Emulsification and Emulsion Stabilising Properties of Fragment Plant Proteins

Cuizhen Chen

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Food Science and Nutrition

January 2023

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

Firstly, I would like to express my deepest appreciation to my primary supervisor Dr Rammile Ettlaie. He was always there to help and encourage me when I was having difficulties in my PhD studies or in life. Without his patient guidance, I would not have been able to complete my studies successfully. Secondly, I would like to extend my sincere thanks to my second supervisor Prof. Brent Murray for his practical suggestions in my experimental work and thesis. In addition, I gratefully acknowledge the assistance of my PhD colleagues from the University of Leeds, especially Dr Yue Ding and Dr Ruixian Han, for their guidance in laboratory techniques.

I would like to thank my parents for providing me with encouragement and support in my life and sponsoring me to study abroad. I would like to thank my girlfriend, Shiyu Tang, for being there for me during the most stressful times. Finally, I would like to thank all my friends for their company during my entire PhD journey.

Publications from thesis

Manuscript in progress:

The relevant paper based on chapters 5 and 6 will be sent to Journal of Colloid and Interface Science.

The relevant paper based on chapters 2, 3 and 4 will be sent to Food Hydrocolloids.

Attended conferences and accepted abstracts

Chen, C., & Ettelaie, R., & Murray, B. A standard method of selecting protein fragments with potential to be suitable emulsion stabilisers for further SCF calculation. *The 4th UK Hydrocolloids Symposium: From Food to Bioprocessing* – University of Leeds, Leeds, West Yorkshire, UK, 12th September 2019.

Chen, C., & Ettelaie, R., & Murray, B. A theoretical technique for screening protein fragments with potential as suitable emulsifiers and stabilisers. *Physics in Food Manufacturing 2020 Conference* – University of Leeds, Leeds, West Yorkshire, UK, 15-17 January 2020.

Chen, C., & Ettelaie, R., & Murray, B. A theoretical technique for screening protein fragments with potential as suitable emulsifiers and stabilisers. *Food Colloids 2022* – Online, 11-13 April 2022.

Abstract

This PhD project investigates the possibility of fragmented plant proteins as suitable and novel food-grade emulsifiers using a combination of theoretical and experimental approaches.

Two innovative fast screening methods (coarse-graining and moving average methods) have been developed for selecting suitable fragments from intact proteins. A di-block-like fragment was found according to the primary structure of β-conglycinin. Then self-consistent field calculations (SCFC) were carried out to verify further the emulsification capacity of the fragment selected by fast screening methods.

In the experiment, soy protein isolate (SPI) was hydrolysed by pepsin under optimal (pH 2.1) and non-optimal (pH 4.7) conditions and the surface activity and emulsion stabilising capacity of the resultant peptides were measured and modelled via SCFC. Hydrolysis at pH 2.1 and 4.7 resulted in a considerable increase in measured surface activity compared to the native (non-hydrolysed) SPI, but the hydrolysate from pH 2.1 was not a good emulsion stabiliser compared to the hydrolysate (particularly a fraction Mw > 10 kDa) at pH 4.7. Furthermore, peptide analysis of the latter (pH 4.7) suggested it was dominated by a fragment of one of the major soy proteins, β-conglycinin of Mw $\approx$ 25 kDa. SCFC calculations confirmed that this peptide should be an excellent stabiliser.

Table of Contents

List of Tables

List of Figures

Chapter 1
General Introduction

1.1 Colloidal state and emulsions

In everyday life, colloidal systems are frequently encountered in a wide range of products, for example, foods, pharmaceuticals and cosmetics. The colloid is a mixture of two or more immiscible substances with at least one relevant linear dimension ranging from 1 nm to 1 μm (Dickinson, 1992). Depending on the nature of the dispersed phase and the dispersion medium, colloids are classified into different categories according to Table 1.1 (Dickinson, 1992).

Table 1.1 Classification of different colloids

		Dispersed phase		
		Solid	Liquid	Gas
Dispersed medium	Gas	Solid dispersion	Solid emulsion	Solid foam
	Solid	Sol or Colloidal dispersion	Emulsion	Foam
	Liquid	Aerosol	Aerosol	-

In its most simple form, a colloid comprises just one dispersed phase and one dispersion medium. However, colloids can often contain more than just one dispersed phase in many practical applications involving foods. For example, cake butter contains oil droplets, fat crystals, starch granules and protein particles (Shepherd and Yoell, 1976). This thesis focuses on the study of emulsions. In its most usual form in foods, an emulsion of oil-in-water (O/W) has a very large number of oil droplets dispersed into a complex aqueous environment that very likely contains macromolecules,

salts and sugars. When the oil phase happens to be the continuous phase, containing a dispersion of many water droplets, the emulsion is referred to as a water-in-oil emulsion (W/O) (Dickinson and Stainsby, 1982). With the further development of colloidal science, more types of emulsions are also being explored, such as multiple emulsions like water-in-oil-in-water W/O/W emulsion (Okochi and Nakano, 2000) and oil-in-water-in-oil O/W/O emulsion (Hwang et al., 2005). These multiple emulsions are generally much harder to produce and keep stable, but when formed, they can offer distinct advantages in the formulation of products (e.g. barrier against oxidation, reducing the amount of oil and fat while keeping the same sensation as O/W emulsions, etc.).

1.2 Stabilisation and Destabilisation of Colloidal Systems

1.2.1 Colloidal instability

Colloidal systems are generally not in the lowest free energy state that a system can adopt, unlike, for example, a well-dissolved solution which has excellent thermodynamical stability and uniformity (Dickinson and Stainsby, 1982). Therefore, colloidal systems are thermodynamically unstable and will eventually separate into individual components after a sufficient period of storage. However, the speed and kinetics of the phase separation can be sufficiently slowed down using a number of methods, including the addition of extra emulsifiers. This then allows for colloids to be stable over the times intended for their use, with other factors such as chemical or biological stability often limiting the shelf life of the product rather than colloidal instability. In such cases, we say that the colloid is kinetically stable. According to Figure 1.1 below, colloids gradually transition from an initially homogeneous state to several unstable states before forming the final phase-separated system.

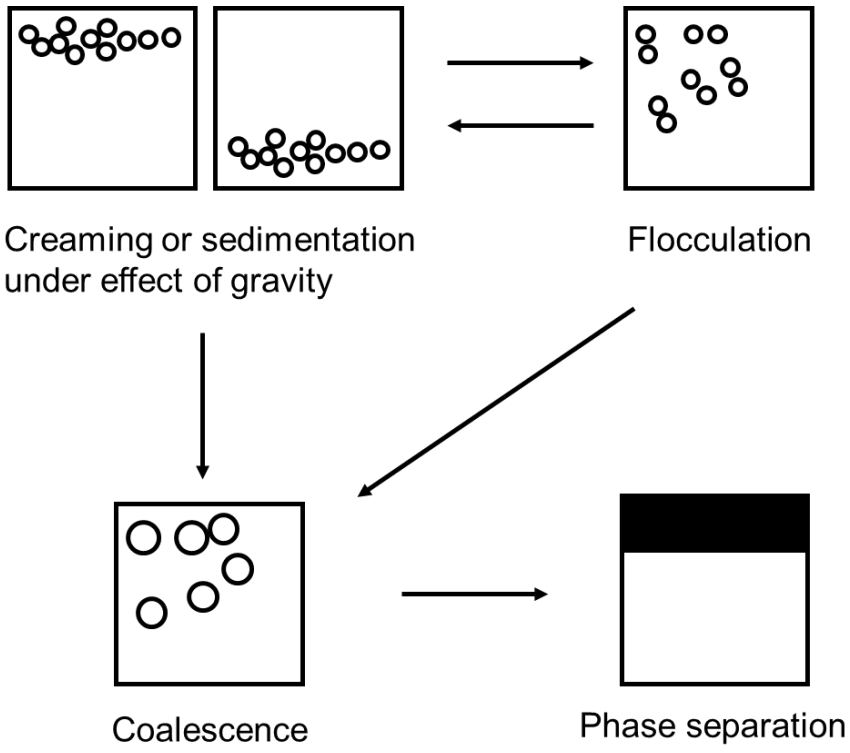


Figure 1.1 Different types of unstable states of the colloidal system

Next, the destabilisation process of a colloid is described briefly below. Firstly, under the influence of gravity, the dispersed phase particle will rise to the top if of lower density or sediment to the bottom if of higher density than the continuous phase. This forms a cream layer at the top (or sediment at the bottom). The creaming rate can be calculated using Stokes' law (Equation 1.1) (Pathak, 2017).

$$v = \frac{2gR^2\left(\rho_c - \rho_0\right)}{9\eta_c} \tag{1.1}$$

where R is the radius of the particle, g is the gravitational acceleration (9.81 m/s^2), ρ_c is the density of the dispersed medium, whereas ρ_0 and η_c are the density and viscosity of the continuous medium, respectively.

The above equation shows that the creaming rate of a colloid is greatly influenced by the size of the dispersed particles (as is the square of R). In order to slow down the creaming rate and thus improve the stability of the colloid, we can treat the emulsion in three ways: by reducing the size of the dispersed particles (e.g. using suitable emulsifiers, as well as deploying

more powerful homogenisers), by reducing the density difference between the dispersed medium and dispersed phase (which in most cases is not practical) and by increasing the viscosity of the dispersed medium (e.g. adding gums or long chain carbohydrates) (Krog, 2002).

The processes of creaming and sedimentation can also be understood as a form of migration of dispersed particles. After this, the dispersed particles may come closer to each other and can form aggregates (aggregation can also happen directly). As defined by La Mer (La Mer, 1966), an aggregate is a collection of two or more solid particles or liquid droplets that are held together by forces of specific strength and origin. Weak aggregation is called flocculation, where the interaction between particles is not strong enough and can easily be broken by a modest external force (e.g., shaking or stirring of the sample). Hence, flocculation is also a synonym for reversible aggregation (Dickinson and Stainsby, 1982). Bridging flocculation and depletion flocculation discussed later, are two common forms of flocculation in food colloids. Bridging flocculation often arises from the emulsifier added to the system. Suppose the added (polymer-like) emulsifier has more than one region that can be adsorbed onto the surface of dispersed droplets. In that case, a bridge may be formed between the two adjacent particles, thus causing the two particles to attract each other. Therefore, for a suitable emulsifier that could provides sufficient repulsive force, it is preferable to have a relatively simple di-block type structure to avoid the formation of bridging flocculation (more discussion of this can be found in Chapters 2 and 4).

Depletion flocculation can occur when too many polymers with weak or no adsorption capacity are added to the dispersion medium (Jones et al., 1989). As the distance between the two particles is reduced, the spatial conformation of the polymers in the gap between the particles decreases. Therefore, in order to reduce the free energy (i.e., increase their conformational entropy), the polymers tend to escape into the dispersed medium outside the particles. This occurs at a distance where the separation between the particles is smaller than the diameter of the polymers, making it difficult for the polymers outside the particles to access the gap. This results in a higher polymer concentration in the dispersion medium relative to the gap between the particles. Due to the osmotic pressure difference, the solvent flows from the low-concentration region to the high concentration, i.e.,

the solvent flows out of the gaps between the particles, causing the particles to come closer to each other and thus form flocs.

In contrast to flocculation, strong and irreversible aggregation is known as coagulation (Dickinson and Stainsby, 1982). The colloidal system is further destabilised by continuous coagulation, and in the case of liquid droplets or foams, this can lead to coalescence. In the process of irreversible coalescence, the small dispersed particles merge to form larger particles until they eventually lead to a phase-separated system (complete destabilisation). Another destabilisation process that can lead to the formation of coarser droplets is Ostwald ripening. Unlike coalescence, Ostwald ripening is a process of dissolution and re-deposition, where small particles/droplets/bubbles constantly lose material to the dispersion medium, and the material diffuses into larger particles/droplets/bubbles. This is mainly because the system tends to form particles that are larger than the original o ones in terms of free energy (Ratke and Voorhees, 2002). This can be understood in terms of the surface-to-volume ratio being less for larger droplets, involving a disproportionally less surface energy per molecule for large as compared to smaller particles.

We will now discuss another set of phenomena of prime importance to the behaviour of colloidal systems, namely Brownian motion and van der Waals forces. These two crucial factors directly impact the stability characteristics of colloids. Brownian motion is the random thermal motion of molecules or particles. The displacement range (*root-mean-square distance*) of the Brownian motion of a particle is mainly influenced by time *t* and its diffusion coefficient *D*, as given by Equation 1.2. The diffusion coefficient can be calculated from Equation 1.3 (Walstra, 2002).

$$\text{For 2D: } \Delta x_1 = \sqrt{2Dt}$$
$$\text{For 3D: } \Delta x_1 = \sqrt{6Dt} \qquad (1.2)$$

$$D = \frac{k_B T}{6\pi\eta_s R} \qquad (1.3)$$

where η_s is the viscosity of the solvent and *R* is the radius of the particle. The van der Waals force is an attractive force between two particles. It is a short-range force, i.e., when two particles are far apart, the van der Waals force is almost non-existent, whereas when two particles are close together, the strength of the van der Waals force increases rapidly (see Figure 1.2).

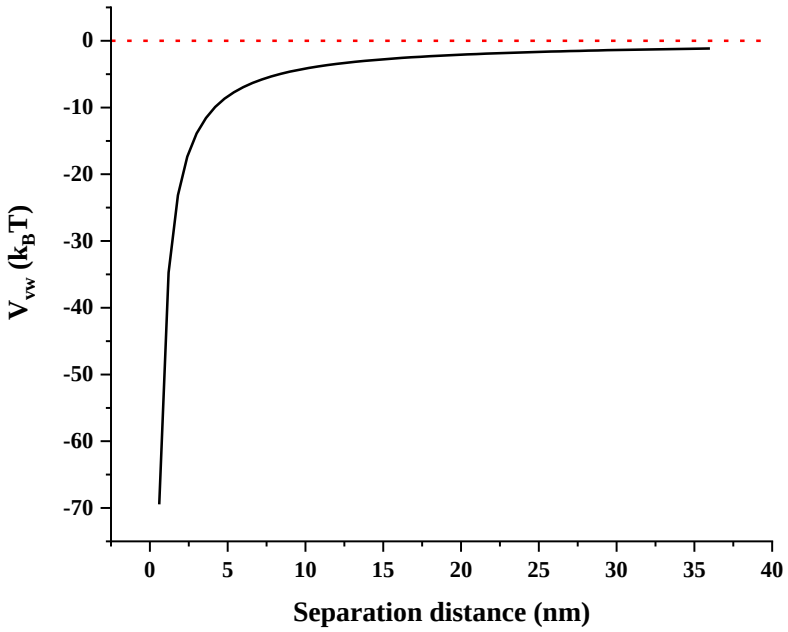


Figure 1.2 Changes in van der Waals interaction energy with distance (spherical particles with a size of 1 μm)

The van der Waals forces can be calculated using Equations 1.4 below (Dickinson and Stainsby, 1982).

$$V_{vw} = -\frac{A_H R}{12r} \tag{1.4}$$

where A_H is the composite Hamaker constant for the dispersed phase in the continuous medium with a typical value of 1 k_BT for edible oils dispersed in water, *R* is the radius of the two spherical droplets, and *r* is the distance between the droplets.

Under Brownian motion, particles collide randomly with each other, and if subject to large van der Waals forces at small separation distances, well

irreversible aggregate. However, the effect of Brownian motion on the stability of colloids is not always unwanted. For example, when two particles are far apart, Brownian motion can easily overcome the effect of the relatively weak van der Waals forces in this situation. Therefore, to maintain a stable colloidal state, we need to create sufficient repulsion forces between the two particles to overcome the attraction due to van der Waals forces as two particles begin to approach each other too closely.

