Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Plant Breeding

Modeling of plant proteins in order to optimize their properties in various application

Joel Markgren

Introductory paper at the Faculty of Landscape Architecture, Horticulture and Crop Production Science, 2017:1 •
Alnarp 2017

Modeling of plant proteins in order to optimize their properties in various application

Joel Markgren
joel.markgren@slu.se

Place of publication: Alnarp
Year of publication 2017
Cover picture: Gunilla Markgren
Title of series: Introductory paper at the Faculty of Landscape Architecture, Horticulture and Crop Production Science
Number of part of series: 2017:1

Online publication: http://epsilon.slu.se
Bibliografic reference:
Markgren, J. (2017). *Modeling of plant proteins in order to optimize their properties in various application*. Alnarp: Sveriges lantbruksuniversitet. (Introductory paper at the Faculty of Landscape Architecture, Horticulture and Crop Production Science, 2017:1).

Keywords: a-gliadin,gliadin, gluten, modeling, proteins, protein, protein modeling, molecular dynamics, md, monte carlo, mc, quantum mechanics, qm, simulated annealing, simulation, ab-initio, folding, energy landscape

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Plant BreedingPublisher

Abstract

Proteins are one of the most dynamic biological macromolecules known on earth. The proteins can alter their properties depending on how they are treated, since they are big chain like, flexible molecules. An example of a protein that can change properties is egg white, which changes color, texture and taste when cooked. In this research project the focus will be on the protein α-gliadin, a sub protein to the gluten protein complex which is known to contribute to fluffy breads when baking. The α-gliadin is interesting since it contains several cysteine residues which under certain conditions are able to form polymers through disulphide bonds. If one could estimate the properties of α-gliadin at different treatments, one would also know its worth for various applications both for food and non-food applications. Using the results of this research for knowing how to produce high quality bio-based plastics is only one of many possible outcomes. To avoid conducting time consuming and material costly lab experiments, this paper will try to describe the possibilities of how to model the protein structure-function relationship and simulate all experiments. Estimating the behavior and properties of a protein often require an estimation of how the atoms in the protein alter positions at the experimental condition. There are several ways of doing that. Estimating the atomic movements due to inter/intra molecular forces, is the way of the *Molecular dynamics* approach. To randomly alter the proteins´ conformations, and seek for the lowest energy potential is the Monte Carlo approach. Estimating the electron and nuclear movements is the Quantum Mechanic approach. Initially, one would go for the most thorough approach, but unfortunately this is seen as impossible due to limitation of computer power when this paper was written. To explore α-gliadins potential properties, one would probably need to combine several known approaches.

Table of contents

List of figures

Abbreviations

Tabell 1. *Abbreviations found in this introductory paper and their corresponding meaning*

Abbreviation	Meaning
LV	Las Vegas
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
NPT	constant amount of particles, pressure and temperature
NVT	constant amount of particles, volume and temperature
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics

1 Introduction

Proteins are flexible organic molecules that can shift conformation and as a result also shift properties, like for example thermal conductivity, solubility and affinity for chemical reactions. This works just as a Swizz army knife, where the tool can change conformation from a cork screw opener to a saw. This capability enables a world of possibilities, if you know how to alter the conformation of your desired protein.

One protein with specific potential is the α-gliadin protein; a sub protein to the gluten proteins found in the grains of wheat plants. This protein can connect to its adjacent proteins and form large networks. These protein networks might be used as polymer like plastic, and are currently studied for these purposes in e.g. a VR-funded project (Johansson 2013). The goal of the research is not only to form plastics out of α-gliadin, but also to investigate the potential of plant protein properties in general.

Investigating the α-gliadin proteins´ chemical potentials can be time demanding if done empirically in a lab. In this project, the idea is to make the experiments in silico by simulating the experiments on computers. In order to make good simulations one needs to model the chemical and physical behaviors of the atoms involved in the protein. In this paper, I will review some of the available methods in the field like molecular and quantum dynamics where one tries to mimic the movements of atoms and electrons. The review will also include the Monte Carlo approach where one identifies a molecule´s most probable position and the Las Vegas approach which increases the accuracy of other methods.

As a fresh PhD student new to the field I write this paper as a part of the course introductory paper, a course meant to provide insight in the field. The work of this paper will be used when planning the initial experiments of my research.

2 Aims and disposition

The aim of this paper is to answer these questions:

“How are functional properties of proteins related to their conformation and structure, and is it possible to determine the function with computer based simulations?”

The article is ordered in the following way:

The paper starts with a basic introduction of protein science, followed by the hypothesis of how a protein is folded, how to order these theories into models and how to validate the theories. Finally, all is summed up with a discussion and conclusions.

3 Background

3.1 Protein composition

Proteins are chemically composed by amino acids linked together by peptide bonds in a chain like pattern and there are often around 20 different standard amino acids found in proteins listed in Horton et al. (2006) and Bränden and Tooze (1999). In addition to the standard amino acids, proteins can be composed of unusual non-standard or synthetic amino acids. The amino acids are in turn composed of a hydrogen atom, an amide and a carboxylic acid group, together with a side chain all bonded to a central carbon atom seen in **figure 1**. The bonds between the central also called "α" carbon and the amide and carboxyl groups (called ψ and φ), illustrated in **figure 1,** and are highly flexible. These two bonds are capable of twisting their angles at 360 degrees, allowing protein changing conformation, in contrast to the peptide bond that is more stiff (Chang 2007).

The flexibility of proteins allows them to fold into three-dimensional shapes, and is often described according to the following four-points of resolution:

- Primary structure: The amino acid sequence (the sequence of how the residues are lined after each other)
- Secondary structure: Local spatial formations of the protein chain
- Tertiary structure: Spatial formation of the entire protein or a larger domain
- Quaternary structure: Spatial relation to surrounding proteins

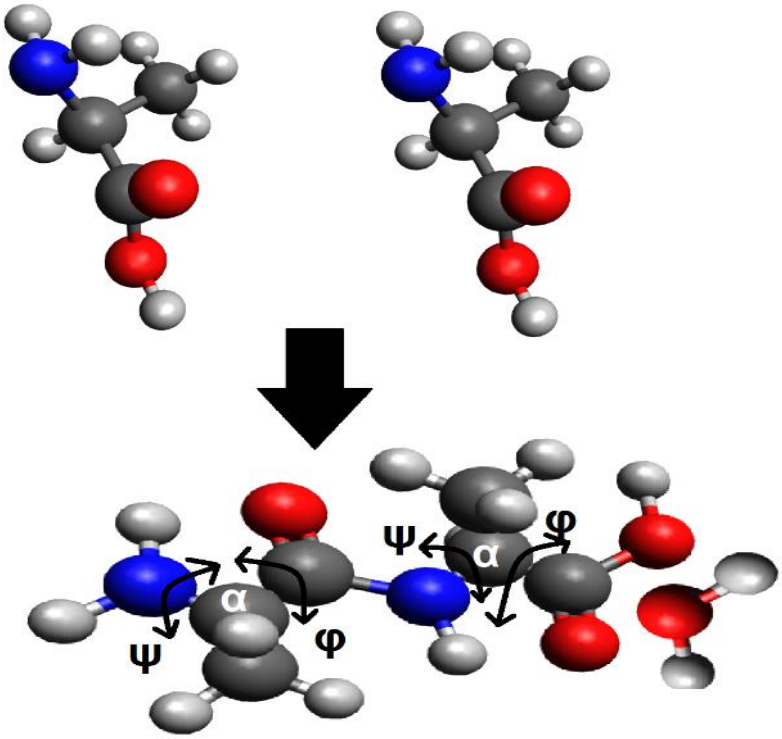

Figure 1. , Illustrates how two random amino acids react and form a peptide bond through condensation (Joel Markgren after Chang 2007, page 1046). The peptide bond is illustrated in the red circle. In the reaction, a water molecule is also produced. The alpha carbons in the peptide chain are marked up with the "α" symbol. The bindings next to the alpha carbon "φ" and "ψ" illustrated with the same symbols in the picture, have the possibility to change angles. The "?" symbols indicate the presence of an amino acid side chains.

The peptide bond linking the amino acids is formed when the carboxylic group and the amide group of two different amino acids condense in an energy demanding reaction creating the bond and a water molecule, illustrated in picture 1 (Horton et al. 2006, Bränden and Tooze 1999). The properties of the amino acids are dependent on their side chain and neighboring amino acids.

3.2 Molecular forces

The individual residues affect each other in a protein, due to inter molecular forces and in some instances, also intra molecular forces. Inter molecular forces affect the relation among molecules, compared to intra molecular forces that affect the individual atoms within a molecule.

3.2.1 Dipole-dipole forces

The electrostatic interactive dipole-dipole force, seen in **equation 1**, is formed between two polar particles (Israelachvili 2011). If the particles´ polarity wear charges of unequal type, they form an attraction toward each other, but if the charges are of equal type, they are repelled. The strength of the energy between the particles depends on the distance between them. Like all the electrostatic forces is it also very similar to the force of Coulomb´s law seen in **equation 2**, which functions in a similar manner but is stronger (Chang 2007, Aims, 1953).

$$\boldsymbol{E} = -\frac{\mathbf{1}}{\mathbf{4\pi\varepsilon}}\frac{\overrightarrow{U_1}\ \overrightarrow{U_2}}{r^3} \qquad \mathbf{1}$$

$$\boldsymbol{F} = \frac{\mathbf{1}}{\mathbf{4\pi\varepsilon}}\frac{Q_1\ Q_2}{r^2} \qquad \mathbf{2}$$

Equation 1 and 2, Dipole-dipole forces and Coulomb´s law, where F is force in Newton (N), Q_x charge of a particle in Coulomb (C), r is distance between particles in meter (m), ε is the product of the relative permittivity constant for the specific medium the particles are interacting in and vacuum permittivity constant (if the interaction is in vacuum, only the vacuum permittivity constant is used) in square Coulomb per Voltmeter for particles (C Vm^{-1}), U is dipole moment in Coulomb meter (C m), E is energy in Joule (J)

3.2.2 Hydrogen bonds

The hydrogen bond is a mixture of dipole-dipole forces and partial covalent bonding. The bond often occurs between polar molecules and more precisely between a hydrogen atom bonded to a highly electronegative atom and a nearby highly electronegative atom. The hydrogen atom in the bond partly shares electrons and is simultaneously attracted to its hydrogen bonded counterpart through dipole-dipole forces (Arunan et al. 2011a, Arunan et al. 2011b).

3.2.3 Hydrophobic effect

When dripping oil in a glass of water, one can see that the oil droplets avoid mixing with the water and almost seem to search for other oil droplets they can merge with. The oil droplets´ avoidance of mixing with the water is called the hydrophobic effect and hinders non-polar agents to mix in polar solvents. The molecules in a polar solvent are linked together by hydrogen bonds like a huge fishing net. On the other hand, the non-polar agent is incapable of forming hydrogen bonds

and instead of becoming a part of the metaphorical fishing net by polar atoms, they find themselves captured in it (Lins and Brasseur 1995).

