structure, folding, size, and shape, according to the simulated results. The main difference between the two proteins was the presence of two CYS residues located in a relatively hydrophilic part of the LMW-GS, while all other CYS residues, both in the α-gliadin and in the LMW-GS, were located in more hydrophobic parts of the molecules. The two CYS located in the more hydrophilic part of the LMW-GS form intermolecular disulfide bonds, while the other CYS in both proteins forms intramolecular disulfide bonds [20,22].

Earlier studies of wheat storage proteins classified LMW-GS as an aggregated type of gliadin, which was corrected in the 1980s by Shewry and co-workers [6,69–71]. Since then, additional similarities have been defined between the two proteins within their primary structures, i.e., amino acid repeats, high level of proline and glutamine, vicinal CYS, etc. [7]. Such similarities have led to the conclusion that gliadins and LMW-GS share a common evolutionary ancestor [72]. Recent genomic studies on the gliadin and LMW-GS loci have elucidated a great extent of repetitive DNA and genes with tandem duplication, thereby challenging sequence assembly studies [73]. Sequencing and analyses of orthologous regions of gliadin and LMW-GS genes have shown relatively large differences between these gene families. The LMW-GS genes were found to be less clustered and separated by more considerable distances, due to the insertion of repetitive blocks and interspread gene loci, as compared to the gliadin genes [74]. Besides primary structure, structural features of the gliadins and the LMW-GS have had limited description, basically due to the aggregation and intrinsically disordered features of the proteins complicating such studies [24]. Therefore, the present study is one of the first of its kind

to contribute to describing the similarities in terms of secondary structure, hydropathy, and folding patterns based on Kratky plots and contact maps for both protein types.

In the present study, it was possible to describe some of the potential background reasons and mechanisms behind the formation of intramolecular disulfide bonds within both proteins. The formation of intramolecular disulfide bonds within both proteins is specific, with particular CYS always forming bonds with each other [5,20–22]. Certain intramolecular disulfide bonds may also affect the formation of other intramolecular disulfide bonds, e.g., the bond between CYS 127 and 162 in LMW-GS may prevent the protein from forming aberrant aggregates [75,76]. Our results indicate that the hydrophobic amino acid segments, adjacent to the CYS involved in intramolecular disulfide bonds, are probably crucial for disulfide bond formation, as also reported for elastin aggregation [77]. Until now, the impact of hydrophobicity in the amino acid chain of the gluten proteins on the formation of crosslinks in the wheat seed has been limitedly experimentally studied. Systematic mutations changing this property along the amino acid chain is an interesting idea for a future study on the topic that would allow a more in-depth understanding of its effect on crosslinking. However, such studies of wheat grain are still challenging to perform experimentally. Moreover, by simulations, where the mutations can easily be incorporated in the amino acid chain, modeling tools need to be developed to be able to describe the effects on crosslinking by the changed hydrophobicity. Furthermore, necessary experimental verifications are also still lacking.

Further describing the potential crosslinking backgrounds for the two proteins, previous studies have indicated that the CYS crosslinking of the LMW-GS takes place as a polymerization directly after the formation of the proteins in the ribosomes of rough ER and transportation to ER lumen [10]. Our results regarding contacts between CYS within the gliadin and LMW-GS molecules, respectively, indicate flexibility; hence, most of the CYS have an opportunity to reach each other for the formation of disulfide bond crosslinks. However, one can presume that the intramolecular disulfide bonds are not formed simultaneously; a bond is likely to be formed immediately after CYS is produced in an intramolecular disulfide bond couple. This rapid chronological bond formation results in intramolecular disulfide bonds, in accordance with previous reports [5,20–22]. Thus, the intramolecular disulfide bond formation can be seen to occur based on the chronological position of CYS and predicted contact propensity for CYS presented in this study, i.e., in α-gliadin, CYS 127 is synthesized first and has frequent contact with the second synthesized CYS 158, where a disulfide bond is formed. Then, CYS 159 is next synthesized and has contacts with the fifth synthesized CYS 249 and forms a disulfide bond, since there is already a bond present between CYS 127 and 158, and so on.

Two of the LMW-GS CYS (CYS 2 and CYS 210) are, according to previous studies, responsible for the formation of intermolecular disulfide bonds [5,22,23]. Following the discussion above, neither CYS 2 nor CYS 210 form intramolecular disulfide bonds due to the lack of neighboring hydrophobic amino acids. The present study clearly shows that CYS 2 is located in a mobile part of the molecule, as shown by a large number of contacts with other CYS in the contact map evaluation. In contrast, CYS 210 is located in a more rigid part of the molecule, as seen from the contact map evaluation. As intermolecular disulfide bonds are most likely also formed directly after synthesis or during synthesis, the N-terminal CYS (CYS 2) crosslinks with surrounding GS at an earlier stage than CYS 210, as has been described using in-vitro studies of LMW-GS proteins [10]. The C-terminal CYS (CYS 210) is more passive, with a slower rate of crosslinking, corresponding to previously reported results [75]. The CYS 210 part of the protein is also more likely transformed into α-helical or β-strand/sheet structures before the intermolecular disulfide bonds are formed, in line with the secondary structure propensity results from this study. In addition to the described crosslinking steps, disulfide genomic differences, with different distances between encoding genes, have also been described as one possible contributor to the differences in intra/intermolecular disulfide bond formation [74].

Intermolecular disulfide bond formation is the basis by which the gluten proteins form their large polymers and involves, together with the LMW-GS, the HMW-GS [20,23]. Intermolecular disulfide bond formation, like intramolecular disulfide bond formation, also seems to be a rather

pre-determined event, with certain CYS at particular GS proteins more commonly crosslinking to each other [7,20,23]. Previous theoretical models have described the directionality of the formation of the polymers (head-to-tail, head-to-head, tail-to-tail) [78], which also indicate the pre-determined formation of the polymers. Determination of the molecular background factors of the intermolecular disulfide bond formation were not the aim of this study, but features described here for the LMW-GS—e.g., hydropathy, time, and chronology of synthesis of the CYS; flexibility of the amino acid chain; structure formation—might play a role in determining the crosslinking. Similarly, as was found for the LMW-GS in the present study, the crosslinking sites of the HMW-GS have been reported to be located in hydrophilic regions of the proteins [79,80]. Moreover, recent investigations have indicted surface hydrophobicity in the *N*-terminal domains of HMW-GS as an important factor for interdisulfide bond formation [81].