1.2.2 Mechanism of colloid stability

The repulsive forces between particles that can stabilise a colloid are usually steric or electrostatic type repulsions. Steric repulsion is commonly seen in the adsorption of long-chain amphiphilic polymers, such as proteins and modified polysaccharides, on the surface of the particles. This forms a so-called hairy interfacial layer around the particle. In the case of protein, for example, hydrophobic blocks adsorb to the particle surface, while hydrophilic blocks extend into the solvent to form the adsorbed layer with a certain thickness. When the two adsorbed layers on a pair of closely spaced particles overlap with each other, a repulsive force is generated, forcing the two particles apart. There are two phenomena contributing to this steric repulsion. The first and the more dominant one is that when the two adsorbed layers overlap, the concentration of solvent molecules at the gap is lower than the concentration in the surrounding solvent, causing an osmotic pressure difference that pushes the particles apart. The second reason is that when the two particles are relatively close together, the spatial conformation of the parts (loops and tails) of the polymer chains that extends into the solvent is reduced. That constitutes an unfavourable reduction in the entropy of the chains, which tends to resist the decrease in the separation distance between the particles (Charles, 1992). In order to obtain a sufficient amount of steric repulsion, three factors should be taken into account when adding polymeric emulsifiers to colloidal systems: (1) The polymer used as an emulsifier should have relatively high adsorption energy to ensure that the polymer can strongly be anchored to the surface. The polymer should also be sufficiently hydrophilic to ensure that it can dissolve in the solvent while remaining as hydrophobic as possible (Ettelaie et al., 2017; Murray et al., 2021). Therefore the dispersed medium that makes up the colloid should be a good solvent so that the polymer can be well dispersed into the system (Obey and Griffiths, 1999) (2) As many polymers (such as proteins) do not have a simple di-block structure, the polymer should adequately cover the

surface of the particles, thus avoiding bridging effects (see the discussion of bridging in section 1.2.1) (3) The polymer concentration should be controlled to a reasonable value. As mentioned above, adding too much polymer may result in depletion flocculation (see the discussion of depletion in section 1.2.1).

Another type of repulsion that can stabilise a colloid is electrostatic repulsion. Electrostatic repulsions arise from the overlap of the electronic double layers of two charged particles. A surface is usually charged, so the solution's opposite charges (counter-ions) tend to adsorb to the surface, forming an inner immobile region (the Inner Stern Layer). In the vicinity of this region, the same charges (co-ions) as the surface are attracted to form an outer mobile region (the Diffuse Layer). The electrical double-layer model was first proposed by Gouy and Chapman (Gouy, 1910; Chapman, 1913). This model ignores the immobile inner charges and assumes that the ions in the double layers are uniformly distributed. In reality, the distribution of ions is not uniform as the volume of the surface is finite, and only limited counter-ions can adsorb the surface. Therefore, the Gouy-Chapman model is unsuitable for applying to places close to the charged surface (Dickinson, 1992b). Subsequently, Stern proposed a new and improved model of the electron double layer (Stern, 1924) (see Figure 1.3).

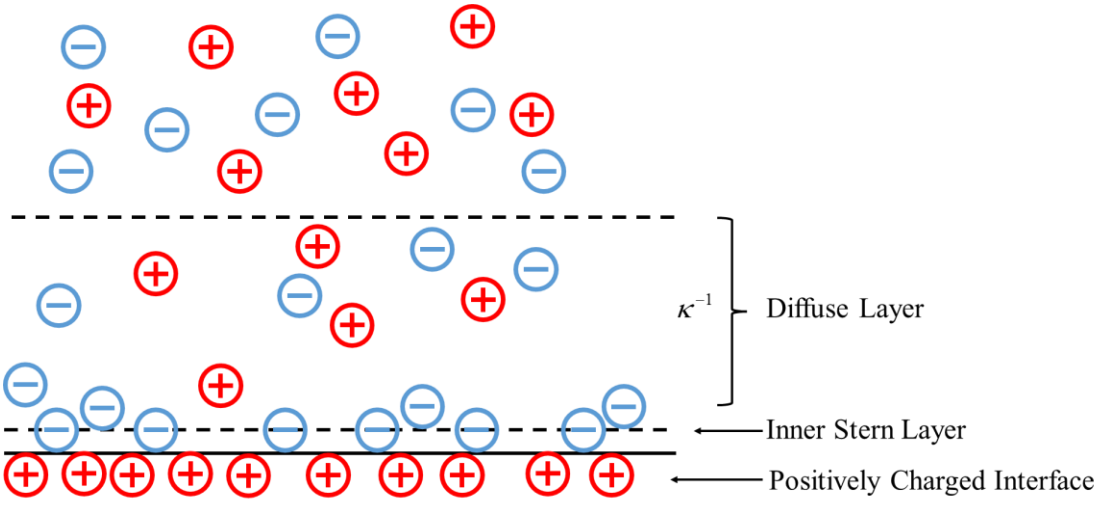


Figure 1.3 The illustration of the Gouy-Chapman-Stern electrical double-layer theory

In this refined model, the distribution of charges is not uniform; in the Stern layer, the potential decreases linearly with distance, while in the diffuse layer, the potential falls exponentially to zero. The thickness of the double diffuse

layer is commonly known as the Debye length κ^{-1}, which is the most crucial component affecting the electrostatic repulsion's potential energy. The Debye length is given by (Miyazawa et al., 2021):

$$\kappa = \sqrt{\frac{2z^2e^2n_0}{\varepsilon_r\varepsilon_0 kT}} \tag{1.5}$$

where n_0 is the concentration of salt in the solution, T is the Kelvins temperature, z is the valency of the ions, ε_r is the relative permittivity of the solution (~ 79 for water), ε_0 is the permittivity of free space (8.85× 10−12 F/m), and k is the Boltzmann constant.

When the electrical double-layer overlap, the charge (ion) concentration in the overlap is significantly higher than the solvent, so the electrostatic repulsion also arises from the difference in osmotic pressure. Based on the above theory of electrified interfaces, four scientists proposed the DLVO theory (Verwey and Overbeek, 1955; Derjaguin, 1993; Derjaguin and Landau, 1993) by further considering the effect of van der Waals forces on the stability of the colloid. DLVO theory considers the two independent van der Waals potential energy E_{vw} and an electrostatic repulsion potential energy E_A between the particles. The value of E_{vw} for spherical particles can be calculated using Equation 1.4, as previously shown (Miyazawa et al., 2021). Since the van der Waals force similarly diminishes with distance, it is evident that the strength E_{vw} decreases as the distance increases.)

Then, we can calculate the potential energy of the electrostatic repulsion E_A according to the following equation (Miyazawa et al., 2021):

$$E_A = \frac{64\pi r n_0 kT\gamma^2}{\kappa^2} e^{-\kappa R} \tag{1.6}$$

where $\gamma = \tan R\left(\frac{ze\psi_0}{4kT}\right)$, z is the bulk concentration and ψ_0 is the surface potential on the particles.

The potential energy of the total interaction between the two charged particles, *E*, is:

$$E = E_{vw} + E_A \tag{1.7}$$

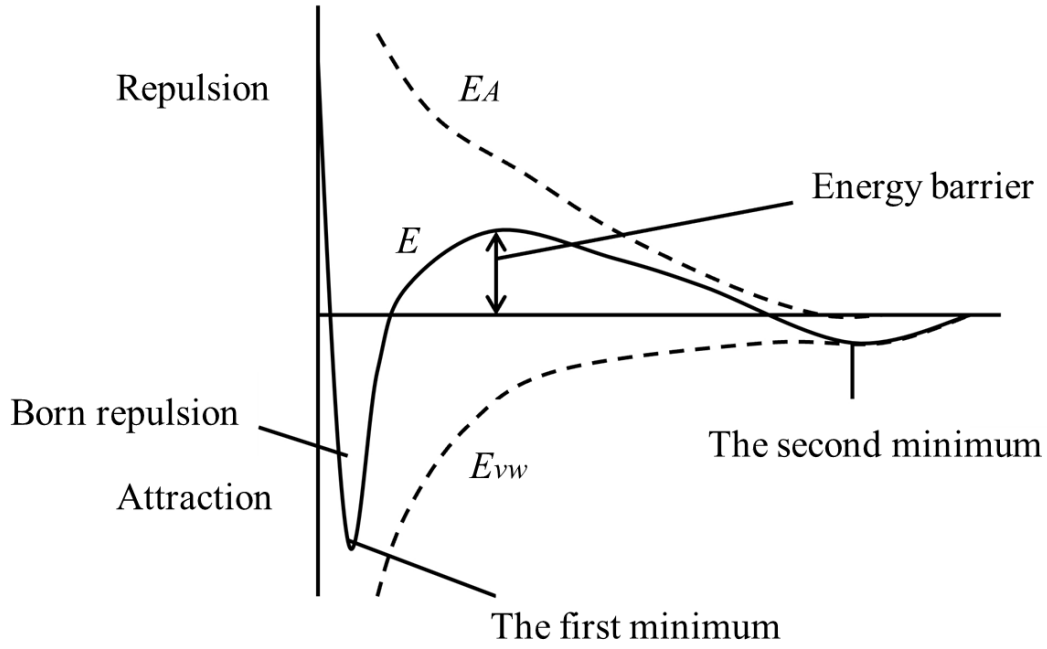


Figure 1.4 Potential energy between a pair of charged spherical particles, plotted against their separation distance, as presented in the DLVO theory

Figure 1.4 illustrates the relationship between E_{vw}, E_A and *E*. When two particles come close together, the first effect felt by the particles is the van der Waals interaction. At this stage, the contribution of the repulsion to the total interaction energy is small, which makes *E* negative. As the distance decreases further, the influence of the E_A gradually becomes more significant, becoming more dominant than the E_{vw} (sometimes leading to a minimal value, known as the secondary minimum). Then *E* gradually increases to a positive value at even closer separations. However, after a certain distance, the influence of E_{vw} again exceeds that of E_A, and *E* becomes negative once again. This produces a peak in the interaction potential, called the energy barrier. The energy barrier is critical for the stability of the colloidal particles. Two approaching particles must cross this energy barrier to get sufficiently close to form an aggregate. When the particles are very close together, *E* drops rapidly to form a deep minimum

(known as the primary minimum). The magnitude of the primary minimum is much larger than the depth of the secondary minimum. When the particles fall into the primary minimum, the formed clusters are very strong, and the aggregation is irreversible (this is referred to as coagulation). With particles trapped in the secondary minimum, the aggregation is much weaker, and clusters formed in the system can often reversibly be broken down (this form of aggregation is known as flocculation).

The influence of electrostatic and steric repulsion on colloidal stability when large-sized fragments of proteins are utilised as emulsifiers will be investigated in this thesis. In particular, near the isoelectric point of the protein (or the polypeptide fragment), steric repulsion plays a much more significant role in stabilising the colloid, whereas electrostatic repulsion is the more dominant contributor at pH values away from pI. In summary, then, the magnitude of the steric repulsion depends mainly on the structure of the emulsifier itself, and it is less strongly affected by external factors. In contrast, the electrostatic repulsion is generally more influenced by environmental attributes, such as temperature, pH and salt concentration (i.e., changes in Debye screening length).

1.3 Fragmented protein as a novel food-grade emulsifier

1.3.1 General introduction of food-grade emulsifiers

Emulsifiers are surface-active substances that prevent the destabilisation of colloids. Emulsifiers are utilised not only in the food industry but also in cosmetics (Calvo et al., 2020), agriculture (Feng et al., 2016), construction (Feng et al., 2016) and many other industries. Due to the requirement that food emulsifiers must be safe for human consumption, the variety of food-grade emulsifiers is significantly more restricted in foods than that of other emulsifiers for wider use. Food emulsifiers can be broadly divided into low-molecular-weight and macromolecular-type emulsifiers (Dickinson, 1992b). Low-molecular-weight emulsifiers are a class of molecules containing amphiphilic groups, such as monoglycerides residues, diglycerides residues, sodium lactate, lecithin, etc. For example, monoglycerides (E471) have a hydrophilic glycerol residue and a hydrophobic fatty acid residue (two hydrophobic groups for diglycerides and three for triglycerides, see Figure 1.5).

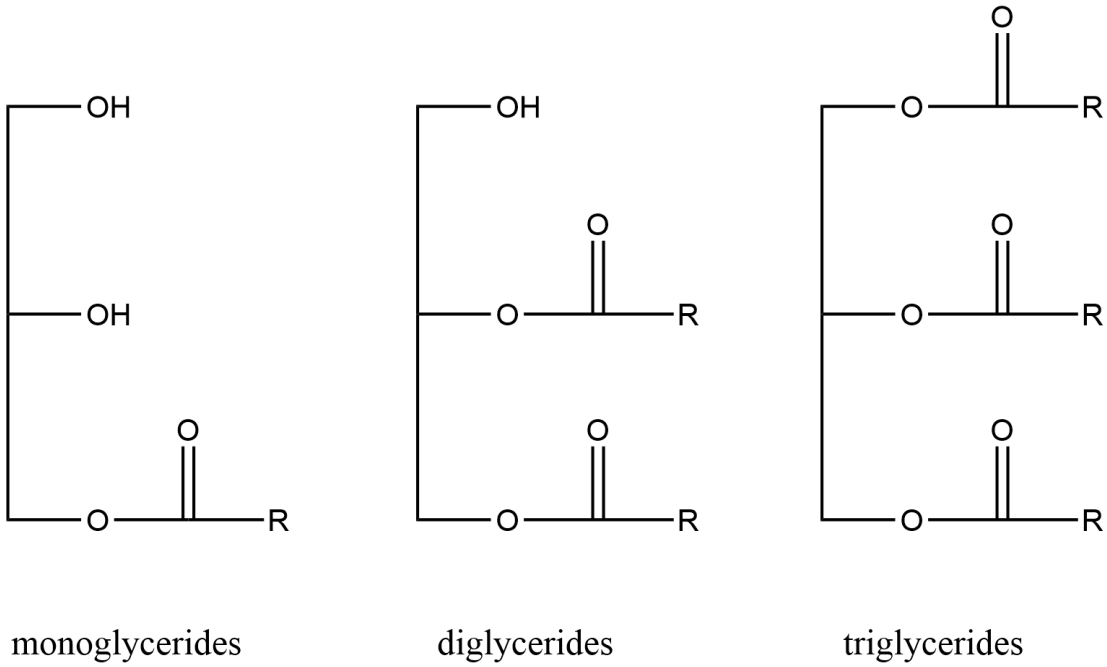


Figure 1.5 A diagram of the chemical structure of monoglycerides, diglycerides and triglycerides

Due to their higher ability to pack at the interface, low-molecular-weight emulsifiers usually have a high surface activity and also are able to adsorb more quickly onto the surface of the droplets due to their smaller sizes. In doing so, they reduce the interfacial surface energy. However, low-molecular-weight emulsifiers, and more specifically those without an electrical charge, typically do not generate a long ranged repulsion because of their size as well as the ease of adsorption-desorption (Dickinson, 1992b). Moreover, earlier research has demonstrated that numerous commercial low-molecular-weight emulsifiers negatively impact gastrointestinal and metabolic health (Cox et al., 2020).

Compared to low-molecular-weight emulsifiers, macromolecular emulsifiers do not possess as high a surface activity, meaning that they are not able to reduce the interfacial tension at the same level as the low molecular surfactants. Nonetheless, due to their much higher adsorption energies, once a protein has successfully adsorbed to the surface of a particle, it cannot be easily desorbed. This fact, together with their larger size, allows them to impart long-term colloidal stability in some extent. Biopolymers, such as proteins, and different conjugates derived from polysaccharide + proteins (Yadav et al., 2010; Xue et al., 2017; Ding et al., 2021) as well as certain

polysaccharides (Yusoff and Murray, 2011; Mu et al., 2021), are the most prevalent macromolecular emulsifiers encountered in food research. Although macromolecular emulsifiers may appear more acceptable for use in the food business as being safe and effective, the preparation of biopolymers involving conjugates still requires some extra chemicals that may have potentially detrimental effects on human health (Li, C. et al., 2015).