3.2.4 Van der Waals forces

The three Van der Waals forces; Debye forces and Keesom interactions, London dispersion forces seen in **equation 3-5**, are also similar to Coulomb´s law, where a force is formed between two charged particles.

$$\boldsymbol{E} = -\frac{\boldsymbol{U_1^2 a_2 + U_2^2 a_1}}{\boldsymbol{(4\pi\varepsilon)^2 r^6}} \quad \textbf{3}$$

$$\boldsymbol{E} = -\frac{\boldsymbol{U_1^2 U_2^2}}{\boldsymbol{3(4\pi\varepsilon)^2 k_b T r^6}} \quad \textbf{4}$$

$$\boldsymbol{E} = -\frac{\boldsymbol{3a_2 a_1}}{\boldsymbol{2(4\pi\varepsilon)^2 r^6}} \frac{\boldsymbol{h v_1 v_2}}{\boldsymbol{(v_1 + v_2)}} \quad \textbf{5}$$

Equation 3,4,5, Debye forces, Keesom interactions, London dispersion force , a is the polarizability of a particle in Coulomb, square meter per Volt (C m^2V^{-1}), h is the Planck's constant in Joule seconds (Js) and v_x is a particle´s Ionization frequency in Hertz (Hz), Kb is Boltzman´s constan Joules per Kelvin ($1.381*10^{23}$ J/K) T is temperature in Kelvin, U is dipole moment in Coulomb meter (C m), ε is the product of the relative permittivity constant for the specific medium the particles are interacting in and vacuum permittivity constant (if the interaction is in vacuum, only the vacuum permittivity constant is used) in square Coulomb per Voltmeter for particles (C Vm^{-1}).

In more detail, The Debye forces are the interaction between particles with an induced polarity by surrounding particles, **equation 3** (Israelachvili 2011, page 99). The Keesom interactions occur between two permanent dipole particles, **equation 4** (Israelachvili 2011, page 85). The London dispersion force is described as dipole-dipole interactions formed by spontaneous and random polar formation in molecules, **equation 5** (Israelachvili 2011, page 107-109).

3.2.5 Pauli exclusion

How close the molecules can come to each other before they collide are approximated by the Lennard-Jones potential, seen in **equation 6**, where the molecules are attracted by van der Waals forces and repelled by the Pauli Exclusion Principle (Jones 1924, Israelachvili 2011). The Pauli Exclusion Principle hinders the possibility for protons or electrons (fermions) to occupy the same quantum state around the same atom (Pichler et al. 2015).

$$\boldsymbol{E = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]} \quad \textbf{6}$$

Equation 6, Lennard-Jones potential, where σ is the distance when the potential forces are at minimum (m), € is the maximum potential (J).

3.2.6 Disulfide bonding

The disulfide bond is formed between cysteine residues, if they are close to each other in an oxidative environment, see formula in **equation 7.** More precisely, the bond is between the sulfur atoms in the involved cysteine residues (Bränden and Tooze 1999).

$$2 - CH_2SH + 0.5O_2 \leftrightarrow CH_2 - S - S - CH_2 + H_2O \quad \textbf{7}$$

Equation 7, Description of disulfide bonding, where 2-CH_2SH are two individual cysteine residues that are unlinked, $0.5O_2$ indicates an oxidative environment, -S-S- is a disulfide bond. The reaction can be conducted with other oxidative reagents than oxygen.

The disulfide bond can link parts that are distant in the primary structure but are close in the secondary or tertiary structure. The disulfide bonding can also link two different protein chains together, forming a quaternary structure.

3.3 Secondary structure motifs

Proteins form many different shapes, since many parts of the protein are movable. In fact, all atoms in the protein are movable to some extent, and can change their torsional, dihedral and constraint bond angles to their connecting atoms. The meaning of the three angle types are described below:

- The torsional angle is the way an atom is allowed to twist itself
- The constraint angle is the way the atom is allowed to move away or approach its neighboring connected atom
- The dihedral angle is the way the atom is allowed to move at any direction that do not involve twisting or a shift in distance to the adjacent connected atoms

The conformation of the protein is most impacted by the dihedral ψ and φ angles associated to the α-carbon in the individual residues. Motif resembling structures can appear in the protein depending on the conformation. In many cases, these motifs can be found or resemble motifs in other proteins (Bork 1996), (Richardson 1994). Even if it is common with motif structures in proteins, there are some proteins without any clear motif which are random structures.

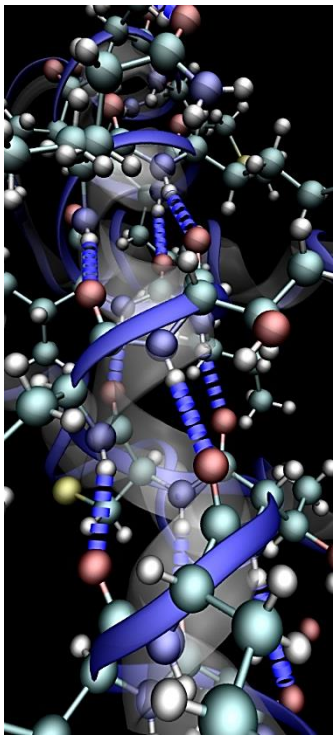

Figure 2. An illustration of an α-helix in Granulysin from human cytolytic lymphocytes (Joel Markgren after Anderson et al. 2002). The dotted lines are hydrogen bonds between oxygen and hydrogen atoms. The transparent spirals illustrate how the chain is like a helix structure. The blue band illustrates the protein´s backbone structure.

The three most common motifs are the α-helix, β-turns and β-sheets, and they will be described briefly together with the loops and the Zink fingers. There are also many more motifs then those mentioned here, for example the "nests", "β-hairpin" and "coiled-coils, but they are omitted in this review since the point is only to briefly describe proteins´ ability to form structural motifs.

3.3.1 The α-helix

The **α-helix**, seen in **figure 2**, is a staircase resembling structure, first observed by Pauling et al. (1951). The shape is formed when the protein backbone makes a turn every 3,6 residues. The conformation is stabilized due to the formation of hydrogen bonds every fourth residue, between the oxygen atoms in a carboxylic acid and the hydrogen atom in an amide group. The structure gains extra stability if the side chains of the involved residues are small, since large side chains increase the risk of collision with neighbouring residues (Kamtekar and Hecht 1995).

There are three variants of the **α-helix** helixes, the 3.6, π and 3_{10} helixes. The 3.6 is described above and is considered as the "normal" type. The 3_{10} helix is a tighter version, where the helix makes a turn every 3.2 residues and a stabilizing hydrogen bond occurs every third residue (Toniolo and Benedetti 1991). The π helix is a more relaxed structure, turning every 4.4 residues, but the hydrogen bond only occurs every fifth residue, making it compared to the other helical types rather unstable (Riek and Graham 2011).

3.3.2 The β-sheets

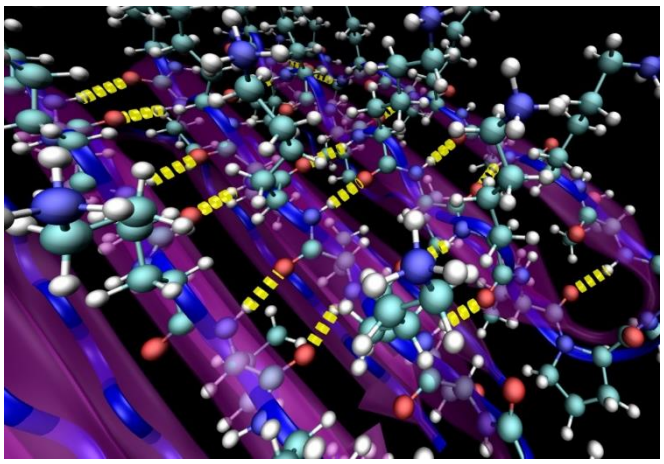

Figure 3. An illustration of a β-sheet in an MAX1 peptide fibril (Joel Markgren after Nagy-smith et al. 2015). The yellow hydrogen bonds can be seen between the red oxygen and the white hydrogen atoms. The teal atoms are carbon, the blue are nitrogen. The purple transparent arrows illustrate the shape of the β-sheet. The blue band illustrates the protein´s backbone structure.

The **β-sheets** are flatbed looking motifs. They are constructed of the protein backbone lying side by side with itself or other proteins connected by hydrogen bonds. More specifically, the hydrogen bond is between the hydrogen from the amide groups and the oxygen from the carboxylic groups in the adjacent protein chain as illustrated in **figure 3**.

3.3.3 The Loops

The loops are structures named after their motifs. One special loop is the **omega loop**; a motif structure that is similar to the Greek Ω (omega) symbol. The omega loop is often found in active domains and binding sites of proteins (more explained in section 3.5. “Protein activation”), and is often stabilized by irregular hydrogen bonding between the residues (Fetrow 1995).

3.3.4 The Zinc fingers

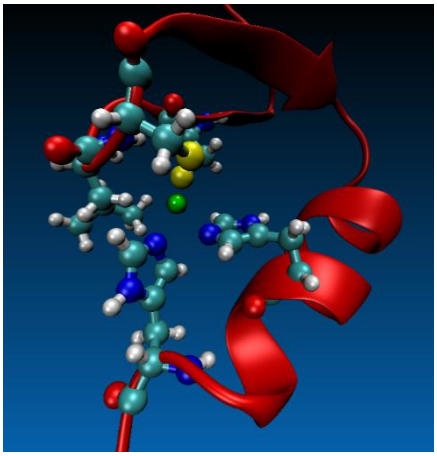

Figure 4. A Zink finger motif with a zinc ion (the green ion) fixed in a three-Cys2His2 domain. Found in a mouse testis zinc finger protein (Joel Markgren after Chou et al. 2010).

Some motifs involve non-organic components, like the Zinc finger motif, where a zinc ion or another metal ion is fixed in a small domain, seen in **figure 4**. The zinc ion provides structural strength to the domain, resulting in a stabilization of fragile motifs in active domains (Klug and Schwabe 1995).

3.4 Folding

A protein folds into its structural conformation, due to intra and inter molecular forces from the protein structure itself or its surroundings. The movements occur when the individual atoms are accelareted by the affecting forces according to Newton´s second equation, **equation 8**.

$$\boldsymbol{F = ma} \qquad \textbf{8}$$

Equation 8. Newton second law, explaining how a force is composed of accelerating matter, where m is Kilograms (Kg) and a is acceleration (Kg s^{-2})

The folding movements will go on until the protein is in a steady state where the protein backbone conformation no longer alters. The necessary force for moving atoms that can alter the comformation are then higher than the gain of new movements, also called the native state (Sali et al. 1994).