The α-gliadin model applied in the present study corresponded well with previous experimental data on the protein. The α-gliadin has, in correspondence with simulations in the present study, been experimentally characterized with a similar size, secondary structure, and lack of a particular fold conformation and is described as a disordered protein with a secondary structure mostly occupied by random coils and turns [82–86]. In particular, previous work determines the size of α-gliadin R_g to be between 35.5 and 41 Å [84,86], which is within the R_g distribution range mostly populated by the simulations. Previous studies [84,86] have described the protein to fit an elliptical model, while here, the protein shapes correspond to an intermediate structure between a globular and expanded SARW shape, possibly an elliptical shape. These potential model differences may be related to force field, solvent type, and/or concentrations of proteins and solvents applied. To our knowledge, similar experimental data as for the gliadin, describing the native monomeric "wild-type" LMW-GS before aggregation, is not available in the literature. This lack of data may be the result of the extensive aggregation through disulfide crosslinking that the protein undergoes directly after/during synthesis [10,76], suggesting that there are no obtainable monomeric native LMW-GS. Only in a few studies, the structural features have been investigated of more hydrophobic IDPs, although those indicate that these proteins are expanded or in theta conditions in an aqueous environment [87], which is in line with the results presented here. Furthermore, polyampholyte and polyelectrolyte IDPs can also have an expanded shape [67,68]. However, neither α-gliadin nor LMW-GS can be classified as polyampholytes or polyelectrolytes since the charged amino acids had only a minor effect on the model structures presented here.

The present study investigated the structural properties of one α-gliadin and one LMW-GS, chosen as good representatives of the gliadin and LMW-GS types of storage proteins in wheat. Despite the rather high level of diversity within the two protein type groups, we hypothesize enough similarities between different proteins within these groups to draw general conclusions based on the results of the present study. Based on the results in the present study and previous findings described above, we propose a model describing the synthesis and folding of the two wheat storage proteins, the gliadins and the LMW-GS, and their assembly into protein bodies (Figure 9). The proteins are synthesized in the cell at the rough ER (Figure 9a), where folding occurs, and intramolecular disulfide bonds are formed. The hydrophobicity of the amino acid peptide sections that are next to the CYS regulates the predetermined intramolecular disulfide bonds to be formed. Thus, the intramolecular disulfide bonds of the protein are formed sequentially and rapidly after each associated CYS is synthesized (Figure 9b,c). After this, the LMW-GS forms intermolecular disulfide bonds (Figure 9e,f), starting with the formation of bonds involving the CYS present in the more flexible *N*-terminal of the protein. The CYS forming intermolecular disulfide bonds are located next to amino acid peptide sections that are less hydrophobic than those close to CYS forming intramolecular disulfide bonds. This is the reason that glutenin polymerization is affected by cytosolic conditions in the cell, e.g., redox conditions. Since external conditions such as humidity and precipitation affect the cytosolic conditions in the cell, the cellular location in which the polymerization occurs explains differences in the size distribution of storage protein polymers in wheat. Such variation in polymer distribution is known to impact wheat quality [88,89]. Additionally, other external conditions, such as temperature, drought, and plant

development time, to mention a few factors, are known to affect wheat storage protein polymerization during grain maturation, and maturity impacts the conditions of the cell both directly and indirectly by differing plant responses [8,90]. Chaperons and foldases might assist in the folding process of the storage proteins, although this has not been evaluated by us, and the results of this study indicate good opportunities for the correct folding and formation of disulfide bonds without such contributions. However, when forming protein bodies, chaperons and foldases might provide functions for correct packing and transport signaling, which has been described for homologous storage proteins in other grass species [14,91]. The formation of protein bodies is most probably a result of similar hydrophobic interactions and crosslinking, as described above (Figure 9g). An increasing concentration and amount of hydrophobic groups contribute to proteins more prone to self-associating [77,92], which is speculated to be an initial stage of protein body formation for gliadins [93,94]. Folded and aggregated proteins are then transported via different routes to the assembly in PBs at various places in the cell, depending on the stages of grain development.

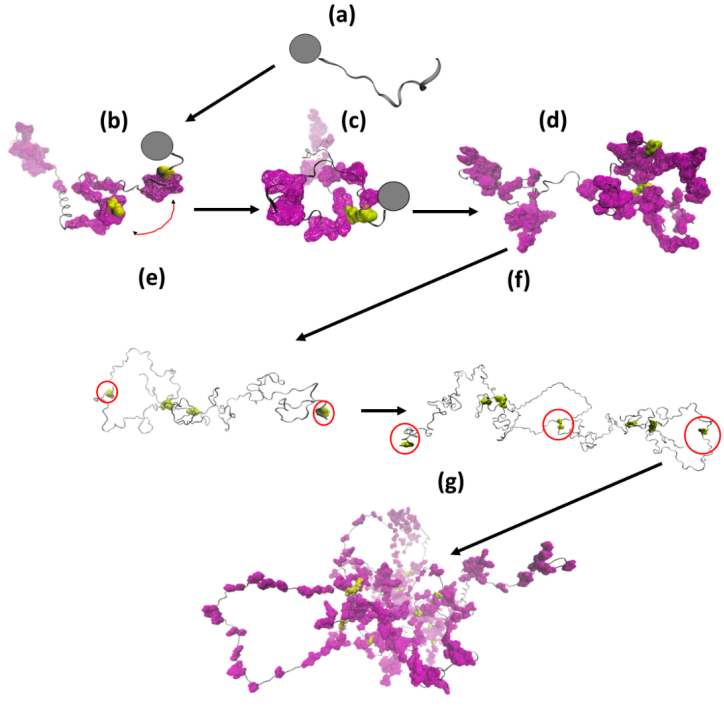


Figure 9. A proposed model for the synthesis, folding, disulfide bond formation, and storage of α-gliadin and low molecular weight glutenin subunits in the wheat grain: (**a**) start of the synthesis of the α-gliadin at the ribosome (dark circle); (**b**) as soon as cysteines residues (yellow surfaces) are synthesized, adjacent sections of amino acids influence the 3D-shape of the molecule, and those which are hydrophobic (purple surfaces) bring cysteine residues close; (**c**) closeness of cysteine residues contributes to the formation of intramolecular disulfide bonds (yellow areas); (**d**) a fully synthesized α-gliadin with disulfide bonded cysteine residues (yellow areas) and a disordered fold; (**e**) a fully synthesized LMW-GS that has undergone the same formation of intramolecular disulfide bonds (yellow areas) but with two cysteine residues (with neighboring sections of amino acids that are less hydrophobic) that are not crosslinked (here encircled in red); (**f**) intermolecular disulfide bonds are formed after protein synthesis for the LMW-GS; (**g**) protein bodies are formed by glutenins and gliadins through intermolecular disulfide bond crosslinks (LMW-GS) and hydrophobic interactions.