In addition to the two types of emulsifiers discussed above, Pickering emulsifiers, mainly composed of nanoscale-sized components such as nanoscale starch particles, organic crystals, prolamins, cellulose, and microgel particles, have grown increasingly popular as emulsifiers in recent years (Murray, 2019). The most significant difference between Pickering emulsions and classical emulsions is that Pickering emulsions are stabilised by solid particles, acting as emulsifiers at the interface of the two liquid phases. In contrast, classical emulsions are stabilised by adsorbed layers of molecular surfactants (Yang et al., 2017). In general, Pickering emulsions are highly resistant to coagulation and other destabilising processes such as Ostwald ripening and are overall also less affected by external factors. However, Pickering emulsifiers have very stringent requirements for their particle size and surface chemistry, which can present a challenge in the food industry (Binks and Lumsdon, 2001; Ge et al., 2017).

Consequently, this study focuses on emulsifiers for protein-based materials by combining the benefits and drawbacks of the previously mentioned emulsifiers.

1.3.2 Protein-based emulsifiers

Before introducing emulsifiers for protein-based materials, a distinction between stabilisers and emulsifiers should be made. Stabilisers, unlike emulsifiers, are only helpful in providing the colloids with long-term stability, preventing the aggregation of the particles, and in the case of emulsion droplets resisting coalescence too. Some stabilisers, such as pure polysaccharides and gums, cannot effectively reduce surface tension and thus form smaller particles. They provide colloidal stability by modifying the bulk properties (namely viscosity) of the dispersion medium rather than any particle-particle interactions. Proteins, on the other hand, are advantageous as emulsifiers because they combine the surface properties of emulsifiers while can act as true colloid stabilisers (Dickinson, 1992b). That is to say

that proteins manipulate the colloidal interactions between particles, inducing repulsion. Since natural proteins have a lower surface activity than small molecule emulsifiers, it is challenging to use them purely as emulsifiers to achieve small droplets in many situations. However, under appropriate conditions, protein can provide emulsions with excellent longer-term stability against coalescence (e.g., by a selection of a suitable protein and/or modification of the protein) (Dickinson, 1992b; Dalgleish, 1997; Damodaran, 2006; Dickinson, 2010).

Animal proteins, such as α_{S1}-casein and β-casein, are widely used as natural emulsifiers in the food industry. This is primarily because both α_{S1}-casein and β-casein are chain-like proteins with less secondary structure. In addition, α_{S1}-casein roughly resembles a tri-block-like polymer, whereas β-casein is closer to a di-block-like polymer (Euston and Horne, 2005), which as such, can provide excellent steric repulsion against attractive van der Walls forces. Many early studies have reported the excellent emulsification ability of α-casein (Poon et al., 2001) and β-casein (Caessens et al., 1999). However, there are limitations to using casein and other animal proteins as emulsifiers in the food industry. First, animal proteins are typically more expensive than plant proteins. Moreover, it has been proven that high animal protein intake can increase the chances of being obese (Lin et al., 2011) and the risk of inflammatory bowel disease (Jantchou et al., 2010). Another research also demonstrated a positive association between animal protein intake and cardiovascular mortality (Song et al., 2016). Also, a meat-based diet is not eco-friendly. Production of animal protein requires increased greenhouse gas emissions and needs the utilisation of proportionately more land, water and energy (Pimentel and Pimentel, 2003; Henchion et al., 2017) for the same quantity of protein. According to the relevant survey report (Vasileska and Rechkoska, 2012), the demand for livestock products worldwide is rapidly increasing every year, which causes a severe sustainability issue. Furthermore, animal-based food additives are unacceptable for some strict vegetarians or religious believers.

As an alternative to animal proteins, plant proteins may be an alternative for also being reasonable emulsifiers that make small droplets and provide long-term stability, with the added advantages of high availability, low production costs and environmental friendliness. Plant proteins can also be used as bioactive ingredients in food-grade delivery, such as protection, encapsulation and controlled release type applications (Wan et al., 2015).

However, natural proteins (especially plant proteins) are usually globular macromolecules with a compact structure. The presence of hydrophobic interactions induces proteins to minimise the exposure of hydrophobic groups to water, thereby decreasing their free energy (Pace et al., 1996; Voet et al., 2016). As a result, the use of natural plant proteins as emulsifiers poses two major challenges. One is that the absence of hydrophobic groups on the surface makes the adsorption of plant proteins on the particle surface slow and insufficient. The chains have to unfold before they can adsorb, hindering fast adsorption kinetics. The other issue is that when globular proteins are adsorbed on the surface, a thick adsorbed layer is normally not formed due to chains lying relatively flat on the surface. Such chain configurations the not suited to the provision of desirable long-ranged steric repulsion. Thus, suitable denaturation of globular proteins has a generally beneficial effect on enhancing emulsification capabilities (Dickinson, 1992b). This conclusion has been confirmed by many studies where proteins have been denatured in different ways (Jahaniaval et al., 2000; Poon et al., 2001; McClements, 2004b; Raikos, 2010; Li, F. et al., 2011). However, care has to be exercised as denatured vegetable-based proteins have a tendency to form interchain covalent bonds, such as disulfide bonds (Jang and Swaisgood, 1990). This can cause gel formation or aggregation of protein molecules, which is not at all desirable for emulsification purposes.

Suitable enzymatic hydrolysis of natural proteins can also enhance their emulsification properties. Many studies have suggested that certain fragmented proteins show a higher emulsifying capacity than untreated natural proteins (Aoki and Inami, 1980; Vojdani and Whitaker, 1998; Zhang et al., 2015; Chen et al., 2016; Ren et al., 2017). However, some studies have shown that fragmented proteins are not always suitable emulsifiers (Singh and Dalgleish, 1998). Factors that produce suitable plant-based fragmented proteins are not fully understood and well explored. The degree of hydrolysis is affected by many factors, such as pH, time, temperature, protein concentration, and the type and amount of enzyme used (Noman et al., 2018), which can significantly influence the emulsification ability of fragmented proteins. In summary, fragmented plant protein may potentially serve to provide emulsifiers that offer the same benefits as animal and natural plant proteins.

1.3.3 Fragmented soya protein as a food-grade emulsifier

The use of soya protein as a raw material for emulsification has received much attention due to its high nutritional value, wide availability and low cost. Droplets stabilised by soya protein isolate (SPI) are somewhat resistant to coalescence in solution (Puppo et al., 2005). However, SPI is usually not used as an efficient food emulsifier, particularly against droplet aggregation. As a result, many researchers have attempted to improve the emulsification properties of natural soy proteins by modifying them with various treatments, such as high pressure (Puppo et al., 2005), ultrasound (Chen et al., 2011a) and by making micro-gel particles from them (Matsumiya and Murray, 2016). In addition to the methods described above, the hydrolysis of soy protein can also potentially improve its emulsification properties, this being the focus of the current thesis. Previous studies have found improved emulsification capacity in reducing droplet size or provding long-term emulsion stability of soya protein hydrolysates under suitable enzymatic treatment with pancreatin (Qi et al., 1997), trypsin (Ochiai et al., 1982; Chobert et al., 1988), papain (Wu et al., 1998; Li, W. et al., 2016) and proteases mixture (Chen et al., 2011b). In contrast, it has also been reported that soy protein hydrolysis does not necessarily enhance emulsification properties (Jung et al., 2004). Therefore, the suitability of fragmented polypeptide chains depends on many factors that are not systematically examined. In order to obtain a more precise fragmented soy protein with good emulsification properties and hence also to better understand the mechanism by which fragmented proteins stabilise emulsions, this PhD study attempts to investigate the interfacial behaviour of fragmented soya protein from both a theoretical (Chapter 3 - 4) and also experimental (Chapter 5 - 6) perspectives.

1.4 Aims and objectives

With the improvement of knowledge of people regarding their health, food safety issues and the impact of foods on well-being are becoming increasingly more significant concerns. However, it is difficult to balance safety and cost-effectiveness with common emulsifiers in the food industry. For example, casein is a reasonable consumer-friendly natural emulsifier, but it is a challenge for industrial production due to its high cost. Our work in this thesis aims to investigate the possibility of using fragmented plant proteins as an innovative emulsifier by combining theoretical and experimental approaches. It is expected that in the future, fragmented

protein emulsifiers could partially replace common animal-based emulsifiers and other macromolecular emulsifiers.

To attain our aims, we have established the following objectives:

1. Determine the raw protein specifically used to prepare fragmented protein emulsifiers.
2. Develop innovative methods for fast screening and selecting fragmented protein as suitable emulsifiers and evaluate the efficacy of such approaches compared to the practical work.
3. Apply self-consistent field calculation (SCFC) to further evaluate in more detail the different interfacial properties of the protein fragments screened as potentially suitable in step (2). On this basis, we will discuss the challenges and solutions of translating theory into practice.
4. Experimentally investigate the ability of diverse protein hydrolysates obtained from soya protein to reduce the surface tension at air-water interfaces so that we can home in on some hydrolysates that may serve as suitable emulsifiers.
5. Investigate the stability for coalescence and aggregation of emulsions prepared by different hydrolysates using droplet size measurement and visual inspection of creaming.
6. Sequencing analysis of the components of the hydrolysates that showed excellent emulsification capabilities, together in conjunction with the results of SCF calculations, provide additional evidence for the conclusion that fragmented proteins could also be suitable emulsifiers.

1.5 Thesis outline

This thesis contains seven chapters and two appendices, as follows.

Chapter 1 introduces the fundamentals of colloid science, focusing primarily on destabilisation and stabilisation mechanism in colloidal systems. The scientific background of this thesis is then described by following a discussion of common emulsifiers, followed by protein-based emulsifiers, then animal protein emulsifiers, plant protein emulsifiers, and finally,

fragmented plant protein emulsifiers. The necessity of developing innovative emulsifiers based on fragmented protein is explained.

Chapter 2 shows the framework of the research methodology used in this thesis. We introduce the detailed principles of the SCF calculation method, starting from single-component systems and moving to multi-component systems.

Chapter 3 presents two fast methods for screening fragmented proteins to filter out and identify potentially suitable emulsifiers. These methods are novel and were developed as part of the work in this thesis. Moreover, the application of our methods suggested a particular fragment involving a specific sequence of amino acid residues originating from soybean protein may have excellent emulsification ability.

Chapter 4 applies SCF calculations to further analyse the interfacial properties of the soya protein-derived fragments identified in Chapter 3. The adsorption amount, volume fraction, and spatial conformation of fragments on a hydrophilic-hydrophobic interface are studied. Furthermore, we investigate, calculate and discuss in detail the colloidally induced interaction potentials by two adjacent protein fragment layers, adsorbed on a pair of neighbouring particles, as these approach each other. At the end of this chapter, the advantages and disadvantages of the fast screening and the SCF calculation methods are summarised and discussed.

Chapter 5 combines theoretical and experimental results. It first explains the difficulty of obtaining peptides with a specific sequence, as identified through theoretical calculations, in a practical setting. It then emphasises the importance of the size of the polypeptide on its emulsification ability. A revised experimentation protocol was developed based on the results of the theoretical calculations. In Chapter 5, the surface tension of the different hydrolysates was measured to identify the experimental materials used to prepare fragment protein stabilised emulsions for work presented in Chapter 6.

Chapter 6 presents the emulsification performance of two soy protein hydrolysates obtained under different conditions. We show that one of the hydrolysate samples performed very well in the emulsion experiment. This was experimentally sequenced, and the results were then once again used in SCF calculations for further analysis of its interfacial and emulsifying properties.

Chapter 7 summarises and discusses the results from Chapters 3 to 6. Combined with current results, this chapter provides an outlook and discussion of future work.

Appendix A presents the source code of some of the programs used in this thesis, mainly including the procedure of the SCF calculation method and the modified SCF method for special situations. Appendix B presents the figures for the relation between surface tension and the length of the measuring time.

List of References

Aoki, H. and Inami, M. 1980. Emulsifying properties of soy protein: Characteristics of 7S and IIS proteins.

Binks, B. and Lumsdon, S. J. L. 2001. Pickering emulsions stabilized by monodisperse latex particles: effects of particle size. 17(15), pp.4540-4547.

Caessens, P. W., Gruppen, H., Slangen, C. J., Visser, S. and Voragen, A. G. 1999. Functionality of β-casein peptides: Importance of amphipathicity for emulsion-stabilizing properties. Journal of Agricultural Food Chemistry. 47(5), pp.1856-1862.

Calvo, F., Gómez, J. M., Ricardez-Sandoval, L. and Alvarez, O. 2020. Integrated design of emulsified cosmetic products: A review. Chemical Engineering Research Design. 161, pp.279-303.

Chapman, D. L. 1913. LI. A contribution to the theory of electrocapillarity. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 25(148), pp.475-481.

Charles, S. 1992. Magnetic fluids (ferrofluids). Magnetic properties of fine particles. Elsevier, pp.267-276.

Chen, L., Chen, J., Ren, J. and Zhao, M. 2011a. Effects of Ultrasound Pretreatment on the Enzymatic Hydrolysis of Soy Protein Isolates and on the Emulsifying Properties of Hydrolysates. Journal of Agricultural and Food Chemistry. 59(6), pp.2600-2609.

Chen, L., Chen, J., Ren, J. and Zhao, M. 2011b. Modifications of soy protein isolates using combined extrusion pre-treatment and controlled enzymatic hydrolysis for improved emulsifying properties. Food Hydrocolloids. 25(5), pp.887-897.

Chen, L., Chen, J., Yu, L. and Wu, K. 2016. Improved emulsifying capabilities of hydrolysates of soy protein isolate pretreated with high pressure microfluidization. LWT - Food Science and Technology. 69, pp.1-8.

Chobert, J. M., Bertrand-Harb, C. and Nicolas, M. G. 1988. Solubility and emulsifying properties of caseins and whey proteins modified enzymically by trypsin. Journal of Agricultural and Food Chemistry. 36(5), pp.883-892.

Cox, S., Sandall, A., Smith, L., Rossi, M. and Whelan, K. 2020. Food additive emulsifiers: a review of their role in foods, legislation and classifications, presence in food supply, dietary exposure, and safety assessment. Nutrition Reviews. 79(6), pp.726-741.

Dalgleish, D. G. 1997. Adsorption of protein and the stability of emulsions. Trends in Food Science and Technology. 8(1), pp.1-6.

Damodaran, S. 2006. Protein Stabilization of Emulsions and Foams. Journal of Food Science. 70(3), pp.R54-R66.

Derjaguin, B. 1993. A theory of interaction of particles in presence of electric double layers and the stability of lyophobe colloids and disperse systems. Progress in Surface Science. 43(1-4), pp.1-14.

Derjaguin, B. and Landau, L. 1993. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. Progress in Surface Science. 43(1-4), pp.30-59.

Dickinson, E. 1992. Introduction to food colloids. Oxford university press.

Dickinson, E. 2010. Flocculation of protein-stabilized oil-in-water emulsions. Colloids and Surfaces B: Biointerfaces. 81(1), pp.130-140.

Dickinson, E. and Stainsby, G. 1982. Colloids in food. Applied Science Publishers.

Ding, Y., Chen, L., Shi, Y., Akhtar, M., Chen, J. and Ettelaie, R. 2021. Emulsifying and emulsion stabilizing properties of soy protein hydrolysates, covalently bonded to polysaccharides: The impact of enzyme choice and the degree of hydrolysis. Food Hydrocolloids. 113, p106519.

Ettelaie, R., Zengin, A. and Lishchuk, S. V. 2017. Novel food grade dispersants: Review of recent progress. Current Opinion in Colloid Interface Science. 28, pp.46-55.

Euston, S. R. and Horne, D. S. 2005. Simulating the self-association of caseins. Food Hydrocolloids. 19(3), pp.379-386.

Feng, J., Shi, Y., Yu, Q., Sun, C. and Yang, G. 2016. Effect of emulsifying process on stability of pesticide nanoemulsions. Colloids Surfaces A: Physicochemical Engineering Aspects. 497, pp.286-292.

Ge, S., Xiong, L., Li, M., Liu, J., Yang, J., Chang, R., Liang, C. and Sun, Q. 2017. Characterizations of Pickering emulsions stabilized by starch nanoparticles: Influence of starch variety and particle size. Food Chemistry. 234, pp.339-347.

Gouy, M. 1910. Sur la constitution de la charge électrique à la surface d'un électrolyte. J. Phys. Theor. Appl. 9(1), pp.457-468.

Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M. and Tiwari, B. 2017. Future Protein Supply and Demand: Strategies and Factors Influencing a Sustainable Equilibrium. Foods (Basel, Switzerland). 6(7), p53.

Hwang, Y.-J., Oh, C. and Oh, S.-G. 2005. Controlled release of retinol from silica particles prepared in O/W/O emulsion: The effects of surfactants and polymers. Journal of Controlled Release. 106(3), pp.339-349.

Jahaniaval, F., Kakuda, Y., Abraham, V. and Marcone, M. F. 2000. Soluble protein fractions from pH and heat treated sodium caseinate:

Physicochemical and functional properties. Food Research International. 33(8), pp.637-647.
Jang, H. D. and Swaisgood, H. E. 1990. Disulfide Bond Formation Between Thermally Denatured β-Lactoglobulin and κ-Casein in Casein Micelles. Journal of Dairy Science. 73(4), pp.900-904.
Jantchou, P., Morois, S., Clavel-Chapelon, F., Boutron-Ruault, M.-C. and Carbonnel, F. 2010. Animal protein intake and risk of inflammatory bowel disease: The E3N prospective study. American journal of gastroenterology. 105(10), pp.2195-2201.
Jones, A., Vincent, B. J. C. and Surfaces. 1989. Depletion flocculation in dispersions of sterically-stabilised particles 2. Modifications to theory and further studies. 42(1), pp.113-138.
Jung, S., Roussel‐Philippe, C., Briggs, J. L., Murphy, P. A. and Johnson, L. A. 2004. Limited hydrolysis of soy proteins with endo‐and exoproteases. Journal of the American Oil Chemists' Society. 81(10), p953.
Krog, N. 2002. Emulsifiers and emulsions in dairy foods. In: Roginski, H. ed. Encyclopedia of Dairy Sciences. Oxford: Elsevier, pp.891-900.
La Mer, V. K. 1966. Filtration of colloidal dispersions flocculated by anionic and cationic polyelectrolytes. Discussions of the Faraday Society. 42, pp.248-254.
Li, C., Wang, J., Shi, J., Huang, X., Peng, Q. and Xue, F. 2015. Encapsulation of tomato oleoresin using soy protein isolate-gum aracia conjugates as emulsifier and coating materials. Food Hydrocolloids. 45, pp.301-308.
Li, F., Kong, X., Zhang, C. and Hua, Y. 2011. Effect of heat treatment on the properties of soy protein‐stabilised emulsions. International journal of food science technology. 46(8), pp.1554-1560.
Li, W., Wang, Y., Zhao, H., He, Z., Zeng, M., Qin, F. and Chen, J. 2016. Improvement of emulsifying properties of soy protein through selective hydrolysis: Interfacial shear rheology of adsorption layer. Food Hydrocolloids. 60, pp.453-460.
Lin, Y., Bolca, S., Vandevijvere, S., De Vriese, S., Mouratidou, T., De Neve, M., Polet, A., Van Oyen, H., Van Camp, J. and De Backer, G. 2011. Plant and animal protein intake and its association with overweight and obesity among the Belgian population. British journal of nutrition. 105(7), pp.1106-1116.
Matsumiya, K. and Murray, B. S. 2016. Soybean protein isolate gel particles as foaming and emulsifying agents. Food Hydrocolloids. 60, pp.206-215.

McClements, D. J. 2004a. Food emulsions: principles, practices, and techniques. CRC press.

McClements, D. J. 2004b. Protein-stabilized emulsions. Current Opinion in Colloid & Interface Science. 9(5), pp.305-313.

Miyazawa, T., Itaya, M., Burdeos, G., Nakagawa, K. and Miyazawa, T. 2021. A Critical Review of the Use of Surfactant-Coated Nanoparticles in Nanomedicine and Food Nanotechnology. International Journal of Nanomedicine. Volume 16, pp.3937-3999.

Mu, M., Karthik, P., Chen, J., Holmes, M. and Ettelaie, R. 2021. Effect of amylose and amylopectin content on the colloidal behaviour of emulsions stabilised by OSA-Modified starch. Food Hydrocolloids. 111, p106363.

Murray, B. S. 2019. Pickering emulsions for food and drinks. Current Opinion in Food Science. 27, pp.57-63.

Murray, B. S., Ettelaie, R., Sarkar, A., Mackie, A. R. and Dickinson, E. 2021. The perfect hydrocolloid stabilizer: Imagination versus reality. Food Hydrocolloids. 117, p106696.

Noman, A., Xu, Y., Al-Bukhaiti, W. Q., Abed, S. M., Ali, A. H., Ramadhan, A. H. and Xia, W. 2018. Influence of enzymatic hydrolysis conditions on the degree of hydrolysis and functional properties of protein hydrolysate obtained from Chinese sturgeon (Acipenser sinensis) by using papain enzyme. Process Biochemistry. 67(September 2017), pp.19-28.

Obey, T. M. and Griffiths, P. 1999. Polymer adsorption: fundamentals. Principles of polymer science and technology in cosmetic and personal care. Marcel Dekker Inc., pp.51-71.

Ochiai, K., Kamata, Y. and Shibasaki, K. 1982. Effect of tryptic digestion on emulsifying properties of soy protein. Agricultural Biological Chemistry. 46(1), pp.91-96.

Okochi, H. and Nakano, M. 2000. Preparation and evaluation of w/o/w type emulsions containing vancomycin. Advanced drug delivery reviews. 45(1), pp.5-26.

Pace, C. N., Shirley, B. A., McNutt, M. and Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. The FASEB journal. 10(1), pp.75-83.

Pathak, M. 2017. Chapter 5 - Nanoemulsions and Their Stability for Enhancing Functional Properties of Food Ingredients. In: Oprea, A.E. and Grumezescu, A.M. eds. Nanotechnology Applications in Food. Academic Press, pp.87-106.

Pimentel, D. and Pimentel, M. 2003. Sustainability of meat-based and plant-based diets and the environment. The American Journal of Clinical Nutrition. 78(3), pp.660S-663S.

Poon, S., Clarke, A. E. and Schultz, C. J. 2001. Effect of denaturants on the emulsifying activity of proteins. Journal of Agricultural Food Chemistry 49(1), pp.281-286.

Puppo, M. C., Speroni, F., Chapleau, N., de Lamballerie, M., Añón, M. C. and Anton, M. 2005. Effect of high-pressure treatment on emulsifying properties of soybean proteins. Food Hydrocolloids. 19(2), pp.289-296.

Qi, M., Hettiarachchy, N. and Kalapathy, U. 1997. Solubility and emulsifying properties of soy protein isolates modified by pancreatin. Journal of Food Science. 62(6), pp.1110-1115.

Raikos, V. 2010. Effect of heat treatment on milk protein functionality at emulsion interfaces. A review. Food Hydrocolloids. 24(4), pp.259-265.

Ratke, L. and Voorhees, P. W. 2002. Growth and coarsening: Ostwald ripening in material processing. Springer Science & Business Media.

Ren, J., Song, C. L., Zhang, H. Y., Kopparapu, N. K. and Zheng, X. Q. 2017. Effect of Hydrolysis Degree on Structural and Interfacial Properties of Sunflower Protein Isolates. Journal of Food Processing and Preservation. 41(1), pp.1-7.

Shepherd, I. and Yoell, R. 1976. Cake emulsions. Food emulsions.

Singh, A. M. and Dalgleish, D. G. 1998. The Emulsifying Properties of Hydrolyzates of Whey Proteins. Journal of Dairy Science. 81(4), pp.918-924.

Song, M., Fung, T. T., Hu, F. B., Willett, W. C., Longo, V. D., Chan, A. T. and Giovannucci, E. L. 2016. Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality. JAMA Internal Medicine. 176(10), pp.1453-1463.

Stern, O. 1924. Zur theorie der elektrolytischen doppelschicht. Zeitschrift für Elektrochemie und angewandte physikalische Chemie. 30(21‐22), pp.508-516.

Vasileska, A. and Rechkoska, G. 2012. Global and Regional Food Consumption Patterns and Trends. Procedia - Social and Behavioral Sciences. 44, pp.363-369.

Verwey, E. J. W. and Overbeek, J. T. G. 1955. Theory of the stability of lyophobic colloids. Journal of Colloid Science. 10(2), pp.224-225.

Voet, D., Voet, J. G. and Pratt, C. W. 2016. Fundamentals of biochemistry: life at the molecular level. John Wiley & Sons.

Vojdani, F. and Whitaker, J. R. 1998. Limited proteolysis of α-lactalbumin and whey protein isolate: Effect on their functional properties. ACS Publications.

Walstra, P. 2002. Physical chemistry of foods. CRC Press.

Wan, Z.-L., Guo, J. and Yang, X.-Q. 2015. Plant protein-based delivery systems for bioactive ingredients in foods. Food Function
6(9), pp.2876-2889.

Wu, W. U., Hettiarachchy, N. S. and Qi, M. 1998. Hydrophobicity, solubility, and emulsifying properties of soy protein peptides prepared by papain modification and ultrafiltration. Journal of the American Oil Chemists' Society. 75(7), pp.845-850.

Xue, L.-H., Xie, C.-Y., Meng, S.-X., Bai, R.-X., Yang, X., Wang, Y., Wang, S., Binks, B. P., Guo, T. and Meng, T. 2017. Polymer–protein conjugate particles with biocatalytic activity for stabilization of water-in-water emulsions. ACS Macro Letters. 6(7), pp.679-683.

Yadav, M. P., Parris, N., Johnston, D. B., Onwulata, C. I. and Hicks, K. B. 2010. Corn fiber gum and milk protein conjugates with improved emulsion stability. Carbohydrate Polymers. 81(2), pp.476-483.

Yang, Y., Fang, Z., Chen, X., Zhang, W., Xie, Y., Chen, Y., Liu, Z. and Yuan, W. 2017. An overview of Pickering emulsions: solid-particle materials, classification, morphology, and applications. Frontiers in pharmacology. 8, p287.

Yusoff, A. and Murray, B. S. 2011. Modified starch granules as particle-stabilizers of oil-in-water emulsions. Food Hydrocolloids. 25(1), pp.42-55.

Zhang, Y., Tan, C., Abbas, S., Eric, K., Xia, S. and Zhang, X. 2015. Modified SPI improves the emulsion properties and oxidative stability of fish oil microcapsules. Food Hydrocolloids. 51, pp.108-117.

Chapter 2
Methodology

2.1 Introduction

The theoretical part of this work consists of two main parts. Its operational flow can be seen in Figure 2.1. Firstly, the innovative fast screening methods filter out a sufficient number of candidate protein fragments with the potential to be suitable emulsifiers in providing sufficient repulsion among droplets based on their overall arrangement of primary amino acid sequence. These fragments are then further analysed in more detail using self-consistent field calculations (SCFC) to identify their interfacial adsorption behaviour and predicted their ability to provide sufficient steric repulsion necessary for them to perform well as emulsifiers. Those that pass this second stage of calculations are then suggested to contribute to future experiments.

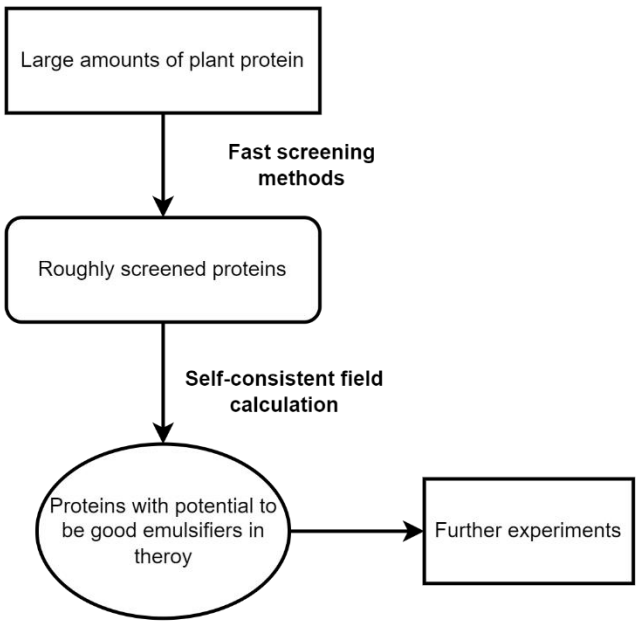


Figure 2.1 Process of identifying protein fragments that may be potentially good emulsifiers by a quick screening method followed by more detailed (but computer resource intensive) SCF calculation

The programs in the theoretical section below were developed in Python and Fortran programming languages, with some of the source codes presented in Appendix A at the end of the thesis. The specific sequences of all the primary structures of the proteins used in this study were obtained from the previous papers and NCBI database.

2.2 Fast screening methods

As mentioned in Chapter 1, protein fragments can reduce or improve emulsification properties to varying degrees compared to intact proteins. It is also not feasible to perform experimental work on all the many different types of fragments that can be derived from one type of protein, let alone from many different types. Therefore, it is necessary to predict the emulsification performance of different protein fragments through theoretical analysis first before further, more focused practical experiments are conducted. The theoretical work will, on the one hand, determine the direction of the experiment and, on the other hand, significantly reduce the time required for the experimentation.

To demonstrate the above, let us focus on just one protein. The number of possible primary structures of peptides it can form is enormous. For example, if there is a protein consisting of 150 amino acids, and supposing that a peptide fragment is obtained from this protein in a set of 50-150 amino acids (for simplicity, a peptide of less than 50 amino acids is too short to be considered here for the moment). Then, permutation of up to

$$\sum_{i=0}^{150-50}\left(150-50-i+1\right)=\frac{1}{2}(150-50+1)(150-50+2)=5151$$

5151 protein fragments would be obtained. Assuming that we need 10 minutes to analyse a possible fragment using more detailed SCF calculations, it would take up about 36 days to analyse the entire permutation derived from just this one type of protein. Moreover, focusing only on one protein in our research is not desirable, as this method tries to rapidly screen fragments from many different proteins. It is impractical to try to exhaust the entire set of primary structures. Therefore, this is the motivation and the need for developing fast screening methods.

Before we describe the specific workings of our rapid screening methods below, we need to understand the basis and rationale for screening protein fragments suitable as emulsifiers. In past theoretical studies, α_{s1}-casein and β-casein were considered excellent natural protein emulsifiers (Dickinson, 1997; Dickinson et al., 1998; Dickinson, 2001). The reason behind this is thought to be often that the former is considered to act somewhat like a tri-block-like polymer, while the latter has adsorption behaviour associated with di-block-like polymers (Euston and Horne, 2005).

Table 2.1 Classification of different types of amino acid residues into five groups

(Ettelaie et al., 2014)

Type	Residues
1 Hydrophobic	Pro, Ile, Leu, Val, Phe, Ala, Met, Trp
2 Polar (non-charged)	Gln, Asn, Ser, Thr, Tyr
3 Positive	Arg, Lys, N-terminus
4 Positive	His
5 Negative	Glu, Asp, C-terminus

This work divides amino acids into two categories based on their interfacial properties: hydrophilic amino acids (amino acids II to V in Table 2.1) and hydrophobic (amino acids I in Table 2.1) residues. A fragment consisting of consecutive hydrophobic or hydrophilic residues is called a block. Perfect di-block fragments will consist of just a single hydrophobic and a single hydrophilic block, as demonstrated in Figure 2.2, while tri-block-like fragments are of two kinds. One is composed of a hydrophilic block with two hydrophobic blocks at each end (as is the case with α_{s1}-casein), and the other is composed of a hydrophobic block and two hydrophilic blocks.