The energy needed for the folding reactions are governed by the thermodynamical rules of Gibb´s free energy, **equation 9**. If a chemical reaction releases a lot of energy, it will result in a stable compound that is very unlikely to return to its former compound without an energy influence from the surrounding environment (Chang 2007).

$$\Delta G = \Delta H - T\Delta S \quad \mathbf{9}$$

Equation 9. Gibbs free energy, where G is Gibbs free Energy in joules per mol (J/mol), S is entropy measured in Joules per Kelvin (J/K), H is entalphy in Joule per mol(J/mol).

An example of the thermodynamic rules of Gibb´s free energy is that ashes from a burned piece of wood are unlikely to spontaneously reverse the combustion process and again become a piece of wood. In protein folding, this means that large movements that release large amount of energy, are less likely to be reversed. A parable to the thermodynamic process in protein folding is to visualize a golf ball in a landscape like the one in **figure 5**. If you push the ball down from a slope it will roll down toward the bottom of the slope, and convert its potential energy to movements. The protein will fold in a likewise manner where it starts off in a high potential energy state and starts moving until it reaches its native state. In reality, as it is pictured in **figure 5**, the path down for the golf ball and also for the protein folding, the way is filled with obstacles. And sometimes, a protein might not manage to fold itself without help from outside (Dobson 2003, Alm and Baker 2003, Sali et al. 1994, Dill and Chan 1997, Coluzza 2015).

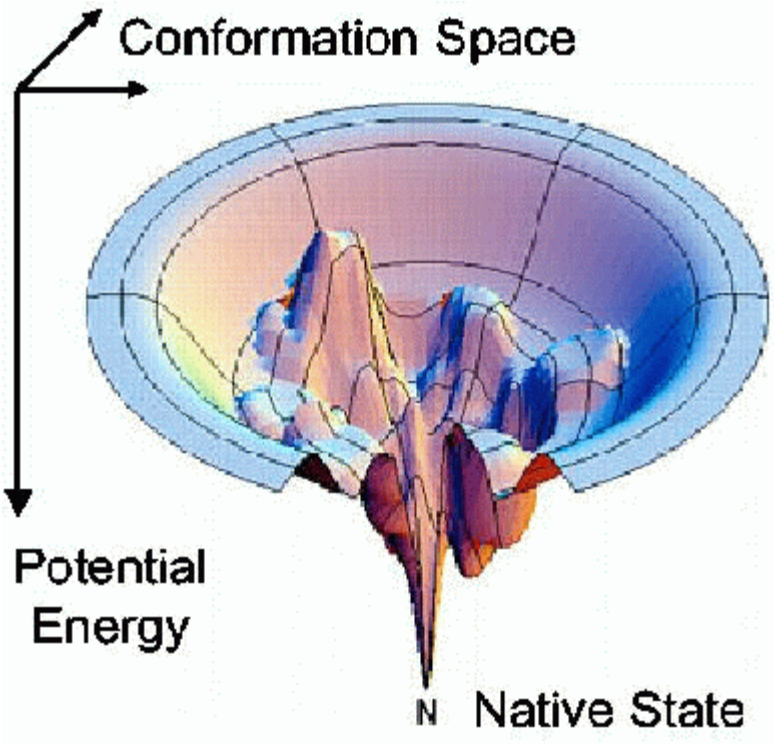

Figure 5. Esthetic visualization of the folding energy landscape for a protein. Where the folding starts at the high altitudes and rolls down toward the lowest point. Figure from Dill and Chan 1997.

An obstacle for the folding of a protein is its own backbone. There are φ and ψ angles for some residues that are not possible in certain conformations in a protein, since these angles make the protein collide with itself. The allowed movements of φ and ψ angles can be mapped out by a Ramachandran plot (Ramachandran et al. 1963). If we think of our golf ball parallel again (that we put on the edge of a slope), one could see this obstacle as high walls forming a maze pattern in the slope down to the absolute bottom.

A protein might not fold in the same way if its folding procedure would be repeated in an identical environment as the first time, although it still would most probably reach its native state (Levinthal 1968, Zwanzig et al. 1992). One reason why it would not repeat its folding is due to the energy of the surrounding environment, can cause random movements in the fold. An energy rich environment can transfer energy to the fold and enable the protein to fold toward less likely conformations, in the same manner as a golf ball is more likely to roll uppwards a slope if pushed by a strong wind. The way surrounding environmental energy affects protein conformation is described by the Boltzman distribution **equation 10**, where a high temperature increases the probabilities for less likely and more energy demanding energy states to occur (Sali et al. 1994, Banushkina and Krivov 2015).

$$\boldsymbol{P_i} = \frac{e^{-(\frac{\varepsilon_i}{kT})}}{\sum_{i=1}^{M} e^{-(\frac{\varepsilon_i}{kT})}} \qquad \textbf{10}$$

Equation 10. Boltzmann´s distribution. Explains the probabilities of energetic states due to temperature and terms of energy, where P_i is the probability of a certain state, ε_i is the energy of a state in Joule or electrovolts (J, eV), M is the amount of possible states.

It is not necessary that the folding goes directly to the native state, sometimes it pauses at specific intermediate states, only to continue after an unknown amount of time. The pausing behavior is due to the fact that intermolecular forces are not strong enough to continue folding and the energy from the surrounding is not sufficent to reverse the folding. Intermediates are often found when large sections of a protein have completed folding, and another section is about to start (Bai et al. 1995, Banushkina and Krivov 2015). To exit the intermediate state, the protein switches conformation through obtaining energy from the surrounding until a distinct force is formed that continues the folding (Dobson 2003). Back to our dropped golf ball parallel in the landscape of **figure 5**, one could say that these intermediate states could be seen as flat regions where the ball does not find a place to roll downward, and is only slowly pushed by the wind until it finds a slope and can roll downhill again.

A protein might fail reaching the native state. The protein fails if it gets trapped in an intermediate state or an energy minimum similar to the native state (also called

alternative state) (Tokuriki et al. 2009). Leaving these trapped states will require energy from the surrounding, in order to unfold the protein and refold it correctly. In some instances, these traps are so severe that the protein will probably never exit them, and the protein is considered as misfolded. A misfolded protein can in some instances start a reaction called aggregation and form so called fibrills, which are proteins composed of a β-strand structures (described more in section 3.5.4. "Protein-protein interactions, chaperons and fibrils") (Dobson 2003).

3.5 Protein activation

In many instances, a protein is like a small mechanical machine. Chemical components can react with reactive sites in proteins, resulting in structural changes within the protein. These changed proteins can in turn react with another chemical or protein, and then return to their original conformation.

In the field of biology, the science of how proteins react with other compounds is a hot topic. Many biological processes are regulated by different type of compounds commonly called ligands. The ligands trigger a structural change of proteins which in turn alter the interactions among cellular components. By regulating the ligands, it is possible to regulate intra cellular interactions.

3.5.1 Ligands and pharmacophores

A protein can change its structural conformation, when reacting with a ligand. The conformational shift is due to a search for a new native state since the ligand alters the conditions for inter molecular bonds in the protein (Kobilka 2007, Russo et al. 1996, Lambright et al. 1994).

The reactive site of a protein where the ligand binds, are specific domains where there is a surface with residues that bind to the compound, like a key to a lock (Kobilka 2007).

The ligand itself can be of any possible compound that can react with a protein, for example metal ions, proteins and chemical substances. The ligands do not necessarily have to be a molecule that naturally regulates the protein activity. It is possible to replace them with other compounds, so called pharmacophores. Many medical drugs are often pharmacaphores due to the body lacking natural ligands. However, a pharmacaphore might not always accomplish an identical protein activation as the ligand, partly explaining why some drugs do not have full effect (Kenakin 2001).

3.5.2 Light and electricity's activation

Light and electrical impulses can also shift protein structural conformations, and thereby work as ligands. Classical examples of this are the photoreceptors in the human eyes that manage to change conformation at different types of light, and also the nervous system that sends electrical shocks for rapid signaling between many cell groups distantly scattered within the human body.

A structural shift occurs, if a photon with the right wave length excites one or several residues within the protein that are involved in bindings important for the native structure. This makes the electrons within the individual residue to alter positions and thereby their inter-molecular properties are altered as well, making the protein find a new native state. The protein is in most cases capable to shift back to the original native state when the excited residues relax to normal state (Pletnev et al. 2012).

An electric current can create electromagnetic fields within a protein, leading to increased dielectrically polarization of the residues. The polarization makes the residues become more repellant or attractive to each other and this can result in a loss of intra molecular bonds and a search for a new native state. Some of the secondary structure motifs, like the α-helix, work as active sites for electrical impulses, since they can function as inducible coils and thereby increase the electromagnetic field. Not only are the electromagnetic fields affecting the individual protein, they can also affect surrounding proteins especially if they have large domains of α-helical structures (Monajjemi 2015).

3.5.3 Solvents effect on protein

In polar solvents like water, hydrogen bonds are formed between the solvent and polar residues on the protein. Polar solvents also result in polar residues on the protein surface whereas nonpolar residues are found tightly packed in the center forming a hydrofobic core (Cheung et al. 2002). In non-polar solvents, the protein reacts in the opposite way, and the polar residues are faced inwards forming a hydrofilic core, or as in some cases, forming α-helical structures, where the charged sides of the polar residues are facing into the helix (Pace et al. 2004).

The salt concentration of a solvent influences the stability of a protein depending on salt type and concentration. The salt ions´ ligand bind to the protein and either stabilize or destabilizes the native state (Chi et al. 2003).

pH regulates the protonation of the protein, and can alter the chemical properties of individual types of residues. The pH influenced residues can then alter intra molecular bonding properties and the protein might switch to an alternative native state (Chatterjee 2015).

In general, a solvent filled with many other proteins or larger molecules, effects the protein structure by the “crowding” effect, where the molecules will collide and attract one another in a random pattern. The collision effects are often seen as a force that enhance the native state by squeezing the protein, as long as the collision is not too strong and literally breaks the protein bindings. Attractive forces are seen as destabilizing, since they are pulling the peptide chain and therefore promote unfolding (Wang et al. 2012).

3.5.4 Protein-Protein interactions, chaperons and fibrils

Proteins can dock into each other and work as ligands, like any ordinal chemical compound. The protein-protein interactions can quickly lead to cascade effects. In **figure 6**, the start of a cascade is illustrated, where protein 1 interacts with protein 2 leading to inhibition of the reactions between protein 2 and 3 and a facilitation between protein 1 and 4. This will in turn affect several other protein-protein interactions.

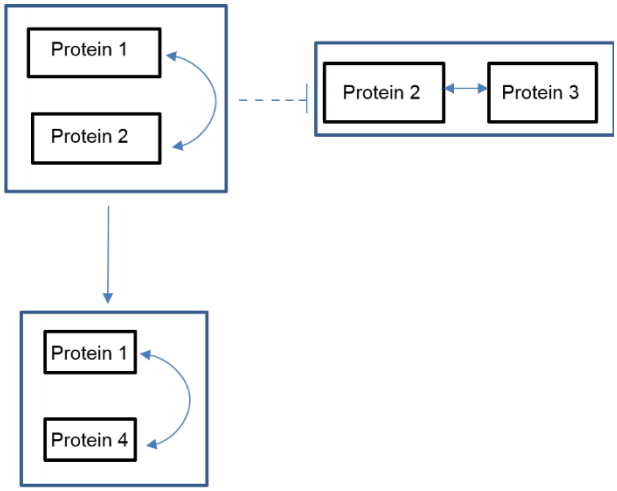

Figure 6. Illustrating the relations between proteins, how the interactions between two proteins can facilitate or inhibit the reaction between other proteins.