5. Conclusions

Our models of gliadins and LMW-GS are strikingly similar when it comes to secondary structure, size, and shape. The major difference between the two proteins are two CYS in the LMW-GS, positioned in parts of the protein with lower hydrophobicity compared to the rest of the CYS in the LMW-GS and also all CYS in the gliadins. These two CYS form intermolecular disulfide bonds, while the rest of the CYS residues form intramolecular disulfide bonds. The hydrophobicity of the adjacent section amino acids, the chronology of synthesis of the amino acids, and the flexibility of the amino acid chain for the contact between CYS to occur are the major determinants for when and how different intra- and intermolecular disulfide bonds are formed, following a predetermined structure. Generally, intramolecular disulfide bonds are formed first, among CYS with adjacent sections of hydrophobic amino acids, and in the order of synthesis, if allowed by amino acid chain flexibility. After this, intermolecular bonds form the polymers, where CYS in the mobile part of the amino acid chain plays a more active role than CYS in the more rigid part of the protein. Cytosolic conditions influence the rate and complexity of intermolecular disulfide bonds formed.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/10/8/1095/s1. References [95,96] are cited in the supplementary materials.

Author Contributions: Conceptualization, J.M., M.H., F.R., and E.J.; Data curation, J.M.; Formal analysis, J.M.; Funding acquisition, M.H. and E.J.; Investigation, J.M.; Methodology, J.M. and M.S.; Project administration, E.J.; Resources, J.M. and M.H.; Software, J.M. and M.S.; Supervision, M.H., F.R., M.S., and E.J.; Validation, J.M., M.H., M.S., and E.J.; Visualization, J.M.; Writing—original draft, J.M.; Writing—review and editing, M.H., F.R., M.S., and E.J. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by the Swedish Research Council, grant number 2013-05991, and Trees and Crops for the Future TC4F. The computational power was provided by Swedish National Infrastructure for Computation HPC2N Kebnekaise Umeå.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Tosi, P.; Gritsch, C.S.; He, J.; Shewry, P.R. Distribution of gluten proteins in bread wheat (Triticum aestivum) grain. *Ann. Bot.* **2011**, *108*, 23–35. [CrossRef] [PubMed]
2. Wrigley, C. Giant proteins with flour power. *Nature* **1996**, *381*, 738–739. [CrossRef] [PubMed]
3. de Souza, C.; Soares, R.; Balbuena, T.S.; Arruda, P. Structure, Organization, and Expression of the Alpha Prolamin Multigenic Family Bring New Insights into the Evolutionary Relationships among Grasses. *Plant Genome* **2015**, *8*, 1–11.
4. Xu, J.-H.; Messing, J. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. *Theor. Appl. Genet.* **2009**, *119*, 1397. [CrossRef] [PubMed]
5. Shewry, P.R.; Napier, J.A.; Tatham, A.S. Seed storage proteins: Structures and biosynthesis. *Plant Cell* **1995**, *7*, 945–956.
6. Shewry, P.R.; Tatham, A.S.; Forde, J.; Kreis, M.; Miflin, B.J. The classification and nomenclature of wheat gluten proteins: A reassessment. *J. Cereal Sci.* **1986**, *4*, 97–106. [CrossRef]
7. Wieser, H. Chemistry of gluten proteins. *Food Microbiol.* **2007**, *24*, 115–119. [CrossRef]
8. Johansson, E.; Malik, A.H.; Hussain, A.; Rasheed, F.; Newson, W.R.; Plivelic, T.; Hedenqvist, M.S.; Gällstedt, M.; Kuktaite, R. Wheat Gluten Polymer Structures: The Impact of Genotype, Environment, and Processing on Their Functionality in Various Applications. *Cereal Chem.* **2013**, *90*, 367–376. [CrossRef]
9. Tosi, P.; Parker, M.; Gritsch, C.S.; Carzaniga, R.; Martin, B.; Shewry, P.R. Trafficking of storage proteins in developing grain of wheat. *J. Exp. Bot.* **2009**, *60*, 979–991. [CrossRef]
10. Lombardi, A.; Marshall, R.S.; Castellazzi, C.L.; Ceriotti, A. Redox regulation of glutenin subunit assembly in the plant endoplasmic reticulum. *Plant J.* **2012**, *72*, 1015–1026. [CrossRef]