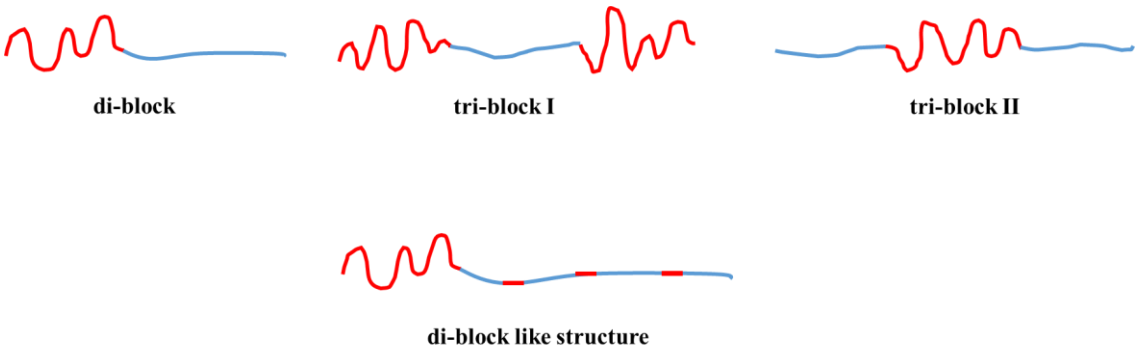


Figure 2.2 Primary structures of perfect di-block and tri-block polymers (red represents the hydrophobic and blue the hydrophilic blocks)

Proteins made up entirely of two or three consecutive blocks (i.e. ideal di-block or tri-block structures) do not exist in nature. Even for α_{s1}-casein or β-casein, it is only that the density of the same type of amino acid (i.e. hydrophobic or hydrophilic) is higher in certain extended sections along the protein backbone. A largely hydrophilic part may still contain a small number of more hydrophobic residues and vice versa. In a rapid screening approach, we try to find protein fragments that have a similarity to the di-block or tri-block structure. While it is very easy to identify perfect di-block, tri-block, or even higher multi-block sequences, it is not trivial to do so for structures that resemble such sequences but not exactly (see the di-block-like structure in Figure 2.2). The reasons for our efforts to find both of these structures are explained in Figure 2.3 below.

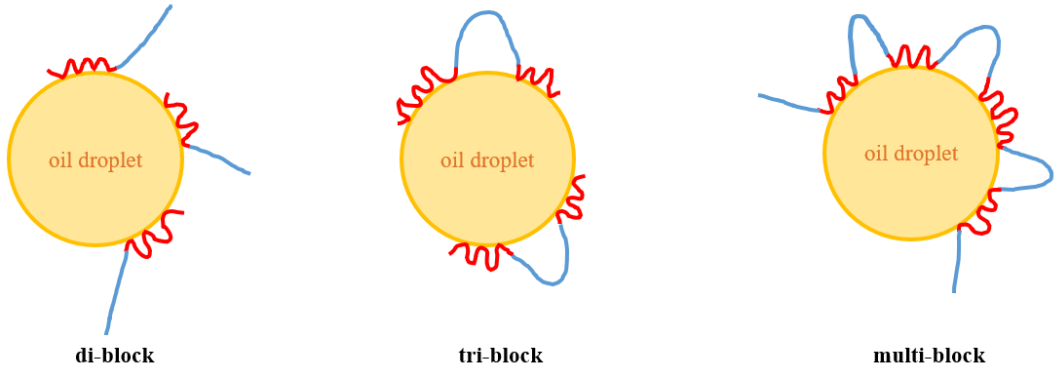


Figure 2.3 Predicted behaviour of protein fragments with different numbers of blocks at hydrophobic-hydrophilic interfaces

In Figure 2.3, the di-block fragment will form a dangling end protruding into the solution and a train lying flat on the surface when it adsorbs to the surface of the droplets; the hydrophobic group will adsorb to the surface,

while the hydrophilic group will extend into the aqueous solution. However, when the number of fragment blocks increases to three (i.e., tri-block), we encounter two situations: (1) if the distance between the two particles is sufficiently large, the protein fragments form loop conformations upon adsorption on the surface of a single particle, with two trains (blocks) lying flat on the surface. In this case, the hydrophilic group of this fragment does not extend completely into the aqueous solution, and the thickness of the adsorbed layer that the tri-block fragment can form will be thinner than that of a di-block fragment of the same length and chemical composition. This affects the range of the steric repulsive force that the emulsifier can provide. (2) If the separation distance between two particles is sufficiently small, then adsorbed trains can reside on the surface of both particles simultaneously, eventually forming a bridge-like conformation. This leads to the phenomenon of bridging flocculation. As the number of blocks continues to increase, protein fragment-based emulsifiers will have more difficulty forming thick adsorption layers. They will also be more likely to form bridges between two droplets at close particle-particle separations, thus reducing emulsification capacity. The rapid screening method aims to identify protein fragments with a structure that resembles some extent, the three blocks similar to that of α_{s1}-casein or di-block more like β-casein. In order to make the screening as quick as possible, the fast method focuses only on the distribution of hydrophobic and hydrophilic groups in the protein chain, leaving a more detailed examination of SCF-type calculations once fragments pass this initial screening.

2.2.1 Coarse graining method

The coarse graining method is the first of the quick screening methods that we devised in the course of the present work to filter out the fragments which are very unlikely to be suitble emulsifiers. The name coarse graining itself appears in many different (and unrelated) simulation techniques and may have been inspired by the reality of the coarse drying process of grains, where the grains with suitable size are selected by a screening machine. Similarly, we used the definitions in Table 2.1 to separate different kinds of amino acid residues into two groups, hydrophobic and hydrophilic. Subsequently, for a protein chain of length n, a new shorter fragment of length n/2 is obtained by combining every two adjacent amino acids in a single entity. The algorithm for combining amino acids into a single unit can be seen in Figure 2.4a below.

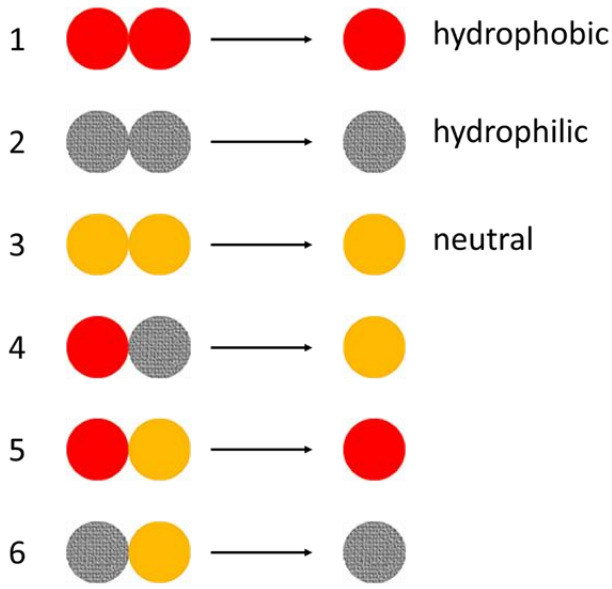


Figure 2.4a The algorithmic procedure for combining two amino acid units into a single entity in our coarse graining method

Undoubtedly, adding two hydrophobic/hydrophilic groups must be a hydrophobic/hydrophilic group. It is important to note that a neutral group is created when a hydrophobic group is combined with a hydrophilic group. The term neutral here does not mean that, in practice, the electric charge of the fragment is neutral or that it is not highly hydrophobic. In the coarse-graining method, the term ‘neutral’ means zero density of hydrophobic and hydrophilic groups in the combined unit. We start the process of coarse-graining the proteins to be examined by following the procedure shown in Figure 2.4b.

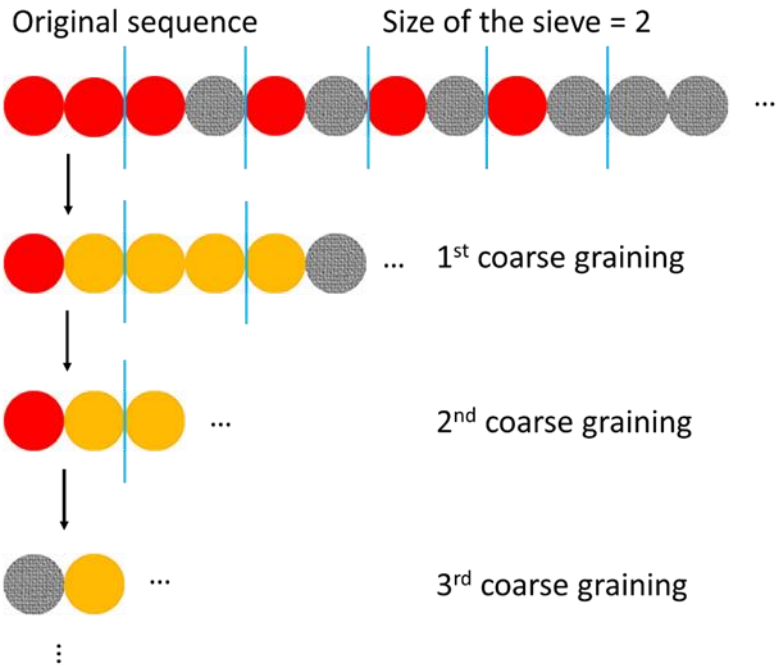


Figure 2.4b Flow chart of our coarse-graining method at different steps of coarse-graining

As the coarse-grain method proceeds, the length of the fragment decreases as units are combined into larger and larger entities. Eventually, of course, we end up with just a single until. However, it does not make much sense to analyse a super-short (only 1 to 3 in length) fragment. Interesting results arise at the intermediate stages of this process. Therefore, we need to identify a suitable point at the end of the coarse-graining process. In this study, we first calculate the length of the longest and consecutive hydrophilic block in each step of the coarse-graining method and then divide this length by the total length of the current fragment to calculate the proportion of fragments with the longest hydrophilic block (Equation 2.1). Finally, we take the point at which this ratio is at its maximum as the endpoint.

$$P_{hydrophilic} = \frac{\text{Length of the longest hydrophilic block}}{\text{Length of the fragment}} \tag{2.1}$$

In order to verify the reliability of this method, here we examine it with α_{s1}-casein and β-casein as subjects (see Figure 2.5).

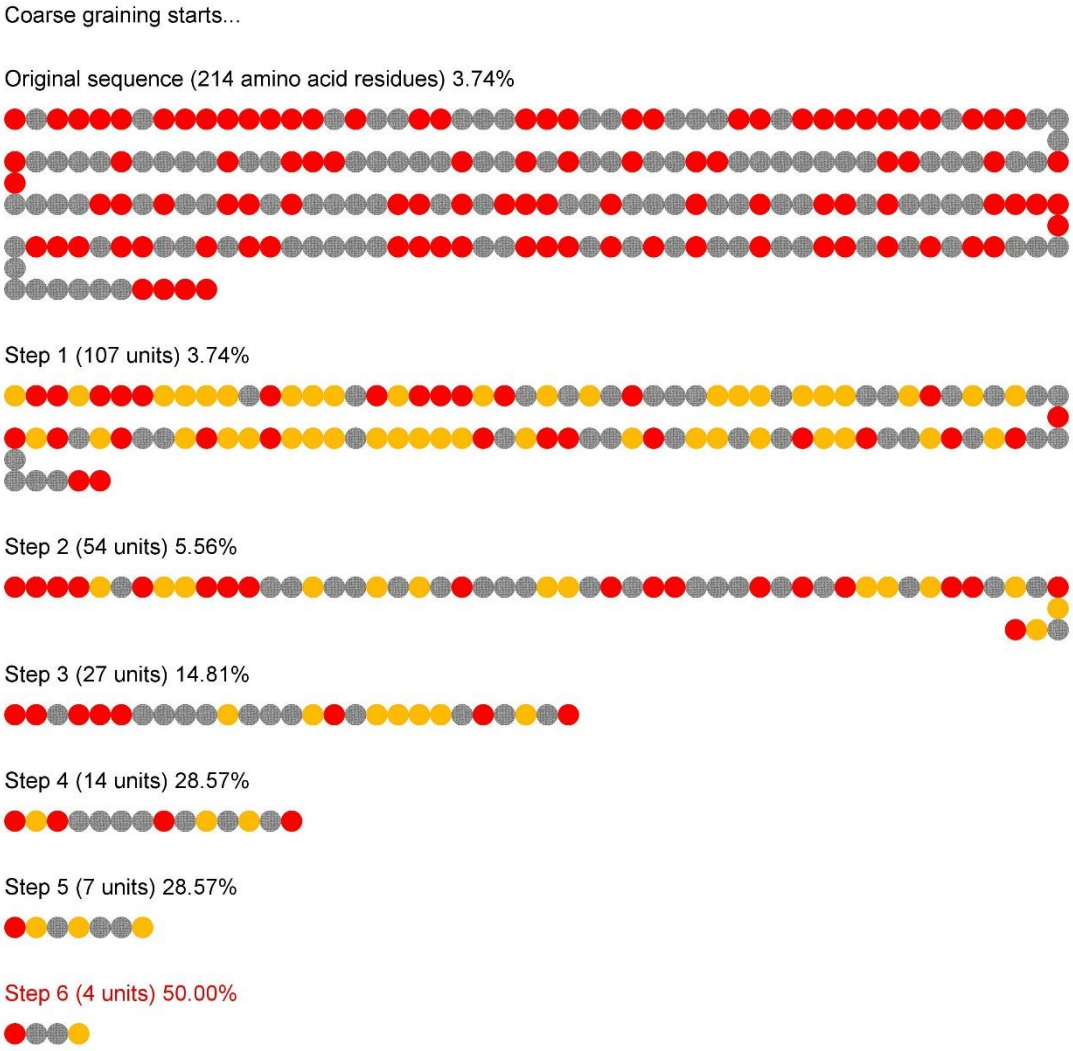


Figure 2.5a The process of the coarse-graining method as applied to α_{s1}-casein

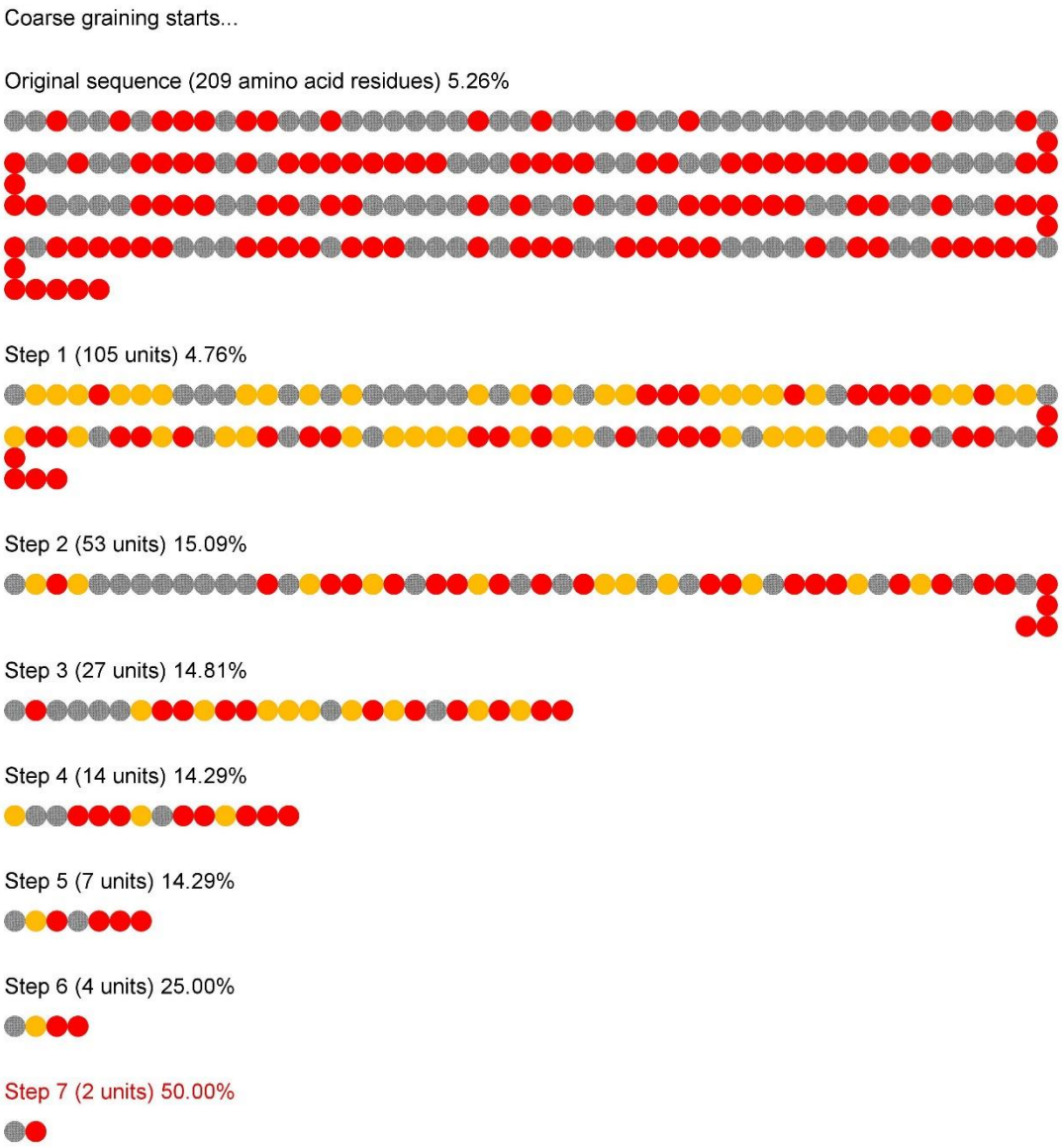


Figure 2.5b The process of the coarse-graining method as applied to β-casein

Let us first take α_{s1}-casein as an example, which after six steps, has been reduced from its original length of 214 to just four units. At this point, we can see that α_{s1}-casein does approximate the structure of a tri-block: it is hydrophobic near the N-terminus, neutral near the C-terminus, and the middle part is relatively hydrophilic. This result is also consistent with empirical knowledge and practical findings (Horne, 2017). Furthermore, this result shows that the hydrophobicity near the N-terminus is higher than that of the C-terminus side. Its primary structure is more apparent after the coarse-graining method. Turning to β-casein, when analysed directly from step 6 results (see Fig 5b), it is seen that this is a relatively hydrophobic

protein with a di-block structure. There are two hydrophobic groups and one neutral group amongst the four units arising in step 6.