Two extreme cases of opposite types of protein-protein activators are the amyloid-fibrils and the chaperons, where the first one totally destructs a protein´s conformation and the other one attempts to refold it correctly.

The amyloid-fibrils are proteins dominated by β-sheet structures, that look like a fibril. They are often the result of a protein that has misfolded. These fibrills can affect surrounding proteins by degrading them to new fibrils, where the protein folding is directed in a competitive reaction resulting in major faulty structural changes (Neudecker 2012). The fibrils are known to be extremly cytotoxic and are often responsible for several diseases like Alzheimers, Creutzfeldt–Jakob disease and cancer. They are often very stable, and can propagate at rapid speed in cellular systems, making them difficult to counteract with medical drugs (Dobson 2001, Bieschke et al. 2004).

On the other hand, chaperones facilitate the protein reaching its natural state. This is thought to be done by several mechanisms. One way of facilitating, is by shielding the folding protein from the surrounding environment and thereby hinder misfolding. The chaperone is also capable of unfolding regions of β-sheet structures in the treated protein and letting the protein refold, in order to minimize the risk of the protein being misfolded into a amyloid-fibril (Zhang and Kelly 2014). Chaperones are often most active in stressful conditions, for example during heatschocks where proteins are more prone to alter their conformations. The disadvantage is that high levels of chaperones hinder apoptosis, and are therefore partly responsible for making cancer tumors prone toward drug resistance in chemotherapy treatments (Mayer et al. 2005).

3.6 Protein denaturing

Denaturation is a process where the secondary and the tertiary structures of proteins often are severely damaged, sometimes so severe that the protein forms a random coil (Krishnamani et al. 2012). However, the primary structure is still intact, but many of the binding forces have been disrupted and disulphide bonds might thereafter be established at random positions locking the protein from folding. In addition, the denaturised proteins can start form bonds with neighbouring proteins in a random pattern, resulting in a mess.

IUPAC provides a more precise explanation of denaturation in a paper of Vert et al. (2012):

> "Process of partial or total alteration of the native secondary, and/or tertiary, and/or quaternary structures of proteins or nucleic acids resulting in a loss of bioactivity.
>
> Note 1: Modified from the definition given in [2].
>
> Note 2: Denaturation can occur when proteins and nucleic acids are subjected to elevated temperature or to extremes of pH, or to nonphysiological concentrations of salt, organic solvents, urea, or other chemical agents.

Note 3: An enzyme loses its catalytic activity when it is denaturized."- Vert et al. (2012)

The properties of a denaturised protein differ widely from a native one. An example of this is when an egg is boiled. The high temperature denaturizes the proteins in the egg resulting in colour, taste and texture changes. The denaturation result can differ depending on the denaturation agent, where for example a physical force can produce a denaturized protein that is less distinguishable from the native state compared to when the protein is denaturized by a chaotropic agent like urea which disrupts the hydrogen bonds (Stirnemann et al. 2014).

A boiled egg does not go back to a liquid one if it is put back in the refrigerator, indicating that the denaturized structure is stable. The egg does not spontaneously become uncooked, but this could be done if vortexed together with an optimized amount of a chaotropic agent like urea or ethanol (Yuan et al. 2015). The intermolecular bonds and disulphide crosslinks between the different peptide chains are broken by both the centrifugal forces in the vortex and the chaotrich agent. The individual peptide bonds are then folding to a native state, since the centrifugal forces are not strong enough for hindering a folding of the individual protein.

4 Modeling

With a model, one can attempt to predict the outcome of situations that so far are not tested, in the same manner as aerodynamics properties are tested with small airplane models in air tunnels before one attempts to build the big airplanes. In protein structural science, mainly two types of methods are used for modeling protein folding; the molecular dynamics and the Monte Carlo methods. Alone, they are often rather rough methods, but they can many times be improved by the Las Vegas and the Quantum mechanical methods, all described together with a few alternative methods in the following sections.

4.1 Molecular dynamics

The molecular dynamics approach (**MD**), intends to model the protein folding by iteratively solving the Newton's second law of motion for all atoms (**equation 8**). With one of the common molecular dynamics software "GROMACS", the ways of integrating the movements with one of either leap-frog or velocity Verlet integrator algorithms (Abraham et al. 2016). Where the leap-frog algorithm basically updates positions at time step t and velocities at time step t - $\frac{1}{2}\Delta t$ by computing the force. In the velocity Verlet algorithm one updates positions and velocities by calculating the force during the same time step. Estimations on acceleration, speed and directions are done for all atoms. Behaviour properties like intermolecular forces and the allowed torsion, dihedral and constrain angles for all molecules and atoms, are provided by a parameter file called force field. There are several different types of force fields available, adjusted for different kinds of molecules. If necessary, in most cases, it is possible to modify the force fields (Brooks et al. 1983, Adcock and McCammon 2006).

A protein system often requires preparation before it can be simulated properly in MD. This involves energy minimization, where one tries to avoid stearic clashes

in a system followed by constant amount of particles, volume and temperature (**NVT)** and constant amount of particles, pressure and temperature (**NPT**) equilibrations where one attempts to invoke desired temperature and pressure to the system.

4.1.1 Energy minimization

When energy minimizing, one moves atoms in the system in order to avoid steric clashes and detect an energy minima for the simulation to explore (Adcock and McCammon 2006). Examples of minimization methods used in the search are the Newthon-Rapson method for optimisation together with steepest descent methods. The Newton-Rapson method illustrated in **equation 11** seeks for an energy minima/maxima or saddle point for every iterative step, by estimating the gradient and hessian of the curve through the Taylor´s theorem, which is a theorem methodology, where it is possible to estimate the curve (Press et al. 1997, mathematicalmonk´s 2011). The steepest descent, seen in **equation 12**, on the other hand, has a focus on following the slope while iterating. This is sometimes a time consuming, but often also robust method (Abraham et al. 2016, Ng 2013). There are several more optimisation iteration methods that can be used like gradient descent or the biconjugate gradient, but in this case I chose only to describe the Newton- Rapson and steepest descent to illustrate the iteration process.

$$\mathbf{x_{t+1} = x_t + \Delta x \rightarrow 0 = \delta/\delta \Delta x (f(x_n) + f'(x_n)\Delta x + 0.5 f''(x_n)\Delta x^2 = f'(x_n) + f''(x_n)\Delta x} \quad \mathbf{11}$$

$$\boldsymbol{x_{t+1} = x_t - y_t f'(x_n), t \geq 0} \quad \mathbf{12}$$

Equation 11& 12, respective describing the Newton-Rapson optimization method and steep decent method, where t is current time step, t+1 is next time step, ΔX is the difference between current and next time step and γ is a parameter for controlling the size of each iteration step.

4.1.2 NVT and NPT equilibration

An NVT simulation basically invokes movements in all atoms in the system, until they reach the desired temperature. The invoked movements are initially random and are later regulated up or down by a thermostat algorithm. The temperature itself is estimated by averaging all atom movements in the system (Andersen 1980).

The system is allowed to change its volume depending on a pressure barostat algorithm in NPT simulations after NVT. This means that the walls of the box are allowed to move to form a desired pressure in the system (Andersen 1980).

The equilibration steps might at first glimpse seem to be very basal, but it takes time for a system to be equilibrated properly. If the system is poorly equilibrated, the temperature and/or pressure of the system can start to misbehave. They could for example start escalating or become unstable with large fluctuations leading to an increased risk of artefacts in the result.

4.1.3 Shortcomings and shortcuts with MD

The time required for MD simulations to find native state or alternative native states, depends on what conformation state it starts with, how large the protein is and the system setup (e.g type of solvent, temperature interacting molecules etc.) (Adcock and McCammon 2006).

In case the availability of computational power is too low, the model can be simplified. One way of doing this is by a procedure called coarse graining, where side chains on the residues are modelled as a single atom or groups of residues are modelled as one residue (Adcock and McCammon 2006). Another way of simplifying is by letting intermolecular forces be adjusted with a bias toward a known conformation (Paci et al. 2002).

4.2 Quantum mechanical approaches

Extremely thorough estimations of molecular movements and chemical reactions are possible by estimating the positions of the nucleus and electrons in the protein through quantum mechanics (QM). The estimations are done by solving the time dependent Schrödinger equation seen in **equation 13,** which describes particles in the form of a wave equation depending on time. The information from the equation can then be used for estimating the probabilities to find an electron and a nucleus at a given place, and its momentum (Brunk and Rothlisberger 2015).

$$\mathbf{i}\hbar\left(\frac{\boldsymbol{\delta}}{\boldsymbol{\delta t}}\right)\boldsymbol{\Psi}(\mathbf{r},\mathbf{t}) = \mathbf{H}(\mathbf{r},\mathbf{t})\boldsymbol{\Psi}(\mathbf{r},\mathbf{t}) \qquad \mathbf{13}$$

Equation 13, The Time Dependent Schrödinger equation, where i is the unit for imaginary numbers, ħ is the Dirac´s constant in Joules and seconds (Js), Ψ(r,t) is the wave function describing a particle´s location and momentum in time, H is the Hamiltonian (energy operator) of the system.

The quantum world is often described as a strange world, where particles are capable to do things that are impossible in classical physics. The Hamilton- Jacobi

equation, which describes particle movements in classical physics, **equation 14**, is one method that can be used in attempts to bridge the quantum world to the world of classic physics, and enables the possibilities to calculate the movements of atoms not changing energy levels (Brunk and Rothlisberger 2015). The Schrödinger equation (**equation 13**) is rather similar to the Hamilton- Jacobi equation, and can be transformed into it. The transformation requires a number of steps of calculation and that the Dirac's constant approaches 0 (**ħ->0**) in **equation 15**, meaning that the estimation of the location and momentum of the quantum particle becomes more probable. However, in reality, according to the Heisenberg uncertainty principle, the constant can never really reach zero.

$$\mathbf{H} + \frac{\delta \mathbf{S}}{\delta \mathbf{t}} \sim \mathbf{i}\hbar\left(\frac{\delta}{\delta \mathbf{t}}\right)\boldsymbol{\Psi}(\mathbf{r}, \mathbf{t}) \qquad \mathbf{14}$$

$$\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{p_x}) \geq \frac{\hbar}{\mathbf{2}} \rightarrow \boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{p_x}) > \mathbf{0} \qquad \mathbf{15}$$

Equation 14,15 the Hamilton Jacob-equation similarities to the Schrödinger equation and the Heisenberg uncertainty principle, where H is the Hamiltonian of the equation, S is a set of coordinates at specific time points, t is time, x is location, p_x is momentum, σ is uncertainty

For atoms that change their energy state through excitation or molecular bonding, their movements need to be estimated by other methods like the Ehrenfest Meanfield Dynamics and Trajectory Surface Hopping methods where an extra focus is on the potential energy surface (Brunk and Rothlisberger 2015).