11. Johansson, E.; Prieto-Linde, M.L.; Gissén, C. Influences of weather, cultivar and fertiliser rate on grain protein polymer accumulation in field-grown winter wheat, and relations to grain water content and falling number. *J. Sci. Food Agric.* **2008**, *88*, 2011–2018. [CrossRef]
12. Vitale, A.; Ceriotti, A. Protein quality control mechanisms and protein storage in the endoplasmic reticulum. A conflict of interests? *Plant Physiol.* **2004**, *136*, 3420–3426. [CrossRef] [PubMed]
13. DuPont, F.M.; Hurkman, W.J.; Tanaka, C.K.; Chan, R. BiP, HSP70, NDK and PDI in wheat endosperm. I. Accumulation of mRNA and protein during grain development. *Physiol. Plant.* **1998**, *103*, 70–79. [CrossRef]
14. Vitale, A.; Boston, R.S. Endoplasmic Reticulum Quality Control and the Unfolded Protein Response: Insights from Plants. *Traffic* **2008**, *9*, 1581–1588. [CrossRef]
15. Loussert, C.; Popineau, Y.; Mangavel, C. Protein bodies ontogeny and localization of prolamin components in the developing endosperm of wheat caryopses. *J. Cereal Sci.* **2008**, *47*, 445–456. [CrossRef]
16. Donovan, G.; Lee, J.; Longhurst, T. Cell-Free Synthesis of Wheat Prolamins. *Aust. J. Plant Physiol.* **1982**, *9*, 59–68. [CrossRef]
17. Greene, F.C. In Vitro Synthesis of Wheat (Triticum aestivum L.) Storage Proteins. *Plant Physiol.* **1981**, *68*, 778–783. [CrossRef]
18. Rubin, R.; Levanony, H.; Galili, G. Evidence for the Presence of Two Different Types of Protein Bodies in Wheat Endosperm. *Plant Physiol.* **1992**, *99*, 718–724. [CrossRef]
19. Arcalis, E.; Marcel, S.; Altmann, F.; Kolarich, D.; Drakakaki, G.; Fischer, R.; Christou, P.; Stoger, E. Unexpected Deposition Patterns of Recombinant Proteins in Post-Endoplasmic Reticulum Compartments of Wheat Endosperm. *Plant Physiol.* **2004**, *136*, 3457–3466. [CrossRef]
20. Keck, B.; Köhler, P.; Wieser, H. Disulphide bonds in wheat gluten: Cystine peptides derived from gluten proteins following peptic and thermolytic digestion. *Z. Lebensm. Unters. Forsch.* **1995**, *200*, 432–439. [CrossRef]
21. Müller, S.; Wieser, H. The location of disulphide bonds in α-type gliadins. *J. Cereal Sci.* **1995**, *22*, 21–27. [CrossRef]
22. Shewry, P.R.; Tatham, A.S. Disulphide Bonds in Wheat Gluten Proteins. *J. Cereal Sci.* **1997**, *25*, 207–227. [CrossRef]
23. Köhler, P.; Belitz, H.-D.; Wieser, H. Disulphide bonds in wheat gluten: Further cystine peptides from high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin and from γ-gliadins. *Z. Lebensm. Unters. Forsch.* **1993**, *196*, 239–247. [CrossRef]
24. Rasheed, F.; Markgren, J.; Hedenqvist, M.; Johansson, E. Modeling to Understand Plant Protein Structure-Function Relationships—Implications for Seed Storage Proteins. *Molecules* **2020**, *25*, 873. [CrossRef] [PubMed]
25. Theillet, F.-X.; Kalmar, L.; Tompa, P.; Han, K.-H.; Selenko, P.; Dunker, A.K.; Daughdrill, G.W.; Uversky, V.N. The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord. Prot.* **2013**, *1*, e24360. [CrossRef]
26. Dill, K.A.; Chan, H.S. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **1997**, *4*, 10–19. [CrossRef]
27. Rasheed, F.; Newson, W.R.; Plivelic, T.S.; Kuktaite, R.; Hedenqvist, M.S.; Gällstedt, M.; Johansson, E. Structural architecture and solubility of native and modified gliadin and glutenin proteins: Non-crystalline molecular and atomic organization. *RSC Adv.* **2014**, *4*, 2051–2060. [CrossRef]
28. Cragnell, C.; Rieloff, E.; Skepö, M. Utilizing Coarse-Grained Modeling and Monte Carlo Simulations to Evaluate the Conformational Ensemble of Intrinsically Disordered Proteins and Regions. *J. Mol. Biol.* **2018**, *430*, 2478–2492. [CrossRef]
29. Henriques, J.; Skepö, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the TIP4P-D Water Model and the Representativeness of Protein Disorder Models. *J. Chem. Theory Comput.* **2016**, *12*, 3407–3415. [CrossRef]
30. Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D.E. Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2012**, *134*, 3787–3791. [CrossRef]
31. Terakawa, T.; Takada, S. Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: P53 N-Terminal Domain. *Biophys. J.* **2011**, *101*, 1450–1458. [CrossRef] [PubMed]
32. Rauscher, S.; Pomès, R. Molecular simulations of protein disorder. *Biochem. Cell Biol.* **2010**, *88*, 269–290. [CrossRef] [PubMed]