In summary, these few well-known cases tend to suggest that the results of coarse-graining are in general agreement with the results of the experiments and what was already known for these well-known proteins. It is best to find di-block fragments with a reasonably large size, but considering that there may not be many such di-block-like fragments, we will also additionally consider fragments with a tri-block structure for further analysis. In this study, with this quick screening method, we can quickly identify the structure that a peptide chain is likely to possess. Fragments with more than three blocks with a relatively low potential to become emulsifiers are ignored, and the focus was only on fragments with di-block or tri-block-like structures to be carried forward to the next stage for more detailed calculations.

2.2.2 Moving average method

The moving average method is another possible quick screening method that we devised and employed in this study. The idea for this comes from the fact that the moving average method in statistical analysis is often used to predict the trend of a dependent quantity y with a variable x. It has been widely used in weather analysis or financial forecasting areas. In our study, we have innovatively applied this method to look at the distribution of two types of amino acids in a protein chain. Figure 6 illustrates the basic principle of our moving average method.

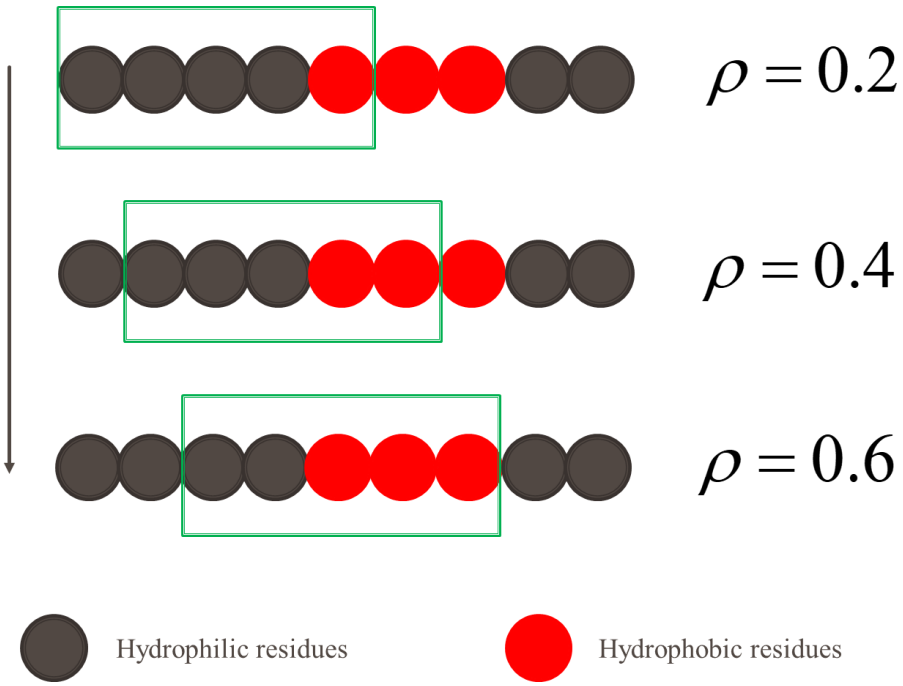


Figure 2.6 The basic idea of the moving average method involves taking a moving average of the number of hydrophobic residues taken over a set of N consecutive monomers (in the case shown S=5); ρ is the percentage of hydrophobic amino acids in a sequence defined in following (Equation 2.2)

First, we need to determine the length S that is needed for taking the average over, as seen in Figure 2.6 (where for example, S was set to 5). Typically in normal statistical analysis, the value S will be chosen to reflect some underlying periodicity. For example, for data measured over many days, S may be chosen to be 7, indicating a likely weekly variation needed to be smoothed out. For data measured over different seasons, S=4. Unfortunately, a primary sequence of proteins has no such obvious periodic variation and choosing the value of S becomes somewhat arbitrary. Nonetheless, smoothing the rapid variation in hydrophobicity moving from one amino acid residue to the next can help reveal underlying trends involving larger blocks. We count the percentage of hydrophobic amino acids in a sequence of this chosen length S (Equation 2.2).

$$\rho = \frac{N_{hydrophobic}}{S} \tag{2.2}$$

where ρ is the density of the hydrophobic residues, S is the averaged block

length and $N_{hydrophobic}$ is the number of hydrophobic residues in the N consecutive residues being considered.

We start at one end of the protein chain to be analysed, finding ρ as given by Equation 2.2 for 1 to S monomers. Then we move to residues from 2 to S+1, repeating the process until the final average is taken for N_{poly}–S to N_{poly} amino acids as one moves from one end of the protein of size N_{poly} to its other end. Thus, a series of {y} can be derived along the chain. The distribution of hydrophobic and hydrophilic amino acids can then be visualised by plotting the number of steps along the chain on the x-axis and the averaged hydrophobicity result ρ, on the y-axis, as demonstrated in figure 2.7.

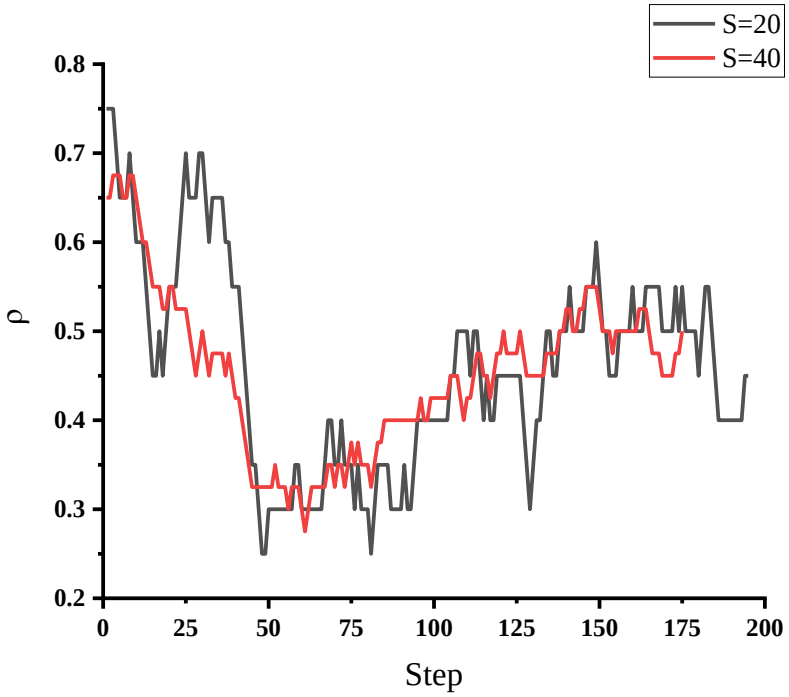


Figure 2.7a Application of moving average to the α_{s1}-casein sequence

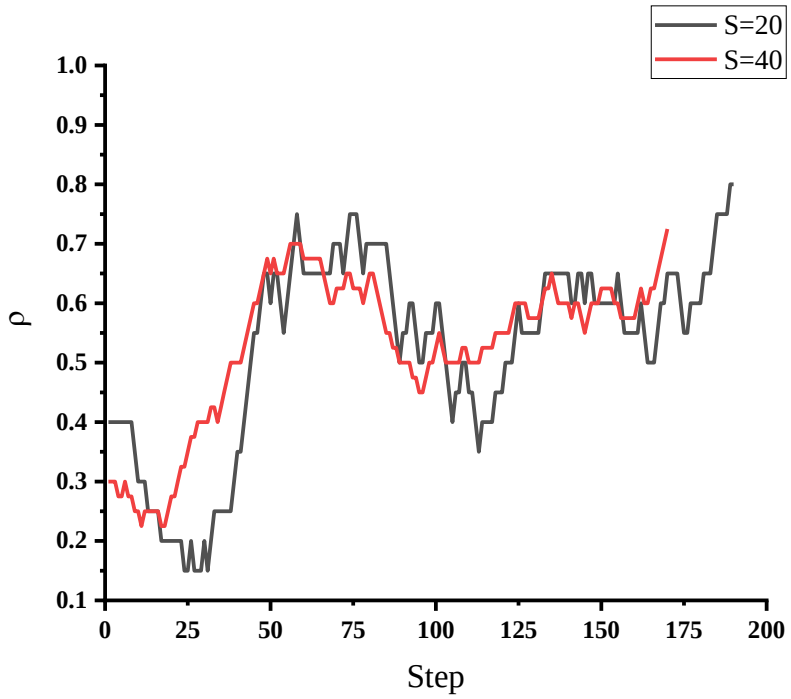


Figure 2.7b Application of moving average to the β-casein sequence

Figure 2.7 Application of moving average to the two types of casein proteins (Ratio of hydrophobic groups vs step along the chain backbone, starting from N terminus)

Here we have selected two values of S for the moving average method. These were 20 and 40. As shown in Figure 2.7, the curve with a step size of 40 is relatively flat, while the curve with a step size of 20 still shows more frequent variations. Nonetheless, the trend in the two curves for S=20 and S=40 are seen to be very similar for both caseins. As defined in Equation 2.2, more points can be obtained in the curve if a shorter step size is chosen. This, though, has less impact on smoothing the variation of hydrophobicity along the original protein chain, and the calculated curve is closer to the actual situation. The disadvantage is that the changes in the curve are then too rapid and complex to visualise, and it is not so easy to identify large blocks that are predominantly hydrophobic or hydrophilic. In the most extreme case, assuming S=1, the calculated curve is exactly equivalent to the completely unprocessed protein chain in the first step of the coarse-graining method described in section 2.2.1.

Conversely, when the average block size S in the moving average method is relatively large (S=40 in this example), we obtain fewer data points and lose more information. However, the advantage of large S is that the distribution

of amino acids is more clear to indentify, and it is easier to distinguish different kinds of blocks directly from the generated graphs. Therefore, choosing a suitable step length is critical in this method (in this thesis, we chose S=40 as the moving step for screening protein fragments).

Next, we analyse the validity of the moving average method. In Figure 2.7a above(with S=40), the hydrophobicity is substantially higher than 0.5 in all areas near the N terminus end of the protein, i.e. from step 1 to step 25, which means that α_{s1}-casein is very hydrophobic near its N terminus side. In contrast, from step 50 to step 90, the ratio of hydrophobic amino acids corresponding to α_{s1}-casein is 0.33-0.38. It can be seen that there is a continuous, quite hydrophilic region on the backbone of α_{s1}-casein in this interval. After this point, i.e. nearer the C-terminus, the ratio of hydrophobic amino acids is close to 0.5. The results of the moving average method suggest that α_{s1}-casein can be considered a protein with a hydrophobic-hydrophilic-neutral structure. This conclusion is consistent with the results of step 6 of the coarse-graining method in Figure 2.5a. Similarly, the analysis of the moving average method for the β-casein structure coincides with the coarse-graining method. As can be seen in Figure 2.7b (with S=40), there is an apparent hydrophilic fragment of β-casein at [1-20] (near the N-terminus). The proportion of hydrophobic amino acids then rises rapidly at [20-54] and remains essentially constant (>0.5) until the end of the moving average method. Thus, the structure of β-casein can be seen as hydrophilic-neutral (tending to be hydrophobic- hydrophobic), showing an overall di-block-like structure. In step 6 of Figure 2.5b, the structure of the transformed β-casein is similarly hydrophilic-neutral-hydrophobic-hydrophobic. In summary, the results obtained by the moving average method as applied to both α_{s1} and β-casein agree with the coarse-graining method while also being roughly similar to the known data for the behaviour of these proteins in experiments (Dickinson and Krishna, 2001; Horne, 2002).

2.2.3 Discussion of coarse-graining and moving average methods

In practice, because coarse-graining and moving average methods were designed by us not to involve any complex operations such as modelling and solving multivariate equations, they produce results very quickly. This meets the need for rapid screening of suitable proteins. In terms of their application, the two methods are slightly different. The moving average method allows

an investigation of the complete protein and produces graphs that help identify one or more potentially suitable fragments by visual inspection of these graphs. The coarse graining method, on the other hand, provides a quick analysis of the composition of a block of sequences and directly determines whether the fragment is the protein sequence that possesses a particular primary structure that we require. Therefore, in the order of using the fast screening method, we generally find some candidate sequences by moving average method and then further verify them by coarse-graining method.

The fast screening methods we designed do have their limitations. Firstly, as mentioned at the beginning of section 2.1.1, rapid screening methods focus only on the distribution of hydrophobic and hydrophilic amino acids in the protein chain and have no way of addressing the effect of the charge of the amino acids on the emulsification properties of the protein. Furthermore, searching for suitable peptide fragments, though made much easier using the average moving technique, still has to be done by visual observation of the generated graphs and therefore is somewhat subjective, especially in identifying precisely the selected protein fragments' starting and ending points for the optimum result. In summary, the essence of both methods is to sacrifice accuracy for speed within reasonable limits. Therefore, the SCF calculation method is needed at the next stage of fragment selection to achieve a more accurate analysis of the selected polypeptides in terms of their emulsification performance. The SCF theory is discussed next in the following section.

2.3 Self-consistent field (SCF) calculation

2.3.1 An introduction to SCF calculations

This study applied the self-consistent field (SCF) calculation scheme to estimate various interfacial properties of various hydrolysed proteins at hydrophobic-hydrophilic interfaces. The SCF calculations were first introduced by Dolan and Edwards (Dolan and Edwards, 1975) to the field of polymers by calculating the interactions induced between two adjacent surfaces by the overlap of adsorbed polymer layers covering them. Later Scheutjens and Fleer (Scheutjens and Fleer, 1979) invented a new, more efficient scheme for implementing such calculations. They could identify

sections of the chains divided into tails, loops and trains for the first time to explore the behaviour of polymers on the interfaces more accurately. They also determined free energy changes resulting from the overlap of such polymer layers, which in turn can be related to the emulsifying and colloid stabilising capacity of macromolecules being studied. Since then, the SCF method has been used in many fields involving polymers at interfaces, including in food colloid research, for example, to simulate the surface coverage and density profile of α_{s1}-casein and β-casein at different pH and ionic strengths (Atkinson et al., 1996; Leermakers et al., 1996; Pinfield et al., 1997). More recently, the SCF method has been applied to a wide range of food colloids, such as protein-polysaccharide (Ettelaie et al., 2008), modified starch (Mu et al., 2021) and fragmented protein (Ettelaie et al., 2014).