In the Ehrenfest Meanfield Dynamics method, atomic movements are estimated by a Newtonian equation of motion for the nucleus, derived from the Schrödinger equation by averaging the electron and nucleus potential energy surfaces (Brunk and Rothlisberger 2015). The method is considered to produce good results as long as the electron and nuclear energy potentials do not differ too rapidly for every time step, since the method is a rather rough estimation of the solutions to the Time Dependent Schrödinger Equation.

The Trajectory Surface hopping uses a metropolis algorithm (described in the section 4.4. "Monte Carlo Methods") to predict how the energy surface is shifting and relates it to possible vector trajectories (Brunk and Rothlisberger 2015).

There are several other ways of modelling the QM properties of molecules, where some other methods worth mentioning are the Hartre-Fock method and density function theory (Sholl and Steckel 2009, page 1-33). Where in Hartree-Fock one creates a matrix for the particle describing the wave function with single electron-wave functions. The density function theory attempts to estimate the electron density by solving the so called "Kohn-Sham equations" to find the wave function for single particles.

4.3 QM/MM modelling

The QM calculations are thorough, but they are costly in terms of calculation time. To speed it up, one can estimate only parts of the protein with QM methods and the rest with MD methods also called Molecular Mechanics (**MM)**. Suspected reaction sites are often modeled with a Quantum mechanics, in order to get a detailed model of the possible chemical reactions.

There are several ways to handle the contact zones between the regions simulated according to QM and those with MM, one of these is the ONIMON methodology (Svensson et al. 1996, Maseras et al. 1995). In the ONIMON methodology, described in **equation 16**, the entire system is MM modelled, the regions QM modelled are explicitly MM modelled and the energy from the QM regions are subtracted from the entire system.

$$\boldsymbol{H_{QM/MM} = H_{QM} + ((H_{MM1} + H_{MM2}) - H_{MM2})} \qquad \textbf{16}$$

Equation 16. Describing the coupling zone between QM and MM areas ($H_{QM/MM}$). H_{QM} is Hamiltonian for quantum mechanics area, H_{MM1} is Hamiltonian for molecular mechanics for the entire system, H_{MM2} is Hamiltonian for molecular mechanics in quantum mechanics areas.

The interaction between QM and MM modeled regions cannot be perfect, since in quantum mechanics one typically studies movements of electrons and nucleus, but in molecular dynamics one studies movements of entire atoms and/or molecules, so there will always be a mismatch.

The QM/MM modeling has been awarded by the Nobel Prize in physics 2013, for its possibilities for medical research. One of the reasons for the award was that the computer costs are not necessarily extremely high, while the methodology is still thorough in the sites of interest in the target molecule (Royal Swedish Academy of Sciences 2013).

4.4 Monte Carlo Methods

Monte Carlo (**MC**) methods applied for protein folding intend to predict the distribution of energy states and their associated conformation of the protein by randomly altering the conformation of the protein.

Molecular bond angles are randomly altered in an MC simulation, accordingly to what is physically possible defined in a force field parameter file (similar to the

ones used in MD methods; see section 4.1. Molecular dynamics). The protein is thereby capable to visit different possible folding states randomly.

A common MC algorithm to study conformational changes of a protein is the Metropolis Hastings algorithm. This algorithm performs a search in the space of protein conformations to determine probability of certain conformations of the protein. The algorithm visits states in a Boltzman distributed manner (**equation 17)**, and generates conformations according to a Markov chain (i.e. each conformation depends on the previous conformation only). To secure the Boltzman distribution, all physical possible protein conformations must be accessible and the transition probability, W(c→c') has to obey (**equation 18**). To fulfill the second criterion the transition probability is divided into a proposal probability and an acceptance probability (**equation** 19). A suggestion is randomly provided from a symmetrical distribution (**equation 20**) for switching the protein conformation. The suggested conformation´s energy levels are then tested toward an acceptance function. The provided suggestion will according to the acceptance function always be accepted if its potential energy is lower than the current one (**equation 21**). If the suggestion has a higher potential energy than the current conformation, the probability of accepting the suggestion depends on temperature and the energy difference (**equation** 21). The thermodynamic beta is defined in **equation 22**. If the suggestion is accepted, the protein will change conformation, otherwise it will keep its conformation until another suggestion is provided (Metropolis et al. 1953, Hastings 1970, Nilsson 2014).

$$\boldsymbol{P(c) \propto e^{-\beta E(c)}} \quad \textbf{17}$$

$$\boldsymbol{W(c \rightarrow c')P(c) = W(c' \rightarrow c)P(c')} \quad \textbf{18}$$

$$\boldsymbol{W(c \rightarrow c') = F(c \rightarrow c') * A(c \rightarrow c')} \quad \textbf{19}$$

$$\boldsymbol{F(c \rightarrow c') = F(c' \rightarrow c)} \quad \textbf{20}$$

$$\mathbf{A(c \rightarrow c') = \min(1, e^{\beta(Ec' - Ec)})} \quad \textbf{21}$$

$$\boldsymbol{\beta = \frac{1}{k_B T}} \quad \textbf{22}$$

Equation 17, 18, 19, 20, 21 and 22. Equations of relevance for the description of the Metropolis Hastings algorithm. Where c is a state in a Markov chain, W is the transition probability, P is a probability function, F is a suggestion function, A is an Acceptance probability, E is energy, T is temperature, k_B is the Boltzmann constant e.g. in Joule per Kelvin (1,38064852 J/K)

Since MC methods are based on a stochastic procedure is it most probable that two simulations will differ. In order to make conclusions regarding probable folding paths and energy minimums, a large number of simulations are therefore required. A large amount of simulations will make it possible to find a convergence toward

an expected energy distribution and its associated possible conformations, according to the law of large numbers. This is seen in **equation 23 and 24**, where the expected number is getting closer to the real number (I) when the number of samples increases (Kalos and Whitlock 2008).

$$\boldsymbol{E(X) = \left(\frac{1}{N}\right) \sum x_i} \quad \textbf{23}$$

$$\boldsymbol{I = E(X)\ if\ N \rightarrow \infty} \quad \textbf{24}$$

Equation 23 and 24. Describes the law of large numbers, E is expected number, I is the actual number in reality, N is amount of samples.

If a large number of simulations are used, the energies will converge toward a normal distribution according to the central limit theorem. The variance between sample means will gain less importance when the number of samples increases, given that there is a finite variance (which there is, since there is a finite amount of conformations) (Kalos and Whitlock 2008). MC simulations, like MD simulations, can get trapped in regions where the energy states correspond to local minimum and the acceptance function does not accept the provided suggestions for escaping. This is illustrated in **figure 7**, where a fictive protein with different energy levels, a global energy minimum and several energy minimum are illustrated, but the path toward the global minimum is not obvious.

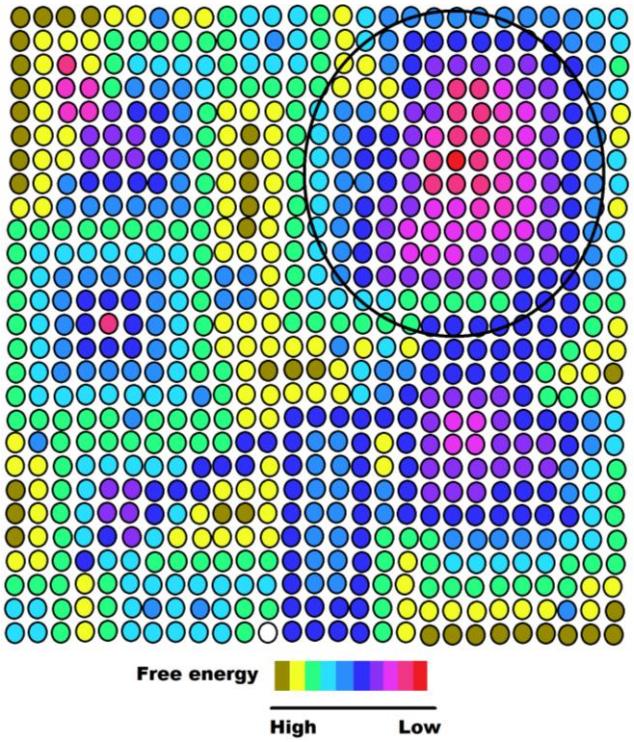

Figure 7. Example of a Markov chain with states of different energy levels. The encircled area is the location of global energy minimum.

4.5 Las Vegas methods

Las Vegas (**LV**) algorithms are methods often used to enhance MD and MC simulations by rapidly searching for plausible global energy minimums. There are two types of LV methods used for predicting protein folding, simulated annealing and Taboo search.

Simulated annealing uses the Metropolis Hasting´s MC, but the temperature in the Metropolis Hasting´s function **equation 17-22** is increased, making the acceptance function more willing to accept radical suggestions. This makes the function capable to wander to all conformations. The temperature then decreases slowly according to a cooling scheme. The function can then manage to escape from areas with states that correspond to energy minimums. To be certain that the procedure ends in the global minimum region, the annealing procedure is repeated (Krikpatrick et al. 1983, Adcock and McCammon 2006).

When the simulated annealing process reaches the freezing point, the Metropolis Hasting´s function will start only to accept suggestions that have lower energy than the current state. This type of behaviour is often rather unreliable in large systems, since they can easily get trapped in energy minimums. Therefore, the simulated annealing procedure is not suited as to point out the exact location of the global energy minimum, but rather to provide information of where to start a MC or MD simulation with a fixed temperature.

The taboo search is an algorithm that always chooses the lowest energy state of adjacent states until reaching a minimum. When reaching a minimum, the algorithm conducts a taboo move. The taboo move allows the algorithm to make long distance moves to a less favourable energy state, allowing an escape from the energy minimum. To avoid returning to the same minimum region, the algorithm often contains memory functions, where it remembers which moves that have been conducted (Fiechter 1994).

In reality, neither simulated annealing nor taboo search are used to find the exact position of the global energy minimum, more often to find a plausible global energy minimum. The only way to find the exact global energy minimum requires testing every single energy state, also called brute force. Brute force is a procedure that often is considered to take unacceptable long time.