33. Herrera, M.G.; Vazquez, D.S.; Sreij, R.; Drechsler, M.; Hertle, Y.; Hellweg, T.; Dodero, V.I. Insights into gliadin supramolecular organization at digestive pH 3.0. *Colloid Surf. B Biointerfaces* **2018**, *165*, 363–370. [CrossRef] [PubMed]
34. Masci, S.; D'Ovidio, R.; Lafiandra, D.; Kasarda, D.D. Characterization of a Low-Molecular-Weight Glutenin Subunit Gene from Bread Wheat and the Corresponding Protein That Represents a Major Subunit of the Glutenin Polymer. *Plant Physiol.* **1998**, *118*, 1147–1158. [CrossRef] [PubMed]
35. Yaşar, F.; Çelik, S.; Köksel, H. Molecular modeling of various peptide sequences of gliadins and low-molecular-weight glutenin subunits. *Food Nahr.* **2003**, *47*, 238–242. [CrossRef] [PubMed]
36. Cazalis, R.; Aussenac, T.; Rhazi, L.; Marin, A.; Gibrat, J.-F. Homology modeling and molecular dynamics simulations of the N-terminal domain of wheat high molecular weight glutenin subunit 10. *Prot. Sci.* **2003**, *12*, 34–43. [CrossRef]
37. Arêas, E.P.G.; Cassiano, M.M. Folding interpenetration in a gliadin model: The role of the characteristic octapeptide motif. *Biophys. Chem.* **2001**, *90*, 135–146. [CrossRef]
38. Hammond-Kosack, M.C.; Holdsworth, M.J.; Bevan, M.W. In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *EMBO J.* **1993**, *12*, 545–554. [CrossRef]
39. Colot, V.; Bartels, D.; Thompson, R.; Flavell, R. Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. *Mol. Gen. Genet.* **1989**, *216*, 81–90. [CrossRef]
40. Kasarda, D.D.; D'Ovidio, R. Deduced Amino Acid Sequence of an α-Gliadin Gene from Spelt Wheat (Spelta) Includes Sequences Active in Celiac Disease. *Cereal Chem. J.* **1999**, *76*, 548–551. [CrossRef]
41. D'Ovidio, R.; Masci, S. The low-molecular-weight glutenin subunits of wheat gluten. *J. Cereal Sci.* **2004**, *39*, 321–339. [CrossRef]
42. Anderson, O.D.; Greene, F.C. The α-gliadin gene family.II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. *Theor. Appl. Genet.* **1997**, *95*, 59–65. [CrossRef]
43. Kasarda, D.D. Structure and properties of alpha-gliadins. *Ann. Technol. Agric.* **1980**, *29*, 151–173.
44. Irbäck, A.; Mohanty, S. PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *J. Comput. Chem.* **2006**, *27*, 1548–1555. [CrossRef]
45. Irbäck, A.; Mitternacht, S.; Mohanty, S. An effective all-atom potential for proteins. *PMC Biophys.* **2009**, *2*, 2. [CrossRef]
46. Fagerberg, E.; Lenton, S.; Skepö, M. Evaluating models of varying complexity of crowded intrinsically disordered protein solutions against SAXS. *J. Chem. Theory Comput.* **2019**, *15*, 6968–6983. [CrossRef]
47. Knop, O.; Boyd, R.J.; Choi, S.C. Sulfur-sulfur bond lengths, or can a bond length be estimated from a single parameter? *J. Am. Chem. Soc.* **1988**, *110*, 7299–7301. [CrossRef]
48. Higashi, L.S.; Lundeen, M.; Seff, K. Empirical relations between disulfide bond lengths, (nitrogen or carbon)-carbon-sulfur-sulfur torsion angles, and substituents in aromatic disulfides. Crystal and molecular structure of 3,3′-dihydroxydi-2-pyridyl disulfide. *J. Am. Chem. Soc.* **1978**, *100*, 8101–8106. [CrossRef]
49. Cragnell, C.; Durand, D.; Cabane, B.; Skepö, M. Coarse-grained modeling of the intrinsically disordered protein Histatin 5 in solution: Monte Carlo simulations in combination with SAXS. *Proteins* **2016**, *84*, 777–791. [CrossRef]
50. Favrin, G.; Irbäck, A.; Sjunnesson, F. Monte Carlo update for chain molecules: Biased Gaussian steps in torsional space. *J. Chem. Phys.* **2001**, *114*, 8154–8158. [CrossRef]
51. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680. [CrossRef] [PubMed]
52. Jurij, R.; Per, L. MOLSIM: A modular molecular simulation software. *J. Comput. Chem.* **2015**, *36*, 1259–1274. [CrossRef] [PubMed]
53. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607. ISBN 978-1-58829-343-5.
54. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [CrossRef]
55. Karplus, P.A. Hydrophobicity regained. *Protein Sci.* **1997**, *6*, 1302–1307. [CrossRef]
56. Rose, G.D.; Geselowitz, A.R.; Lesser, G.J.; Lee, R.H.; Micheal, Z.H. Hydrophobicity of Amino Acid Residues in Globular Proteins. *Science* **1985**, *229*, 834–838. [CrossRef]
57. Flory, P.J. The Configuration of Real Polymer Chains. *J. Chem. Phys.* **1949**, *17*, 303–310. [CrossRef]

58. Ding, F.; Jha, R.K.; Dokholyan, N.V. Scaling Behavior and Structure of Denatured Proteins. *Structure* **2005**, *13*, 1047–1054. [CrossRef]
59. Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Nat. Acad. Sci. USA* **2012**, *109*, 16155–16160. [CrossRef]
60. Hong, L.; Lei, J. Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *J. Pol. Sci. Part B Pol. Phys.* **2009**, *47*, 207–214. [CrossRef]
61. Kikhney, A.G.; Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577. [CrossRef]
62. Schneidman-Duhovny, D.; Hammel, M.; Tainer, J.A.; Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **2016**, *44*, W424–W429. [CrossRef] [PubMed]
63. Schneidman-Duhovny, D.; Hammel, M.; Tainer, J.A.; Sali, A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105*, 962–974. [CrossRef] [PubMed]
64. Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566–579. [CrossRef] [PubMed]
65. R: What is R? Available online: https://www.r-project.org/about.html (accessed on 13 May 2019).
66. VMD—Visual Molecular Dynamics. Available online: https://www.ks.uiuc.edu/Research/vmd/ (accessed on 14 May 2019).
67. Das, R.K.; Ruff, K.M.; Pappu, R.V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112. [CrossRef] [PubMed]
68. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13392–13397. [CrossRef] [PubMed]
69. Tatham, A.S.; Field, J.M.; Smith, S.J.; Shewry, P.R. The conformations of wheat gluten proteins, II, aggregated gliadins and low molecular weight subunits of glutenin. *J. Cereal Sci.* **1987**, *5*, 203–214. [CrossRef]
70. Shewry, P.; Miflin, B.; Lew, E.; Kasarda, D. The Preparation and Characterization of an Aggregated Gliadin Fraction from Wheat. *J. Exp. Bot.* **1983**, *34*, 1403–1410. [CrossRef]
71. Jackson, E.A.; Holt, L.M.; Payne, P.I. Characterisation of high molecular weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. *Theor. Appl. Genet.* **1983**, *66*, 29–37. [CrossRef]
72. Shewry, P.R.; D'Ovidio, R.; Lafiandra, D.; Jenkins, J.A.; Mills, E.N.C.; Békés, F. CHAPTER 8—Wheat Grain Proteins. In *Wheat*, 4th ed.; Khan, K., Shewry, P.R., Eds.; American Associate of Cereal Chemists International, AACC International Press: St Paul, USA, 2009; pp. 223–298. ISBN 978-1-891127-55-7.
73. Huo, N.; Zhu, T.; Altenbach, S.; Dong, L.; Wang, Y.; Mohr, T.; Liu, Z.; Dvorak, J.; Luo, M.-C.; Gu, Y.Q. Dynamic Evolution of α-Gliadin Prolamin Gene Family in Homeologous Genomes of Hexaploid Wheat. *Sci. Rep.* **2018**, *8*, 5181. [CrossRef]
74. Gao, S.; Gu, Y.Q.; Wu, J.; Coleman-Derr, D.; Huo, N.; Crossman, C.; Jia, J.; Zuo, Q.; Ren, Z.; Anderson, O.D.; et al. Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. *Plant Mol. Biol.* **2007**, *65*, 189. [CrossRef]
75. Lombardi, A.; Barbante, A.; Cristina, P.D.; Rosiello, D.; Castellazzi, C.L.; Sbano, L.; Masci, S.; Ceriotti, A. A Relaxed Specificity in Interchain Disulfide Bond Formation Characterizes the Assembly of a Low-Molecular-Weight Glutenin Subunit in the Endoplasmic Reticulum. *Plant Physiol.* **2009**, *149*, 412–423. [CrossRef]
76. Orsi, A.; Sparvoli, F.; Ceriotti, A. Role of Individual Disulfide Bonds in the Structural Maturation of a Low Molecular Weight Glutenin Subunit. *J. Biol. Chem.* **2001**, *276*, 32322–32329. [CrossRef] [PubMed]
77. Rauscher, S.; Pomès, R. The liquid structure of elastin. *eLife* **2017**, *6*, e26526. [CrossRef]
78. Weegels, P.L. Theoretical considerations on the directionality in polymerization mechanism and polymer type. *Macromol. Theory Simul.* **1996**, *5*, 299–303. [CrossRef]
79. Bekkers, A.C.A.P.A.; de Boef, E.; van Dijk, A.A.; Hamer, R.J. Research Note: The Central Domain of High Molecular Weight Glutenin Subunits is Water-Soluble. *J. Cereal Sci.* **1999**, *29*, 109–112. [CrossRef]