The SCF calculations used in the present study are based on the general framework of the so-called Scheutjens-Fleer scheme (Scheutjens and Fleer, 1979). This scheme is implemented in Fortran programming language, which takes about 2-15 minutes to analyse a complete data set. This length of time is not suitable for analysing a very large number of different fragments, which is precisely why the fast screening method is desired before SCF is applied to fragments identified as being potentially more likely to be useful as emulsifiers. In this PhD research, the same scheme was also implemented using the Python language, but due to the interpreted nature of Python language, it runs much slower than the Fortran implemented version. Therefore, the Python version is only presented here as a more readable source code in Appendix A but otherwise was not applied in our study.

2.3.2 One monomer systems

The SCF method is generally used for more complex multi-component systems, so before we describe the traditional SCF calculation in detail, it is helpful to take a brief look at a system with only one monomer. Let us first consider a simple question: Suppose that there is only one type of specie in the bulk solution. This specie is made from only a single monomer with adsorption energy $-\varepsilon$ to a flat interface. The bulk concentration of the monomer in the system is C_{bulk}. The question is, “what is the adsorbed amount, or the number of molecules per unit area n, of this monomer on the surface, at equilibrium with a bulk solution?” To address this question, we first need to introduce the commonly used idea of the lattice model in the SCF approach, where the area of an interface (e.g. droplet surface) is

considered as consisting of a finite number of regularly organized grid points on a lattice (in this work, we consider the cubic lattice).

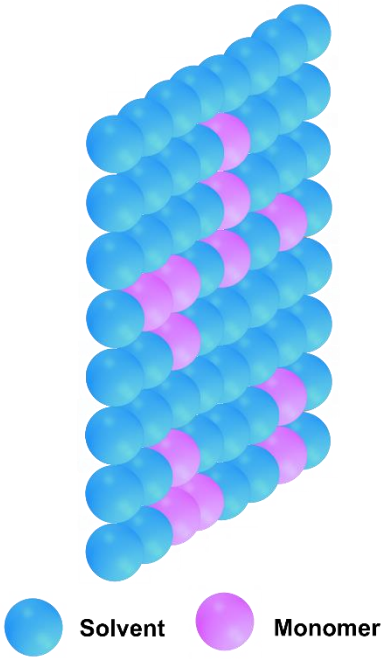


Figure 2.8a A simple lattice model of the surface in the SCF method showing grid sites occupied by either solvent or the monomer

Let us take the number of available lattice grid points available on the surface as N, where obviously N>n. At equilibrium, the number of monomers adsorbed on the surface is n (converted to n_{mol} moles). Thus, the equation for combinatorial arrangements of these n monomers on N lattice sites gives the number of microstates of the system Ω as:

$$\Omega = C_n^N = \frac{N!}{n!(N-n)!}$$

By definition, the entropy in this state is

$$S = k_B \ln \Omega = k_B \ln C_n^N$$

where k_B is Boltzmann's constant.

Substituting entropy into the Helmholtz free energy formula gives:

$$F_{surface} = U - TS = U_{solvent} + n \cdot (-\varepsilon) - Tk_B \ln C_n^N$$

The Helmholtz free energy can also be calculated as the function of chemical potential μ:

$$F_{surface} = n_{\mathrm{mol}} \cdot \mu$$

Assuming a sufficiently dilute solution, according to Henry's Law (Henry, 1803), chemical potential can be expressed as a function of bulk concentration $C_{bulk,}$ so The Helmholtz free energy of the surface is:

$$F_{surface} = n_{mol} \cdot \mu \ = n_{mol} \cdot \left[\mu^* + k_B T \ln C_{bulk} \right]$$

μ^* is also a constant which represents the chemical potential of the pure monomer, the magnitude of which is only related to the nature of the monomeric substance itself.

Now we start to calculate the free energy of the bulk part of the system. From the First law of thermodynamics, we have:

$$F_{bulk} = F_{\text{non-adsort}} - F_{surface}$$

Again, because of the first law of thermodynamics, the system's total energy F_{Total} is the same as the total energy

$$F_{Total} = F_{surface} + F_{bulk} = \left[U_{solvent} + n \cdot (-\varepsilon) - Tk_B \ln C_n^N \right] + \left(F_{non-absort} - n_{mol} \cdot \left[\mu_b^* + k_B T \ln(x_b) \right] \right)$$

When the system reaches equilibrium, the free energy of the system must be at a minimum, so there will be:

$$\frac{\partial F_{Total}}{\partial n} = 0$$

Moreover, because it is difficult to calculate the derivative of a logarithmic function containing factorials, we can use the approximation Stirling formula (Dutka, 1991) to simplify the operation.

When N is very large, It is possible to write:

$$\ln C_N^n \approx N \ln N - n \ln n - \left(N - n\right) \ln \left(N - n\right)$$

So when $\frac{\partial F_{Total}}{\partial n} = 0$, we have:

$$\frac{n}{N-n} = e^{-\varepsilon + \mu^*} C_{bulk}$$

where $e^{-\varepsilon+\mu^*}$ is known as the Henry's constant K

Take the coverage of adorted monomers $\Gamma = \frac{n}{N}$; we can finally obtain a concise expression:

$$\Gamma = \frac{KC_{bulk}}{1 + KC_{bulk}}$$

2.3.3 The classical SCF method

2.3.3.1 Introduction of the concept of mean-field ψ

In the previous section, in a system with only one monomer, we obtained the monomer concentration at equilibrium as a function of the concentration of the monomer in bulk by inter-operating between the free energies. However, as the number of monomers increases, the monomer itself receives various forces (i.e. monomer-monomer interactions, hydrophobic interactions, and charge effects). At this point, the entropy part of the free energy equation becomes considerably more difficult to calculate. To overcome this difficulty, one can consider a none-interacting (no interaction between monomers) system where the interaction between a monomer and its surrounding neighbors is represented by an external field that, in turn, has to be

determined to accurately represent the internal interactions. We note that the statistical mechanic treatment of a set of non-interacting molecules in external fields is far easier to accomplish than that of species that are interacting with each other. Therefore, in traditional SCF calculations, we introduce the concept of field ψ (here, this will be taken as given in units of $k_B T$ where temperature T = 298 K in our study). As stated above, this so-called field can be understood as the sum of all the effects from interactions due to its neighbours on a given monomer. It can also be thought of as putting the monomers in a system without any interactions and subjecting it to this field alone. A field is the function of concentration and internal energy (see equations 2.3 and 2.4). For “non-interaction systems” with a single monomer, this will be

$$\mathrm{E}_\alpha = \phi_\alpha \psi_\alpha \tag{2.3}$$

where E_α is the internal energy of the monomer of type α in the non-interacting system, ϕ_α is the volume fraction of the monomer α and ψ_α is the field of monomer α. The concentration of the monomeric species at any point close to the interface is governed by the Boltzmann equation, namely

$$\phi_\alpha = \Phi_\alpha e^{-\psi_\alpha} \tag{2.4}$$

where Φ_α is the bulk volume fraction of monomer type α in bulk. Equation 2.4 is also related to the segment density function, which will be discussed in subsequent sections.

Based on the field, this concept allows us to recalculate the question from the previous section. We may assume that there are two types of monomers in the system, monomer 1, which can adsorb on the surface with adsorption energy $-\varepsilon$, and monomer 2, which represents the solvent with zero adsorption energy. Assume that the volume fraction of monomers 1 and 2 in surface are ϕ_1 and ϕ_2; while in bulk are ϕ_{10} and ϕ_{20}. Then according to the formula for Helmholtz free energy, for a non-interaction system:

$$-TS = F_{non} - E_{non} = \left[\left(\phi_1 - \phi_{10}\right) - \left(\phi_2 - \phi_{20}\right)\right] - \left(\phi_1\varphi_1 + \phi_2\varphi_2\right)$$

From the conditions, we can see that $\phi_1+\phi_2=1$, so the first term of the above formula would be 0. From Equation 2.2, we have $\psi_1 = -\ln\frac{\phi_1}{\phi_{10}}$ and $\psi_1 = -\ln\frac{\phi_2}{\phi_{20}}$, so

$$-TS = F_{non} - E_{non} = -\ln\frac{\phi_1}{\phi_{10}} - \ln\frac{\phi_2}{\phi_{20}}$$

From the calculation of entropy in the previous section, it is evident that entropy is only related to the number of microstates of matter according to the definition of entrophy ($\mathrm{S} = \ln(\Omega)$). The number of microstates, in turn, is only related to the concentration of the substance in the system. Therefore, the entropy in an actual interacting system (i.e. one where the monomers interact with each other) remains the same as in the non-interaction system. Using this fact allows one to determine the entropy term for the interacting system using this result for the none-interacting system. Now according to the same formula for Helmholtz free energy, by adding the interaction energy term to the entropic term for the system with interaction:

$$\mathrm{F}_{\text{interaction}} = E_{\text{interaction}} - TS = -\varepsilon\phi_1 - TS = -\varepsilon\phi_1 + \ln\frac{\phi_1}{\phi_{10}} + \ln\frac{\phi_2}{\phi_{20}}$$

Similarly, at equilibrium, $\mathrm{F}_{\text{interaction}}$ reaches its minimum value. Since this is a question of conditional extremes, the minimisation has to be conducted with due consideration for the fact that site occupancy (by monomer or solvent) at any grid point is always 1. In applying this, we assume that our solution is non-compressible. To carry out the constraint minimisation, we construct a Lagrangian function L with yet to be determined variable λ using the Lagrangian multiplier method (Hoffmann et al., 1989; Beavis and Dobbs, 1990):

$$\mathrm{L} = \mathrm{F}_{\text{interaction}} + \lambda\left(\phi_1 + \phi_2 - 1\right)$$

Then we can solve the system of equations:

$$\begin{cases} \dfrac{\partial L}{\partial \phi_1} = 0 \\ \dfrac{\partial L}{\partial \phi_2} = 0 \qquad \rightarrow \Gamma = \dfrac{K\phi_{10}}{(1-\phi_{10})+K\phi_{10}} \\ \phi_1 + \phi_2 - 1 = 0 \end{cases}$$

As $\phi_{10} \approx 0, \Gamma \approx \dfrac{K\phi_{10}}{1+K\phi_{10}}$, which agrees with the result in previous section.

So far, we have figured out the concentration of different monomers in a simple system at equilibrium using the extra parameter field. In the following sections, we will continue our discussion of the situation in multi-component systems.

2.3.3.2 Components of the mean-field

In the calculations for this study, we need to calculate the more complex case of two surfaces covered with proteins, in close proximity to each other where the protein layers begin to overlap. An extended lattice model in Figure 2.8b continues to be invoked and described in detail.

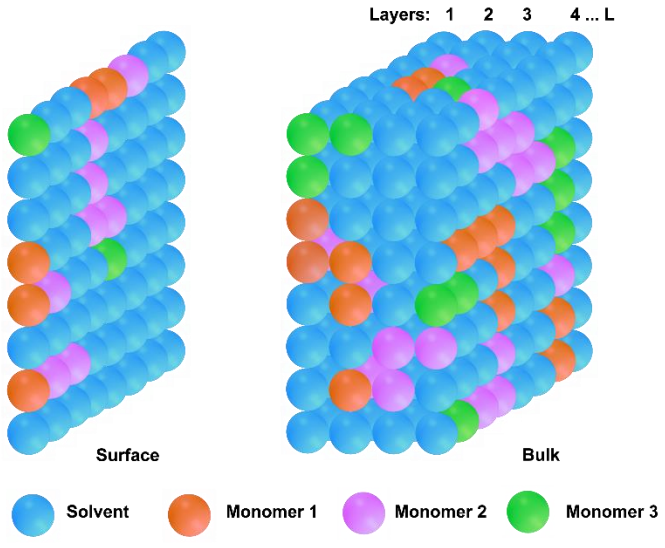


Figure 2.8b A more common lattice model of the surface and bulk in the SCF method showing grid sites occupied by either solvent or monomers

The gap between the two adjacent parallel surfaces is divided into a set of lattice sites as before, with each site occupied by one and only one of eight possible monomers: i.e. five kinds of different possible amino acids (see Table 1 for more detail), two types of ions or solvent molecules. The division of amino acid types into five groups is based on their broad characteristics of these. For example, we group all hydrophobic amino-acid into group 1. Polar but non-charged amino acids are placed in group 2, etc. The distance between two opposite interfaces, L, is considered as consisting of equally separated layers parallel to the surfaces, each with a thickness equal to the size of a grid point, a_0 (the nominal value of a_0 is taken to be the size of a peptide bond ~ 0.3 nm here). We can vary the distance between the surfaces from 2 to a relatively larger number (typically 120 layers (z = 120) in the present study), thus giving a maximum separation distance of ~ 36 nm between the two surfaces. This is considered sufficient in most cases for the two surfaces to be far enough to be considered isolated from each other. All internal interactions between different kinds of monomers, monomers and the solvent and those with ions are expressed by the potential of mean forces in SCF calculations. This potential is calculated at each layer and for each type of monomer (or solvent or ions). The mean potentials, in turn, depend on the variation of the concentration profiles of various monomers $\{\phi_i^a(z)\}$, with $\phi_i^a(z)$ representing the volume fraction of monomers of kind α belonging to molecules (chains) of type i, residing in layer z. Normalised in units of k_BT, the mean potential acting on monomers of type α has three components as given by the equation below:

$$\psi^{\alpha}(z) = \psi_{hc}(z) + q_{\alpha}\psi_{el}(z) + \psi_{\text{int}}^{\alpha}(z) \tag{2.5}$$

First of these $\psi_{hc}(z)$ is a hard-core potential term acting equally on all monomers in a given layer, irrespective of their kind. This is the interaction arising from the crowding of different monomers in the same layer, which enforces and ensures the incompressibility of the system.

$$\sum_{\text{i}} \sum_{\alpha} \phi_i^{\alpha} = 1 \tag{2.6}$$

The term $q_\alpha \psi_{el}(z)$ is a long-ranged electrostatic interaction between charged species, only present if the monomers of type α possess an electrical charge. Its value is proportional to the charge of the monomer of type α, i.e. q_α , and the electric potential at layer z, i.e. $\psi_{el}(z)$. The latter is itself determined by the solution to Poisson's equation (Ettelaie et al., 2008)

$$\varepsilon_r \varepsilon_0 \nabla^2 \psi_{el}(z) = -\sum_\alpha q_\alpha \phi_\alpha(z) \tag{2.7}$$

In which vacuum permittivity $\varepsilon_0 = 8.85 \times 10^{-12}\, F/m$ and the dimensionless parameter $\varepsilon_r = 78.5$ is the relative permittivity of water. ∇^2 is the Laplace Operator, and due to the symmetry of the problem here, we can consider that $\nabla^2 \psi_{el}(z)$ it is the second-order derivative $\psi_{el}(z)$ along a direction perpendicular to the surfaces.

The final component in Equation 2.5 above, i.e. $\psi^a_{int}(z)$, is a short-ranged contribution arising from the interaction of a monomer with other neighbouring monomers around it. Its strength is specified using the so-called Flory-Huggins interaction parameter $\chi_{\alpha\beta}$ between two dissimilar monomer types α and β. The various values of $\chi_{\alpha\beta}$ used in our calculations are shown in Table 2.2 (Ettelaie et al., 2008).