4.6 Artificial intelligence methods

Swarming intelligence is a way of conducting multiple searches of energy states simultaneously in order to find the native state. The methods are named after social

insects; the ant colony algorithm and the bee colony algorithm, employee a number of virtual workers to find the native state. The workers independently search for the lowest minimum often using a Metropolis Hasting´s searching algorithm. After a certain amount of time the workers are allowed to "communicate" with each other. In this procedure, the workers alter their locations to the lowest energy state found. When communicating, a few workers are elected to become scouts. The scouts are sent to distant random energy states to search for an energy minimum. The scouts are then returned in order to communicate if all workers should switch location to one of the recently explored regions. The entire search cycle is then repeated. To avoid returning to previous locations, all workers share a common hive memory, which they update every time they communicate. The algorithm estimates an answer, faster than a normal MC procedure, for small proteins. Unfortunately, it can be tricky to make the algorithm accept the scout's suggestions on new location further away. Therefore, the algorithm can be a bit slow for larger proteins (Shmygelska and Hoos 2005, Zhang and Wu 2012).

4.7 Topology methods

It is possible to determine the folding of a protein by conducting an unfolding estimation of a protein´s known conformation. If one pulls the peptide chain in one direction, the intermolecular bond connections closest to the pulling terminus will break up first and the chain will unfold. This provides the pathways that are possible and probable for the protein to unfold, which in turn provides information of possible pathways of how the protein actually can fold to its native state (Mugler et al. 2014).

When unfolding the protein, it is seen as a chain of several loops, where a loop is defined as the circle formed geometrical figure between two intermolecular bonds connecting two residues. When unfolding the protein, these loops are opened one by one in a specific manner, depending on the topology nature of how residues are connected. There are three ways in which residues connect; in parallel, serial or crossing manners. Parallel connected residues always capture a loop inside their own loops. The serial connections form independent loops next to each other. A crossing connection occurs when one part of a connection is inside another connection´s loop and the other end is not (Mashagi et al. 2014). The parallel connections can only unfold in one direction, where the outer loop unfolds first. Meanwhile, the serial and crossing bonds can be unfolded in a more independent manner, and give rise to several possible unfolding alternatives (Mashagi et al. 2014).

4.8 Homology and ab initio modeling

Based on statistical alignment of the primary structure, estimations of a protein´s structure are done by searching for similarities in other proteins with Homology and ab initio modelling.

Homology modelling is based on the hypothesis that proteins with similar primary structure often have a similar spatial structure. The similarities are often explained by the slow processes of evolution, where mutated proteins do not differ too much from the wildtypes, since that can endanger the functionality of the protein and in the long run even the survival of the individual. When modelled, the primary structure of the modelled protein is aligned against the primary structures of other proteins to identify large regions of high similarity. The proteins with the highest similarities in the interesting regions are then used as templates. The conformation of the template protein is then forced upon the modeled protein. To avoid faulty modeling, the candidate templates are compared to each other in order to estimate which regions that are the most probable to be structurally different between the modeled protein and the template (Cavasotto and Phatak 2009).

Ab initio modeling is a way of estimating the secondary structures by comparing the secondary structure on several proteins with a similar primary structure. In the modeling procedure, the primary structure of the modeled protein is cut into several pieces called threading and are fed in to a neural network. The neural network compares the threads with secondary structures of identical sequences of many proteins with known structures. Finally, the most probable secondary structures are puzzled together and suggested as the final structure of the protein (Pirovano and Heringa 2009, Yang et al. 2015, Kelley et al. 2015).

5 Gluten and gliadins

Gluten consists of the storage proteins of wheat that remains as a sticky mass after washing of wheat flour in water. The storage proteins of the wheat grain compose of 70-80 different types of proteins and beside the proteins, wheat grain contain e.g. lipids, starch and fibers. The gluten is known for making bread fluffy during baking. In the kneading process of making bread, the gluten forms large networks most likely in a so called "train and loop" process where long tails of proteins attach by hydrogen bonding and Van der Waals forces, but also by formation of disulfide links that cross-link proteins together (Belton 1999, Johansson et al. 2013, Shewry et al. 2002). When a fermentation agent is added, the network functions as a membrane that contains the gases from the agent. During baking, the network solidifies due to denaturation and gets even more cross linked due to formation of more disulfide links to increased temperatures. (Johansson et al. 2013).

Producing fluffy breads is not the main purpose of gluten. Primarily produced in the seeds of wheat plants, it is mainly a storage protein for the seeds, and used as a nutrient source during germination. The amount of gluten and gluten composition depends on genetic and environmental factors (Rasheed et al. 2016). Treating plants optimally, providing them with sufficient fertilizers, water, sun and protection from diseases and pests are also important in order to gain desirable gluten content.

The protein composition of gluten consists of two major fractions; the gliadins and the glutenins. The gliadin group is in turn composed of the α/β, γ and ω gliadins, where the α-gliadins are the proteins of interest in this research project. Most of the gliadins are monomeric proteins composed of large sections of repetitive sequences. The α/β, γ gliadins contain up to eight cysteine residues, and can form up to four intra disulfide bonds. The ω gliadins on the other hand do not contain any cysteine residues. The glutenins are large (500 000 up to 10 million MW (Da)) and are cross-linked to each other through disulfide bonds and other covalent bonds (Wieser et al. 2006).

Gliadins can form tough material with properties, similar to plastics and other polymers. The material is formed when cysteine residues in gliadins are inter chain crosslinking in an oxidative environment, under heat, shear forces and additives.

It might be difficult to predict behaviors of α-gliadins since they are intrinsically disordered proteins. These proteins normally have a dynamic spatial conformation. Their potential energy surface is probably very flat, making it difficult for it to fold to its native state.

Some behaviors that are suspected to be attributed for α-gliadin are that they can form supramolecular patterns similar to hexagonal structures. This can be seen in α-gliadins mixed with other gliadins when exposed to elevated temperature pressure and addition of the chemical chaperone "glycerol" (Rasheed et al. 2015). Even if the proteins formed structures it was in form of a glassy amorphous material.

The exact structural conformation of the α-gliadins in either native or normal state is still unknown, since it is difficult extract large enough amount pure protein and it is difficult crystalize it for x-ray techniques to provide high resolution images. Unfortunately, there is only one reported instance where an α-gliadin fraction has been crystallized, and that was in a zero gravity environment onboard the previous Russian space station MIR (Aibara 1995).

6 Discussion

What kind of approach should one have when attempting to model properties of the α-gliadin protein? In **figure 8** a flow chart illustrates a suggested approach to the issue, which will be discussed in the following sections. The suggestion involves how to setup the protein system, either by electron density maps, homology modeling or Ab initio modeling. The protein system needs to be validated to be certain that it is correct. If necessary, the protein system and/or the modeling software needs to be calibrated. When the system is set, the experiments on the system can be conducted with **MD**. The results from the experiments should be validated. Finally, one can make chemical interaction simulations with **QM/MM** on interesting results.

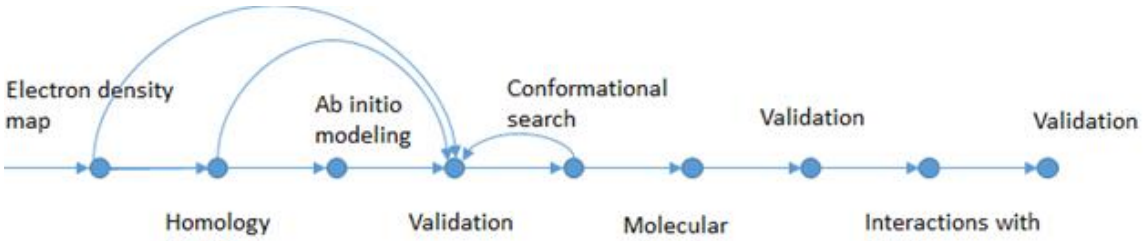

Figure 8. Proposal of flow chart for modelling polymerisation of α-gliadin.

6.1 Electron density map

The protein system of α-gliadin protein needs to be set up, with a solvent in normal room temperature conditions. Just entering the primary structure sequence will provide a long unfolded amino acid chain in a big box for all simulation software. To fold the protein from scratch, and see how it looks like in room temperature is a demanding procedure that probably would take unacceptable long time. If possible, one should try to use gathered data of the spatial structure of the protein and use it as setup data.

The best setup data are often provided from an electron density map of the protein and solvent. This map allows one to arrange the protein conformation like a 3D jigsaw puzzle. The maps are obtained from high resolution x-ray or NMR analyses. Unfortunately, there are no such analyses of α-gliadin, since it fails to crystallize, and other measurement methods have so far not provided pictures of high enough resolutions to provide an electron density map.

6.2 Homology modeling

An attempt to make a homology model of the protein is normally done when one cannot get or do not trust the setup from an electron density map or equaling data from other types of analysis experiment. When looking for templates to the α-gliadin protein in the databases, there is only distantly related proteins available with very low similarities to the primary structure. If the model protein and the template have a lower match than 75%, which is the case for the gliadin protein, it is often considered as too uncertain to use the templates. However, in pharmaceutical science one can use templates with as low similarities as 20% on the primary structure between template and model protein, but with high similarities in regions of the active sites (Young 2009). The α-gliadin protein however, share no known related active sites with any of the templates suggested, making it difficult to claim that the protein could be represented by any templates provided by the protein blast tools at http://www.ncbi.nlm.nih.gov/

.

6.3 Ab initio modelling

The last resort, if homology modelling cannot provide a good system setup, is to turn to ab initio modelling. There are several web based services that can provide

ab initio modelling for academic purposes. Even if an output is generated, most services stress that the output for α-gliadin is of low confidence.

When ab initio services attempt to model proteins larger than 100 residues like the α-gliadins, they often produce less reliable results. Proteins up to 100 residues have successfully been modelled by ab initio modelling at competitions between ab initio service developers. However, there are possibilities to roughly estimate the correctness of the suggestions from the models, even if there are no electron density maps or homologs to double check the correctness. If the provided suggestion behaves as suspected in a solvent (dissolve or stabilise) by look on RMSD values (Root Square Mean Deviation of atomic distance) in an short 2 to 5 ns MD simulation, it can be put forward for validation.

6.4 Validation

To certify that all simulated data are acceptable models of reality or close to reality, they need to be properly validated. In many simulation softwares it is possible to calculate several properties of a protein system. Comparing the model´s properties with properties measured from similar protein systems in reality enables to validate if the model is acceptable. Some of the properties that might be interesting to test at different stages of the experiment are these:

- Polymer stiffness
- Thermal conductivity
- Solubility in solvent
- X-ray diffraction patterns
- Density
- Colour
- Radius of gyration

In the case of α-gliadin, the protein material used for validation is not at the native state, meaning that we are not interested in that the ab initio service successfully models the gliadin´s native state. The protein used for validation has been processed and extracted in several rough extraction steps involving high temperatures and mechanical force. It might be tricky to get a start setup that matches real life data, but it might also not be necessary. There is room for deviation since several of the experiments will be conducted in environments with high temperatures. Environments with high temperature allow the protein to populate less likely conformations according to the Boltzmann **equation 10**.

6.5 Conformational search and calibration

If none of the provided conformations from the provided ab initio services is sufficiently valid, one or several calibrations of MD parameters and conformational searches might enhance the validity. If the system setup conformations are considered valid, one should still consider making at least a conformational search since it might increase the validity. Even if one has used an electron density map for creating a start conformation, one should make a conformational search since it might actually lead to a better fit of the map.