80. Örnebro, J.; Nylander, T.; Eliasson, A.-C.; Shewry, P.R.; Tatham, A.S.; Gilbert, S.M. Adsorption of the High Molecular Weight Glutenin Subunit 1Dx5 Compared to the 58-kDa Central Repetitive Domain and α-gliadins. *J. Cereal Sci.* **2001**, *34*, 141–150. [CrossRef]
81. Wang, J.J.; Liu, G.; Huang, Y.-B.; Zeng, Q.-H.; Hou, Y.; Li, L.; Ou, S.; Zhang, M.; Hu, S.-Q. Dissecting the disulfide Linkage of the N-Terminal Domain of HMW 1Dx5 and Its Contributions to Dough Functionality. *J. Agric. Food Chem.* **2017**, *65*, 6264–6273. [CrossRef]
82. Shewry, P.R.; Miles, M.J.; Thomson, N.H.; Tatham, A.S. Scanning Probe Microscopes—Applications in Cereal Science. *Cereal Chem.* **1997**, *74*, 193–199. [CrossRef]
83. Kasarda, D.D.; Bernardin, J.E.; Gaffield, W. Circular dichroism and optical rotatory dispersion of alpha-gliadin. *Biochemistry* **1968**, *7*, 3950–3957. [CrossRef]
84. Thomson, N.H.; Miles, M.J.; Popineau, Y.; Harries, J.; Shewry, P.; Tatham, A.S. 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—Prot. Struct. Mol. Enzym.* **1999**, *1430*, 359–366. [CrossRef]
85. Tatham, A.S.; Miflin, B.J.; Shewry, P.R. The Beta-Turn Conformation in Wheat Gluten Proteins: Relationship to Gluten Elasticity. *Cereal Chem.* **1985**, *62*, 405–412.
86. Sato, N.; Matsumiya, A.; Higashino, Y.; Funaki, S.; Kitao, Y.; Oba, Y.; Inoue, R.; Arisaka, F.; Sugiyama, M.; Urade, R. 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.* **2015**, *63*, 8715–8721. [CrossRef] [PubMed]
87. Riback, J.A.; Bowman, M.A.; Zmyslowski, A.M.; Knoverek, C.R.; Jumper, J.M.; Hinshaw, J.R.; Kaye, E.B.; Freed, K.F.; Clark, P.L.; Sosnick, T.R. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **2017**, *358*, 238–241. [CrossRef] [PubMed]
88. Malik, A.H.; Prieto-Linde, M.L.; Kuktaite, R.; Andersson, A.; Johansson, E. Individual and interactive effects of cultivar maturation time, nitrogen regime and temperature level on accumulation of wheat grain proteins. *J. Sci. Food Agric.* **2011**, *91*, 2192–2200. [CrossRef]
89. Malik, A.H.; Kuktaite, R.; Johansson, E. Combined effect of genetic and environmental factors on the accumulation of proteins in the wheat grain and their relationship to bread-making quality. *J. Cereal Sci.* **2013**, *57*, 170–174. [CrossRef]
90. Johansson, E.; Branlard, G.; Cuniberti, M.; Flagella, Z.; Hüsken, A.; Nurit, E.; Peña, R.J.; Sissons, M.; Vazquez, D. Genotypic and Environmental Effects on Wheat Technological and Nutritional Quality. In *Wheat Quality for Improving Processing and Human Health*; Igrejas, G., Ikeda, T.M., Guzmán, C., Eds.; Springer International Publishing: Cham, Germeny, 2020; pp. 171–204. ISBN 978-3-030-34163-3.
91. Li, X.; Wu, Y.; Zhang, D.Z.; Gillikin, J.W.; Boston, R.S.; Franceschi, V.R.; Okita, T.W. Rice prolamine protein body biogenesis: A BiP-mediated process. *Science* **1993**, *262*, 1054–1056. [CrossRef]
92. Rieloff, E.; Tully, M.D.; Skepö, M. Assessing the Intricate Balance of Intermolecular Interactions upon Self-Association of Intrinsically Disordered Proteins. *J. Mol. Biol.* **2019**, *431*, 511–523. [CrossRef]
93. Boire, A.; Sanchez, C.; Morel, M.-H.; Lettinga, M.P.; Menut, P. Dynamics of liquid-liquid phase separation of wheat gliadins. *Sci. Rep.* **2018**, *8*. [CrossRef]
94. Sahli, L.; Renard, D.; Solé-Jamault, V.; Giuliani, A.; Boire, A. Role of protein conformation and weak interactions on γ-gliadin liquid-liquid phase separation. *Sci. Rep.* **2019**, *9*, 1–13. [CrossRef]
95. Archer, D.G.; Wang, P. The dielectric constant of water and Debye-Hückel limiting law slopes. *J. Phys. Chem. Ref. Data* **1990**, *19*, 371–411. [CrossRef]
96. Binder, K. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*; Oxford University Press: New York, NY, USA, 2005.

Supplementary A

The primary structure for the α-gliadin used, the sequence in brackets belongs to signal peptide not used in this study:

" (MKTFLILALLAIVATTATTA) VRVPVPQLQPQNPSQQQPQEQVPLVQQQQFLGQQQPFP

PQQPYPQPQPFPSQQPYLQLQPFPQPQLPYSQPQPFRPQQPYPQPQPQYSQPQQPISQQQQ

QQQQQQQQQQQQQQQQILQQILQQQLIPCMDVVLQQHNIAHGRSQVLQQSTYQLLQELCCQ

HLWQIPEQSQCQAIHKVVHAIILHQQQKQQQQPSSQVSFQQPLQQYPLGQGSFRPSQQNPQ

AQGSVQPQQLPQFEEIRNLALQTLPAMCNVYIPPYCTITPFGIFGTN"

Charge distribution map for the α-gliadin, where + refers to positively charged, – negatively charged and · neutral amino acids:

"(·+·····)·+·····-·····

·····+·····

·····-·····+·····-

·····-·····+·····+·····+

·····—·+·····"

Net charge for α-gliadin is + 1.

The primary structure for the LMW-GS with the same notations as for the α-gliadin:

" (MKTFLVFALLAVAATSAIAQMET) RCIPGLERPWQQQPLPPQQTFPQQPLFSQQQQQQL

FPQQPSFSQQQPPFWQQQPPFSQQQPILPQQPPFSQQQQLVLPQQPPFSQQQQPVLPPQQS

PFPQQQQQHQQLVQQQIPVVQPSILQQLNPCKVFLQQQCSPVAMPQRLARSQMLQQSSCHV

MQQQCCQQLPQIPQQSRYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQQ

SQQQLGQQPQQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNVPLYRTTTSVPFGV

GTGVGAY"

Charge distribution map for the LMW-GS:

"(·+·····)+·····-+·····

·····

·····+·····+··+·····

·····+·-··+·····

·····+·····+·····

……"

Net charge for LMW-GS is +7.

Supplementary B

Supplementary B. 1 Model description

In the coarse-grained model [28,49], the amino acids of the proteins are represented by hard spheres (beads) that mimic their excluded volume including the hydration layer, and connected via harmonic bonds. The N- and C-termini are treated explicitly to account for the extra charge of the protein terminals. The bead radius was set to 2 Å providing a realistic contact separation between the charges and accurate Coulomb interaction. The non-bonded spheres interact through a short-ranged attractive interaction as well as electrostatic interactions, where the interparticle electrostatic interactions are described on the Debye–Hückel level. The simulations are performed at constant pH with point charges. Each monomer is either negative, positive, or neutral, depending on the amino acid. All values except the ones for the selected screening lengths are from the work of Cragnell et al. [28], which describes the characteristics of IDPs.

The total potential energy of the system U_{total} contains bonded and non-bonded contributions and is given by:

$$U_{total} = U_{non-bonded} + U_{bonded} = U_{hs} + U_{el} + U_{short} + U_{bonded}, \quad (1)$$

The hard-sphere potential (U_{hs}) is defined as:

$$U_{hs} = \sum_{i,j,i\neq j} u_{ij}^{hs}(r_{ij}), \quad (2)$$

where the distance between particle pairs (r_{ij}) is the centre-to-centre distance. The excluded volume of the amino acid is taken into account through the hard-sphere potential, u_{ij}^{hs} is given by:

$$u_{ij}^{hs}(r_{ij}) = \begin{cases} 0, r_{ij} > R_i + R_j \\ \infty, r_{ij} < R_i + R_j \end{cases}, \quad (3)$$

where R_i and R_j are the radii of the beads. The electrostatic interactions U_{el} are handled through an extended Debye-Hückel potential given by:

$$U_{el} = \sum_{ij} u_{ij}^{el} r_{(ij)} = \sum_{i<j} \frac{Z_i Z_j e^2 \exp\left[-k\left(r_{ij} - (R_i - R_j)\right)\right]}{4\varepsilon_0 \varepsilon_r (1 + kR_i)(1 + kR_j) r_{ij}}, \quad (4)$$

where e is the strength and Z (positive, negative or neutral) is the type of the elementary charge, k is the inverse Debye screening length, ε_0 is the vacuum permittivity, and ε_r is the dielectric constant for water 78.4 [95]. The inverse Debye screening is used for salinity:

$$k^{-1} = \sqrt{\frac{\varepsilon_0 \varepsilon_r k_B T}{2 N_A e^2 I}}, \quad (5)$$

where k_B is the Boltzmann constant, N_A is Avogadro's number, and I is the ionic strength, which in this study was set to 10, 80, 500 or 1000 mM.

The short-ranged attractive interaction between the monomers is included through an approximate arithmetic average overall amino acids, given by:

$$U_{short} = -\sum_{i<j} \frac{\varepsilon}{r^6}, \quad (6)$$

where ε was set to $0.6 \cdot 10^4$ kJÅ^6/mol, based on the work by Cragnell et al. [28], and sets the strength of the interactions based on the polarizability of the proteins. In this model, ε provides an attractive potential of 0.6 kT at closest contact.

The beads in the protein are connected by a harmonic bond potential, U_{bonded}:

$$U_{bonded} = \sum_{i=1}^{N-1} \frac{k_{bond}}{2} \left(r_{i,i+1} - r_0\right)^2, \quad (7)$$

where $r_{i,i+1}$ is the distance between two connected beads, r_0 with a value of 4.1 Å is the equilibrium distance, and k_{bond} with a value of 0.4 N/m describes the spring force constant, based on the work by Cragnell et al. [28].

The proteins are assumed to be flexible, where only the electrostatic interactions between the beads, as well as the volume of the hard spheres, contribute to the rigidity of the protein.

Supplementary B. 2 Monte Carlo Parameters

The coarse-grained protein model chain was inserted in a cubic box of length 1500 Å where periodic boundary conditions (PBC) were applied in all directions. The long-range Coulomb interactions were truncated using the minimum image convention at a cut-off length of 750 Å. Four different types of displacements were allowed: (i) translational displacement of a single bead, (ii) pivot rotation, (iii) translation of the entire chain, and (iv) slithering move, in order to accelerate the examination of the configurational space [96]. The probability of the different trial moves was weighted to enable single particle moves 20 times more often than the other three. Initially, the protein was randomly located in the box, and an initial equilibrium simulation of typical $2 \cdot 10^5$ trial moves/bead was performed for equilibration purposes, whereas the proceeding production run comprised $2 \cdot 10^6$ passes divided into 10 subdivisions. To confirm that the simulations were sampled accurately, the radius of gyration (R_g) probability distribution functions were analyzed. The uncertainty of the R_g values is based on the standard deviation of the total mean and the mean from the ten subdivisions, as described in Cragnell et al. [28].