The conformation search is mostly a MC simulation where one attempts to find conformations closer to the native state. The MC simulation can in turn be used with a simulated annealing or a taboo search, for further increasing the probability to find conformations with less potential energy. It is possible to make the conformational search with an MD simulation together with simulated annealing as well, but it is a more time demanding procedure.

For small molecules and peptides it is possible to use a quantum mechanical approach for conformational analysis. In this way, one estimates all possible conformations and one is provided with the conformation with the lowest energy. For large molecules, like the α-gliadin protein, a MC with simulated annealing would be more advisable. Using quantum mechanical methods for large proteins are considered as impossible with current computational power.

Calibrating the simulation methods might be equally important as conformational search. One of the first choices one is confronted with when conducting MD simulations, is to choose the force field that is most suitable for describing the model protein in the solvent. First I thought that there is probably a minor difference between the force fields, until I saw a comparison of 10 different force fields on the folding of the 32-mer mouse Neh2 peptide on Youtube, by SoftSimu (2012) (Cino et al. 2012). In comparison, only three force fields found the correct peptide structure after 1000 ns, and none of the simulations resembled the other until the final ns. In order to find a suitable force field, one needs to carefully investigate benchmark analysis done by other researchers. For the gluten protein, the Amber ff99SB*-ILDN might be the most suitable force field due to benchmarking by Lindorff-Larsen et al. (2012) on protein structures (Lindorff-Larsen et al. 2010). When conducting MD and MC simulations as well, there are several other parameters, except for force fields that can be fine-tuned, leading to different results.

6.6 Molecular dynamics simulation

How the modelled protein behaves in the simulated experimental conditions can be tested by MD simulations. The MD simulation identifies the protein´s interactions and shift of properties stepwise in the experimental conditions. It is important always to make multiple simulations of an experiment. MD simulations are not necessarily reproducible and can lead to different kinds of results (Gromacs). The irreproducibility lies within rounding of calculations and producing stochastic values, which can differ from individual hardware components in computers. Therefore, one should always conduct several simulations in order to find the most probable results of an experiment.

A MC simulation might arrive at the same result as a MD simulation or even at a better result. The drawback with a MC simulation is that there are no records of the actual atomic movements of the system. The movements are important when providing information of the stability of the system at the experimental conditions. For example how much new conditions change the system, how quickly, how many changes and in what order and so on.

6.7 Chemical reactions with other proteins/molecules

In α-gliadin, it would be of great interest to predict the reactivity of the cysteine residues. Making the residues capable to enter chemical reactions with mixed QM/MM methodology, would make it possible to conduct polymerisation experiments, where the protein undergoes reactions with other proteins. The tricky part is to determine which atoms should be estimated by QM. In many instances, it requires a trial and error procedure where one tries to see how much difference one finds when shifting the locations of the limit. Since these procedures might be time consuming, one carefully chooses which systems provided from the previous MD step should undergo this simulation.

For small molecules, it is common to simulate the entire system solely on quantum mechanical level. For α-gliadin, it is impossible to use only QM approaches for estimating its interaction in an experimental system. There is no computer power to predict systems of that size. In 5 to 15 years it might be possible to apply quantum methods on larger protein systems, with the use of upcoming quantum computers (a type of computer that with today standards have extremely high capacity) that are in development. IBM has recently developed a quantum computer and continues to develop the concept (IBM 2016). If I am lucky in a few years it might be possible to make simulations on just a couple of hours on a quantum computer that now would take months on a powerful work station.

For other proteins, especially in pharmacology, one often attempts to test if ligands react to active sites or simply fit cavities in a protein structure, a procedure often referred to as docking. This is a procedure where one puts the ligand in the active site and investigates formation of intermolecular bonds that stabilize the position of the ligand. To get a good fit, it often requires extensive testing where one docks the ligand in different ways and with different conformations of the ligand and the receptor, in order to find the best fits. When the best fit is found, one needs to test the docking in a MD simulation where one puts the ligand next to the active site and where one lets it be pushed into the receptor to be able to identify if the estimated bindings are possible or not. For α-gliadin, docking procedures might be of interest to test its properties of absorbing other molecules.

7 Conclusions

The aim of this paper was to answer the following questions:

"How are functional properties of proteins related to their structure, and is it possible to determine their properties with computer based simulations?"

After writing this article I think I may dare to answer the questions with the following concluding statements.

The functional properties of a protein lie to a large extent within its structural conformation and its environment. In the same way a machine´s property lies within its design to function and in the environment it is located, where an environment with a power source enables the machine to function.

It is possible to determine different proteins´ properties through computer based simulations. The approach of working with computer based simulations will yield results that should be viewed as rationalistic science. This means one need to take into account the reliability and confidence level of input data, used parameters and variables and produced output data. However, to obtain the necessary computer power for simulate a proteins properties might in several instances be impossible today and for the upcoming ten years. For at least some of the gluten proteins subunits like the α-gliadins, it might be possible to estimate their properties through simulations, but one should not be surprised if it would require large amounts of computer power.

8 Acknowledgments

I wish to thank all my supervisors: prof. Eva Johansson, dr. Faiza Rasheed and prof. Michael Hedenqvist, for taking time to review my work and provide valuable discussion and comments necessary to make me finish this paper and start researching.

Thanks to all researchers I have discussed with in the field. Thanks to prof. Anders Irbäck for introducing me to how the Monte Carlo methods work, thanks to prof. David van der Spoel for discussing on how to use molecular dynamics in my research.

Thanks to language and speech therapist student Susanna Markgren and international languages teacher Elisabeth Sandström for commenting on my text and helping me with spelling and grammar.

9 References

Abraham, M.J., van der Spoel, D., Lindahl, E., Hess, E. and the GROMACS development team (2016). Gromacs User Manual version 2016-rcl, www.gromacs.org, http://manual.gromacs.org/documentation/2016-rc1/manual-2016-rc1.pdf time: 2016-07-14 14:56

Adcock, S.A., McCammon, J.A. (2006). Molecular Dynamics: Survey of methods for simulating the activity of proteins. *Chem Rev*, **106**, 1589-1615

Aibara, S. (1995). Crystallization of wheat γ-gliadin under a microgravity environment using space station MIR. *Journal of Crystal Growth*, **155**, 247-253

Aims, E.S. (1953). Columb´s Law and the quantitative interpretation of reaction rates. *Journal of chemical education,**30**, 351-353*

Alm, E., Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci**,* **96**, 11305-11310

Andersen, H.C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The journal of chemical physics*, **72**, 2384-2393

Anderson, D.H., Sawaya, M.R., Cascio, D., Ernst, W., Modlin, R., Krensky, A., Eisenberg, D. (2002). Granulysin Crystal Structure and a Structure-Derived Lytic Mechanism. *J.Mol.Biol*, **325**, 355-365, http://www.rcsb.org/pdb/explore/explore.do?structureId=1L9L

Arunan, E., Desiraju, G.R., Klein, R.A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D.C., Crabtree, R.H., Danneberg, J.J., Hobza, P., Kjaergaard, H.G., Legon, A.C., Mennucci, B., Nesbitt, D.J. (2011a). Defining the hydrogen bond: An account (IUPAC Technical Report). *Pure Appl. Chem*, **83**, 1619-1636

Arunan, E., Desiraju, G.R., Klein, R.A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D.C., Crabtree, R.H., Danneberg, J.J., Hobza, P., Kjaergaard, H.G., Legon, A.C., Mennucci, B., Nesbitt, D.J. (2011b). Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem, **83**, 1637-1641*

Bai, Y., Sosnick, T.R., Mayne, L., Englander, S.W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192-197

Banushkina, P.V., Krivov, S.V. (2015). High-resolution free energy landscape analysis of protein folding. *Biochemical society transactions*, **43**, 157-161

Belton, P.S. (1999). On the Elasticity of Wheat Gluten. *Journal of Cereal Science*, **29**, 103-107

Bieschke, J., Weber, P., Sarafoff, N., Beekes, M., Giese, A., Kretzschmar, H. (2004). Autocatalytic self-propagation of misfolded prion protein. *Proc Natl Acad Sci U S A*, **101**, 12207-12211

Bork, P., Koonin, E.V. (1996). Protein sequence motifs. *Current opinion in structural biology*, **6**, 366-376

Bränden, C., Tooze, J. (1999). introduction to protein structure. *Garland Pub*

Brooks, B.R., Bruccoleri, R. E., Olafson, B. D., States, D.J., Swaminathan, S., Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry*, **4**, 187-217

Brunk, E., Rothlisberger, U. (2015). Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chemical Reviews*,**115**, 6217-6263

Cavasotto, C.N., Phatak, S.S. (2009). Homology modelling in drug discovery: current trends and applications. *Drug Discovery Today*, **14**, 676-683

Chang, R. (2007). Chemistry 9th edition, *McGrawHill*, United states

Chatterjee, D. (2015). pH dependent protein stability: A quantitative approach based on Kramer`s barrier escape. *Chemical Physics Letters*, **618**, 94-98

Cheung, M.S., Garcia, A.E., Onuchic, J.N. (2002). Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *PNAS*, **99**, 685-690

Chi, E.Y., Krishnan, S., Randoolph, T.W., Carpenter, J.F. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical research*, **20**, 1325-1336

Chou, C-C., Lou, Y-C. , Tang, T.K., Chen, C. (2010). Structure and DNA binding characteristics of the tree-Cys2His2 domain of mouse testis zinc finger protein. *Proteins*, http://www.rcsb.org/pdb/explore/explore.do?structureId=2KVG

Cino, E. A., Choym, W-Y., Karttunen, M. (2012). Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput*, **8**, 2725-2740

Coluzza, I. (2015). Constrained versus unconstrained folding free-energy landscapes. *Molecular physics: An international journal at the interface between chemistry and physics*, **113**, 2905-2912

Dill, K.A., Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nature structural biology*, **4**, 10-18

Dobson, C.M.M. (2003). Protein folding and misfolding. *Nature*, **426**, 884-890

Dobson, D.M. (2001). The structural basis of protein folding and its links with human disease. *Phil. Trans. R Soc. Lond.*,**356**, 133-145

Fetrow, J.S. (1995). Omega loops: nonregular secondary structures significant in protein function and stability. *The FASEB Journal*, **9**, 708-717

Fiechter, C.N. (1994). A parallel taboo search algorithm for large traveling salesman problems. *Discrete Applied Mathematics*, **51**, 243-267

Gromacs, http://www.gromacs.org/Documentation/Terminology/Reproducibility access time 2016-08-29 16:04

Hastings, W.K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109

Hoos, H.H., Stützle, T. (2013). Evaluating Las Vegas Algorithms- Pitfalls and Remedies, *eprint arXiv:1301.7383*, 238-245

Horton, R.H., Moran, L.A., Scrimgeour, K.G., Perry, M.D., Rawn, J.D. (2006). Principles of Biochemistry, 4th ed, *Pearson Prentice hall*, United states

IBM (2016). IBM Makes Quantum Computing Available on IBM Cloud to Accelerate Innovation, http://www-03.ibm.com/press/us/en/pressrelease/49661.wss access time 2016-08-29 16:04

Israelachvili, J.N. (2011). Intermolecular and Surface Forces, *Elsebier Inc*, **Third edition**, 107-130

Johansson, E. (2013). Modeling plant protein structures for optimal performance in various applications. *Vetenskapsrådet*, diarienr 2013-5991. http://vrproj.vr.se/detail.asp?arendeid=102400

Johansson, E., Malik, A.H., Hussain, A., Rasheed, F., Newson, W.R., Plivelic, T.S., 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**, 367-376

Jones, J.E. (1924). On the determination of molecular fields. ||. From the equation of state of a gas. *Proceedings of the Royal Society of London A: Mathematica, Physical and Engineering Sciences*, **106**, 463-477

Kalos, M.H., Whitlock, P.A. (2008). Monte Carlo Methods. *John Wiley & Sons*, 1-199

Kamtekar, S., Hecht, M.H. (1995). The four-helix bundle: what determines a fold?. *The FASEB Journal*, **9**, 1013-1022

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nature Protocols*, **10**, 845-858

Kenakin, T. (2001). Inverse, protean and ligand-selective agonism: matters of receptor conformation. *FASEB*, **15**, 598-611

Klug, A., Schwabe, J.W.R. (1995). Zinc fingers. *The FASEB Journal*, **9**, 597-604

Kobilka, B.K. (2007). G protein coupled receptor structure and activation. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, **1768**, 794-807

Krikpatrick, S., Gelatt, C.D., Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, **220**, 671-680

Krishnamani, V., Hegde, B.G., Langen, R., Lanyi, J.K. (2012). Secondary and Tertiary structure of Bacteriorhodopsin in the SDS Denatured State. *Biochemistry*, **51**, 1051-1060

Lambright, D.G., Noel, J.P., Hamm, H.E., Sigler, P.B. (1994). Structural determinants for activation of the α-subunit of a heterotrimeric G protein. *Nature*, **369**, 621-628

Levinthal, C. (1968). "Are there pathways for protein folding.". *J.Chim. phys*, **65**, 44-45

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., Shaw, D. E., (2012). Systematic Validation of Protein Force Fields against Experimental Data. *PLos one*, **7**, e32131

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, **78**, 1950-1958

Lins, L., Brasseur, R. (1995). The hydrophobic effect in protein folding. *The FASEB Journal*, **9**, 535-540

Maseras, F., Morokuma, K. (1995). IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *Journal of Computational Chemistry*, **16**, 1170-1179

Mashagi, A., Van Wijk, R.J., Tans, S.J. (2014). Circuit Topology of proteins and nucleic acids. *Structure*, **22**, 1227-1237

Mathematicalmonk, (2011). (ML 15.2) Newton's method (for optimization) in multiple dimensions, https://youtu.be/42zJ5xrdOqo,

Mayer, M.P., Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell Mol Life Sci.*,**62**, 670-684

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953). Equation of state calculations by fast computing. *The Journal of Chemical Physics*, **21**, 1087-1092

Monajjemi, M. (2015). Cell membrane causes the lipid bilayers to behave as variable capacitors: A resonance with self-induction of helical proteins. *Biophysical Chemistry*, **207**, 114-127

Mugler, A., Tans, S.J., Mashaghi, A. (2014). Circuit topology of self-interacting chains: implications for folding and unfolding dynamics. *Physical Chemistry Chemical Physics*, **16**, 22537-22544

Nagy-Smith, K., Moore, E., Schneider, J., Tycko, R. (2015). Molecular structure of monomorphic peptide fibrils within a kinetically trapped hydrogel network. *Proc. Natl.Acad.Sci.USA*, **112**, 9816-9821, http://www.rcsb.org/pdb/explore/explore.do?structureId=2N1E

Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundström, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M., Kay, L.E. (2012). Structure of an intermediate state in protein folding and aggregation. *Science*, **336**, 362-366

Ng, A. (2013). 6. Gradient Descent Intuition, https://youtu.be/kWq2k1gPyBs, access time 2016-01-06, 13:19

Nilsson, D. (2014). A Multisequence Monte Carlo Method in a simple Protein Model. *FYTK01*

Pace, N.C., Trviño, S., Prabhakaran, E., Scholtz, J.M. (2004). Protein structure, stability and solubility in water and other solvents. *Phil. Trans. R. Soc. Lond.*, **359**, 1225-1235

Paci, E., Vendruscolo, M., Karplus, M. (2002). Validity of Gō Models: Comparison with a Solvent-Shielded Empirical Energy Decomposition. *Biophysical Journal*, **83**, 3032-3038

Pauling, L., Corey, R.B., Branson, H.R. (1951). The structure of proteins: two hydrogen-bonded helical conformations of the polypeptide chain. *Proc Natl Acad Sci U S A*, **37**,205-211

Pichler, A., Bartalucci,S., Bertolucci, S., Berucci, C., Bragadireanu, M., Cargnelli, M., Clozza, A., Curceanu, C., Paolis, L. De., Matteo, S. Di., DÚffizi, A., Egger, J.P., Guaraldo, C., Iliescu, M., Ishiwatari, T., Laubenstein, M., Marton, J., Milotti, E., Pietreanu, D., Piscicchia, K., Ponta, T., Sbardella, E., Scordo, A., Shi, H., Sirghi, D., Sirghi, F., Sperandio, L., Vazquez-Doce, O., Widmann, E., Zmeskal, J. (2015). VIP 2: experimental tests of the pauli exclusion principle for electrons. *Hyperfine Interact*, **233**, 121-126

Pirovano, W., Heringa, J. (2009). Protein Secondary Structure Prediction, in Carugo, O, Eisenhaber, F (2010) Data Mining Techniques for the Life Sciences. *Methods in Molecular Biology*, 327-348

Pletnev, S., Subach, F.V., Dauter, Z., Wlodawer, A., Verkhusha, V.V. (2012). A Structural Basis for Reversible Photoswitching of Absorbance Spectra in Red Fluorescent Protein rs TagRFP. *Journal of Molecular Biology*,**417**, 144-151

Press, W.H., Teukolsky, S.A., vetterling, W.T., Flannery, B.P., (1997). Root finding and nonlinear sets of equations. in *Numerical recipes in C: The Art of Scientific Computing*, 2 nd edition (Cambridge: Cambridge University Press), 347-393

Ramachandran, G.N., Ramakrishna, C., Sasisekharan, V. (1963). Stereochemistry of polypeptide chain conformations. *Journal of molecular Biology*, **7**, 95-99

Rasheed, F., Kuktaite, R., Hedenqvist, M.S., Gällstedt, M., Plivelic, T.S., Johansson, E. (2016).The use of plants as a "green factory" to produce high strength gluten-based materials. *Green Chemistry*, **18**, 2782-2792

Rasheed, F., Newson, W.R., Plivelic, T.S., Kuktaite, R., Hedenqvist, M.S., Gällstedt, M. and 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.

Richardson, J.S. (1994). Introduction: Protein Motifs. *The FASEB Journal*, **8**, 1237-1239

Riek, R.P., Graham, R.M. (2011). The elusive π-helix. *Journal of structural biology*, **173**, 153-160

Royal Swedish Academy of Sciences (2013). The Nobel Prize in Chemistry 2013. press release 2013-09-09

Russo, A.A., Jeffrey, P.D., Pavletich, N.P. (1996). Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nature Structural Biology*, **3**, 696-700

Sali, A., Shakhnovich, E., Karplus, M. (1994). A lattice model study of the requirements for folding to the native state. *J. Mol. Biol*, **235**, 1614-1636

Shewry, P.R., Halford, N.G., Belton, P.S., Tatham, A.S. (2002). The structure and properties of gluten: an elastic protein from wheat grain. *Phil. Trans. R Soc. Lond.*, **357**, 133-142

Shmygelska, A., Hoos, H.H. (2005). An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem. *BMC Bionformatics*, **6**, 30-52

Sholl, D.S., Steckel, J.A. (2009). Density Functional Theory: A Practical Introduction. *John Wiley & Sons , Inc*, New Jersey Unitead States of America

SoftSimu (2012). Biological and Condensed Matter Group at University of Waterloo, Canada, https://youtu.be/AtDOJnVNC18 access time: 2016-05-05, 17:06

Stirnemann, G., Kang, S-G., Zhou, R., Berne, B.J. (2014). How force unfolding differs from chemical denaturation. *PNAS*, **111**, 3413-3418

Svensson, M., Humbel, S., Froese, R.D.J., Matsubara, T., Sieber, S., Morokuma, K. (1996). A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem*, **100**, 19357-19363

Tokuriki, N., Tawfik, D.S. (2009). Protein Dynamism and evolvability. *Science*, **324**, 203-207

Toniolo, C., Benedetti, E. (1991). The polypeptide 3_{10}-helix. *Trends in biochemical sciences*, **16**, 350-353

Vert, M., Doi, Y., Hellwich, K-H., Hess, M., Hodge, P., Kubisa, P., Rinaudo, M., Schué, F. (2012). Terminology for biorelated polymers and applications (IUPAC Recommendations 2012)*. *Pure Appl. Chem.*, **84**, 377-410

Wang, Y., Sarkar, M., Smith, A.E., Krois, A.S., Pielak, G.J. (2012). Macromolecular crowding and protein stability. *Journal of the American chemical society*, **134**, 16614-16618

Wieser, H., Bushuk, W., MacRitchie, F. (2006). The Polymeric glutenins. In: Wrigley, C, Bekes, F, Bushuk, W. (Eds), Gliadin and Glutenin: the Unique Balance of Wheat Quality. St. Paul American Association of Cereal Chemistry, 213-240

Yang, J., Yan. R., Roy, A., Xu, D., Poisson, J., Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, **12**, 7-8

Young, D.C. (2009). *Computational drug design: a guide for computational and medicinal chemists*. John Wiley & Sons.

Yuan, T.Z., Ormonde, C.F.G., Kudlacek, S.T., Kunche, S., Smith, J.N., Brown, W.A., Pugliese, K.M., Olsen, T.J., Iftikhar, M., Raston, C.L., Weiss, G.A. (2015). Shear-stress-mediated refolding of proteins from aggregates and inclusion bodies. *ChemBioChem*, **16**, 393-396

Zhang, X., Kelly, W. (2014). Chaperonins resculpt folding free energy landscapes to avoid kinetic traps and accelerate protein folding. *Journal of molecular biology*, **426**, 2736-2738

Zhang, Y., Wu, L. (2012). Artificial bee colony for two dimensional protein folding. *Advances in electrical engineering systems*, **1**, 19-23

Zwanzig, R., Szabo, A., Bagchi, B. (1992). Levinthal´s paradox. *Proc. Natl. Acad. Sci. USA*, **89**, 20-22