Supplementary B. 3 Results

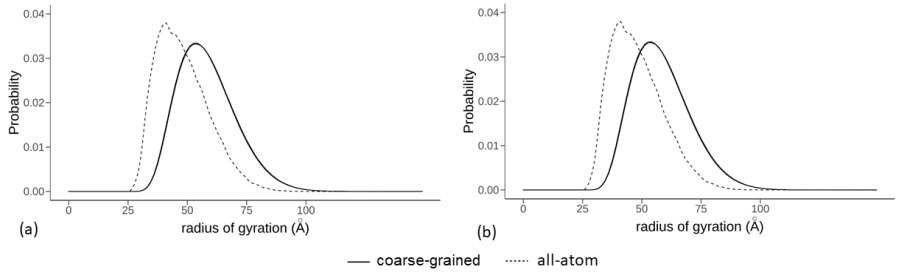


Figure 10. Probability distribution comparison between coarse-grained (solid line), and all-atom without disulphide bonds, dashed line, simulations for (**a**) α-gliadin and (**b**) low-molecular-weight glutenin. The coarse-grained simulations are four plotted curves representing the results from the salt treatment, they are perceived as one curve due to their high similarities.

The R_g distribution from the coarse-grained models shows that the values are larger in comparison to the all-atom simulations without disulphide bonds (Figure 10). Both types of simulation techniques describe that α-gliadin and LMW-GS are similar in terms of R_g.

Supplementary C

Table 2. Amount of collected configurations for each all-atom model at the corresponding temperature.

Temperature	α-gliadin with disulphide bonds	α-gliadin without disulphide bonds	LMW-GS with disulphide bonds	LMW-GS without disulphide bonds
300	80087	80000	80000	80088
309.74	80100	80066	80066	80286
319.8	80125	80100	80170	80500
330.2	80238	80206	80377	80581
340.9	80369	80415	80400	80633
351.99	80528	80544	80504	80720
363.42	80600	80600	80600	80800
375.22	80600	80676	80600	80800
387.4	80657	80700	80600	80800
400	80694	80694	80683	80793

Supplementary D

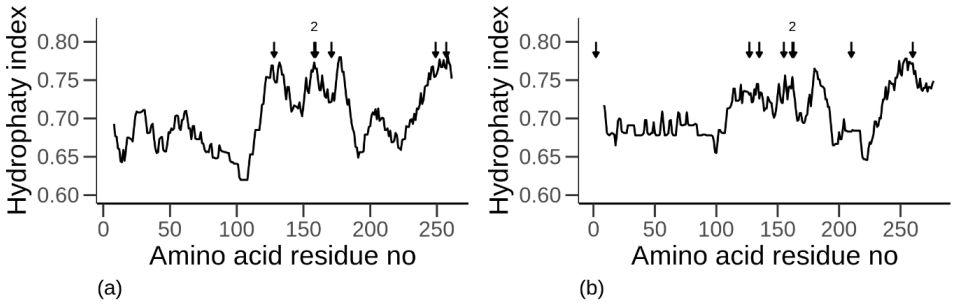


Figure 11. Hydropathy index (hydrophobicity for each amino acid with a window size of 15 amino acids) for each amino acid residue of (**a**) α-gliadin and (**b**) low molecular weight glutenin, indicating hydrophobic (+) and hydrophilic (-) regions in the protein. Cysteine amino acids are marked with an arrow, and the number 2 indicates two vicinal cysteines marked with two arrows at that position.

Figure 11 depicts the hydropathy index for the two proteins α-gliadin and LMW-GS, according to Rose et al. [56].

Supplementary E

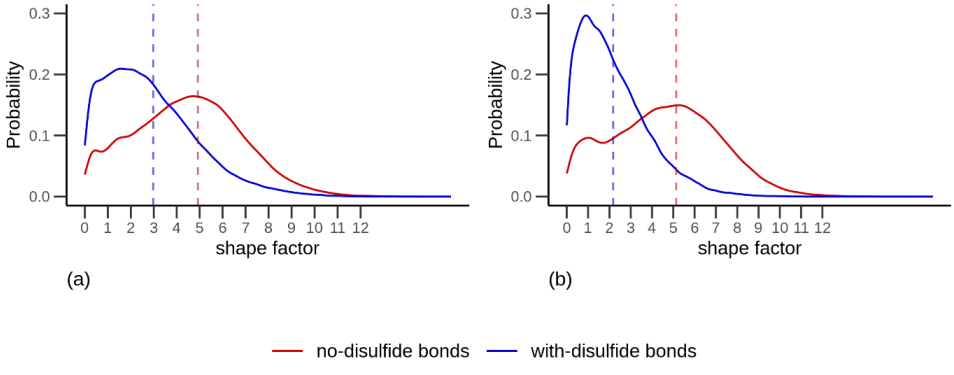


Figure 12. The shape factor of (**a**) α-gliadin and (**b**) LMW-GS, with disulphide bonds (red) and without (blue). Dotted lines highlight the mean values.

The shape factor distribution for α-gliadin and LMW-GS (figure 12), where low numbers indicate an average globular shape and high numbers indicate a rod-like shape.

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2022:10

Gluten proteins are found within the wheat seed, where they form extensive polymeric aggregates. This thesis used a combination of modelling methods and HPLC techniques to describe these proteins' assemblage into oligomers, how they form cross-links, and their water-absorbing capacity. Further, the modelling of gluten and other plant proteins was reviewed, describing that a few categories of plant proteins dominate the field and that modelling of gluten proteins is a pioneering field.

Joel Markgren received his doctoral education at the Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp. He received his MSc in Biology and Agriculture at SLU Uppsala.

Online publication of thesis summary: https://pub.epsilon.slu.se

ISSN 1652-6880
ISBN (print version) 978-91-7760-895-0
ISBN (electronic version) 978-91-7760-896-7