Table 2.2 Flory-Huggins interaction parameters between different kinds of monomers used in our calculations

Monomer type	0	1	2	3	4	5	6	ion+	ion-
0 - solvent	0	1	0	0	0	0	0	-1	-1
1 - hydrophobic residues	1	0	2.0	2.5	2.5	2.5	2.5	2.5	2.5
2 - polar residues	0	2.0	0	0	0	0	0	0	0
3 - positive residues	0	2.5	0	0	0	0	0	0	0
4 - histidine (His)	0	2.5	0	0	0	0	0	0	0
5 - negative residues	0	2.5	0	0	0	0	0	0	0
6 - positive ions	-1	2.5	0	0	0	0	0	0	0

7 - negative ions	-1	2.5	0	0	0	0	0	0	0
s - surface	0	-2.0	0	0	0	0	0	0	0

$\psi_{int}^{a}(z)$ can be calculated as:

$$\psi_{int}^{a}(z) = \sum_{\alpha,\beta} \chi_{\alpha\beta}\left(\left\langle \phi_{\beta}(z)\right\rangle - \Phi_{\beta}\right) + \left(\delta_{z,1} + \delta_{z,L}\right)\chi_{\alpha s} \qquad (2.8)$$

And therefore depends on the variation of density profiles of various types of monomers (as well as ions and the solvent) in the small gap between the two surfaces. The quantity $\left\langle \phi_{\beta}(z)\right\rangle$ represents the average volume fraction of the neighbouring monomers of kind β surrounding a monomer placed in layer z, whereas Φ_{β} is the corresponding value in the bulk solution. The values $\left\langle \phi_{\beta}(z)\right\rangle$ are obtained for each layer through a weighted average of $\phi_{\beta}(z)$ that and its two neighbouring layers. This is calculated as follows

$$\left\langle \phi^{\beta}(z)\right\rangle = \lambda_{-1}\phi_{\beta}(z-1) + \lambda_{0}\phi_{\beta}(z) + \lambda_{+1}\phi_{\beta}(z+1) \qquad (2.9)$$

with the weight factors λ_{-1} = λ_{-1} = 1/6 and λ_{0} = 4/6 for the cubic lattice adopted in our calculations here. The weight factors reflect the number of adjacent sites in each neighbouring layer to any given grid point. The monomers are residing in layers 1 and L also in contact and therefore interact with one or the other of the two surfaces. The last term in Equation 2.7 represents the interaction field between monomers and surfaces for these two layers. The function $\delta_{a,b}$ represents the Kronecker delta function which mathematically is defined so that $\left(\delta_{z,1} + \delta_{z,L}\right)$=1 when z=1 or L, otherwise $\left(\delta_{z,1} + \delta_{z,L}\right)$=0.

2.3.3.3 Segment density functions $G_i(n,z)$

The Semeny density function can be considered as a bridge connecting the values of the volume fractions and the fields. Within a constant normalisation factor, The segment density function $G_i(n,z)$ represents the probability of the

n^{th} monomer of the polymer chain i to be found at layer z. For the case of a single monomer α, it is just the usual Boltzmann factor and is expressed as:

$$G(1,z) = e^{-\psi_\alpha(z)} \tag{2.10}$$

In such a simple case, then, the relation between the volume fraction and the field is given by:

$$\phi_\alpha(z) = \Phi_\alpha G(1,z) \tag{2.11}$$

It can be seen here that Equation 2.11 is the same as Equation 2.4 in section 2.3.3.1. It is intuitively clear from Equations 2.10 and 2.11 that $G_i(n,z)$ is essentially a probability function constructed from the values of fields. Under the action of this probability, the bulk volume fraction of a single monomer is transformed into the corresponding volume fraction under the action of the specified field. However, the above discussion only addresses the case of a single monomer, whereas our study focuses on more complex protein chains made from chains of connected monomers. In order to calculate the value of the segment density function when n>1, we need to return to the properties of the polymer itself, representing this as a Markov chain. In a Markov chain, future behaviour is determined by current values only, independent of past behaviour. In a similar fashion, for polymers that do not interact with each other, the position of a monomer on the backbone of a polymer is related to the position of the previous monomer only. By a similar method of calculating the joint probability, we have:

$$G(2,z) = G(1,z) \times G_{\text{joint}} \tag{2.12}$$

As shown in Figure 2.8b, each monomer of the lattice model applied in this study, residing in the z layer, will have four possible directions connecting it to the adjacent monomers also sitting in the z layer and one possible direction each connecting it to adjacent monomers positioned in the layers z-1 and z+1. Therefore, the value G_{joint} in Equation 2.11 can be calculated as follows:

$$G_{\text{joint}} = \lambda_{-1} G(1, z-1) + \lambda_0 G(1, z) + \lambda_{+1} G(1, z+1)$$

(2.13)

where weight factors $\lambda_{-1} = \lambda_{+1} = 1/6$ and $\lambda_0 = 4/6$ for the cubic lattice adopted in our calculations here. The weight factors reflect the number of adjacent sites in each neighbouring layer to any given grid point. Combining Equations 2.12 and 2.13 leads to the more general equation for $G_i(n, z)$:

$$G_i(n, z) = e^{-\psi_{Tn}(z)} \cdot \left[\lambda_{-1} G(n-1, z-1) + \lambda_0 G(n-1, z) + \lambda_{+1} G(n-1, z+1) \right]$$

(2.14)

where $\psi_{Tn}(z)$ is the field in the *z* layer acting on monomers of type group *T* to which the nth monomer on polymer chain i belongs; We also impose the conditions $G(n,0)$ $G(n,\mathrm{L}) = 0$ which signify that the polymer cannot penetrate the interface to enter the non-aqueous phase.

By recursive application of Equation 2.14, starting from a set $\{G(1,z)\}$ given by (2.10), we can calculate the complete set of $\{G(n,z)\}$ values for all n=1 to N_i (where N_i is the degree of the polymerisation of the polymer). Once the segment density values for all monomers on a polymer chain have been calculated, $G_\alpha(z)$ for the monomer with a specific type α can be further calculated by:

$$G_\alpha(z) = \sum_{n=1}^{Ni} G_{Tn=\alpha}(n, z)$$

(2.15)

where $G_{Tn=\alpha}(n,z)$ is the segment density of the nth monomer belonging to type α on layer z. If the nth monomer happens not to be of type α, then, of course, the value of $G_{Tn=\alpha}(n,z)$ =0.

The volume fraction of monomers of type α residing on layer z can also be calculated in a similar way once $\phi_{Tn=\alpha}(z)$ has been calculated:

$$\phi_{\alpha}(z) = \sum_{n=1}^{Ni} \phi_{Tn=\alpha}(z) \tag{2.16}$$

Here $\phi_{Tn=\alpha}(z)$ is the volume fraction of the nth monomer along the polymer backbone, if it happens to belong to type group α, positioned on layer z.

The general formula for calculating the volume fraction of any type of monomer in any layer can be obtained by combining Equations 2.10 - 2.16. This is known as the composition law (Ettelaie et al., 2016):

$$\phi_{\mathrm{i}}^{\alpha}(z) = \frac{\Phi_i}{N_i} \sum_{n=1}^{Ni} \frac{G_{Tn=\alpha}^{f}(n,z) G_{Tn=\alpha}^{b}(N_i - n + 1, z)}{e^{-\psi_{Tn}(z)}} \tag{2.17}$$

where Φ_i is the bulk concentration of the polymer I;N_i is the length of of polymer *i*

Since polymers (and in particular proteins) are generally not symmetric if viewed from C- and N- terminus, we need to calculate $G_{\alpha}^{f}(z)$ and $G_{\alpha}^{b}(z)$ separately using the two ends of the polymer (i.e. the 1st monomer and the N$_i$th monomer) as starting points and subsequently eliminate a redundant $e^{-\psi_{Tn}(z)}$ according to the definition of $G_i(n,z)$ in Equation 2.14.

2.3.3.4 Creation of a system of equations

In the discussion in sections 2.3.3.1 to 2.3.3.3 above, there is an interconnection between values of the volume fraction and those of the field. That is to say, if volume fractions are known, fields can be determined, and if fields are known, the volume fraction can be computed. However, one does not know the value of either set of quantities. The segment density function calculation makes it possible to calculate the volume fractions from fields, and therefore here we take these (i.e. $\{\psi^{\alpha,\beta\cdots}(z)\}$) as the independent set of variables in the mathematical treatment of the problem. Together with $\{q_{\alpha,\beta\cdots}\psi_{el}(z)\}$, $\{\psi^{\alpha,\beta\cdots}(z)\}$ in Equation 2.7 needs a set of a system of simultaneous (which turns out to be highly non-linear) equations involving

these unknowns. There are altogether a total of LW + L unknowns (L is the number of layers, and W is the number of different groups of monomers from which the polymer chains are made).

Returning to Equation 2.5, the hard-core potential term may be expressed as follows:

$$\psi_{hc}(z) = \psi^{\alpha}(z) - q_{\alpha}\psi_{el}(z) - \psi^{\alpha}_{\text{int}}(z) \tag{2.18}$$

Now recall that hard core potential acts equally on any monomer in a given layer, irrespective of its kind α. It is obvious from (2.18) that layer number z is the only factor that influences the value of the hard-core potential. All of the system's monomers have the same hard-core potential acting on them when the system is in equilibrium. As a result, we can derive the following set of (LW-L) simultaneous equations:

$$\begin{aligned}\psi^{\alpha}(z) - q_{\alpha}\psi_{el}(z) - \psi^{\alpha}_{\text{int}}(z) = \psi^{\beta}(z) - q_{\beta}\psi_{el}(z) - \psi^{\beta}_{\text{int}}(z) = \ldots \\ = \psi^{0}(z) - q_{0}\psi_{el}(z) - \psi^{0}_{\text{int}}(z)\end{aligned} \tag{2.19}$$

where the superscript 0 here stands for solvent.

Now, we need a further 2L equations to satisfy the need for having the same number of equations as unknown variables. We may derive extra L equations based on the incompressibility within each layer in the lattice model (see Equation 2.6). It is more convenient for the numerical solution to express Equation 2.6 in its logarithmic form, i.e.

$$\log\left(\sum_{\text{i}}\sum_{\alpha}\phi_i^{\alpha}\right) = 0 \tag{2.20}$$

Finally, the Poisson equation for the electrostatic potential can be used to generate the final set of L equations (see Equation 2.7). On a lattice model,

a discretisation of the Poisson differential equation for $\psi_{el}(z)$ on the x-axis, with dz=a_0, is used in the SCF calculation. So, using the discretised form of the derivatives, the Laplace operator in Equation 2.7 can be expanded as follows:

$$\varepsilon_r \varepsilon_0 \cdot \frac{\dfrac{\psi_{el}(z+1)-\psi_{el}(z)}{a_0} - \dfrac{\psi_{el}(z)-\psi_{el}(z+1)}{a_0}}{a_0} = -\sum_{\alpha} q_\alpha \phi_\alpha(z)$$

Leading to

$$\frac{2\psi_{el}(z)-\psi_{el}(z-1)-\psi_{el}(z+1)}{{a_0}^2} - \sum_{\alpha} q_\alpha \phi_\alpha(z) / \varepsilon_r \varepsilon_0 = 0 \qquad (2.21)$$

Simultaneous set of non-linear equations 2.19, 2.20 and 2.21 can be solved together to find the values of the fields corresponding to every monomer type in each layer *z*. Consequently, the volume fraction of all monomers in different layers can be calculated by substituting these fields into the equations related to the segment density function, as described in the previous section. It must be noted that it is a challenging and time-consuming process to attempt to solve a system of equations with so many unknowns. For example, in a typical case with 100 layers involving six types of monomers, we will have 700 variables. In this work, we employed the hybrid Powell approach (so-called Powell's dog leg method) to resolve equational problems. Fortunately, this method has been implemented in the publicly available Fortran numerical package Minipack (Moré et al., 1980). The traditional Powell's approach (Powell, 1964) is not used to solve systems of multivariate equations; rather, it is a method for finding local minima of functions (see Figure 9). The hybrid Powell's algorithm sums up all the left-hand sides of the equation system and then squares them to create a new function (assuming all the right-hand parts are zero). Using the traditional Powell approach, the minimum of this function is then determined, and the x-coordinate corresponding to the minimum is the solution to the system of equations.

It is crucial to understand that the hybrid Powell technique does not have a fixed proportion of success when it comes to solving equation systems. It greatly depends on the initial guess starting values when used as a search algorithm for finding a minimum. The algorithm starts from the initial values and searches in the calculated direction. The convergence condition $\left\|x^k - x_0^k\right\| \le \varepsilon$ (in this work, the convergence accuracy $\varepsilon = 10^{-7}$) is verified at the end of each search. In our implementation of SCF, we use what we call a bini number to generate the initial values. By changing the bini number, we can change the starting guess solution.

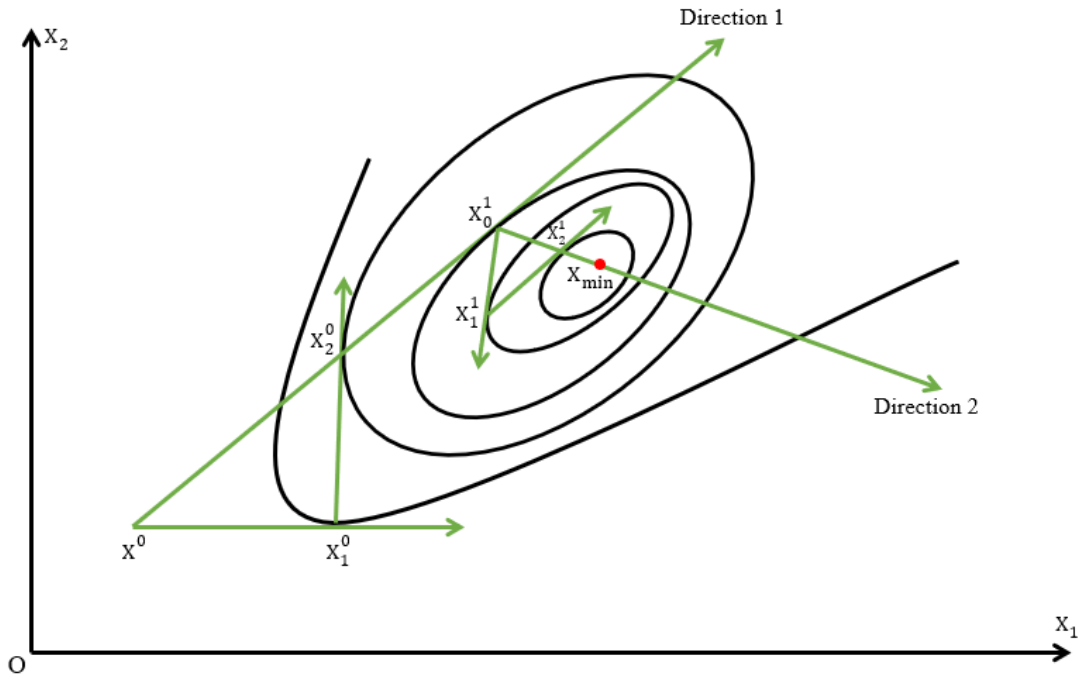


Figure 2.9 The search process of the traditional Powell's algorithm attempts to find a minimum

2.3.3.5 Free energy and Deryagin approximation

The procedure for estimating the free energy of complicated multi-component systems remains similar to that in section 2.3.3.1 once the volume fraction and field values are determined. Firstly, the value of -TS is ascertained using the formula based on the observation that the entropy in the interaction and non-interaction systems are equal because the number of microstates is equal in both systems. The following equation for the free energy (expressed in the unit of k_BT/a_0^2) by adding to the entropy term the total internal energies in the interacting system.

$$\mathrm{F}(L)=\left[-\sum_i \frac{1}{N_i}\sum_{z=1}^{L}\sum_{\alpha}\left(\phi_i^{\alpha}(z)-\Phi_i^{\alpha}\right)\right]-\left[\sum_i\sum_{z=1}^{L}\sum_{\alpha}\psi_{\alpha}(z)\phi_i^{\alpha}(z)\right]+$$
$$\left[\frac{1}{2}\sum_{z=1}^{L}\sum_{\alpha,\beta}\sum_{i,j}\chi_{\alpha\beta}\left(\phi_i^{\alpha}(z)-\Phi_i^{\alpha}\right)\left(\left\langle\phi_j^{\beta}(z)\right\rangle-\Phi_j^{\beta}\right)\right]+$$
$$\left[\frac{1}{2}\sum_{z=1}^{L}\psi^{el}(z)\sum_{\alpha}\sum_{i}q_a\phi_i^{\alpha}(z)\right]+\sum_i\sum_{\alpha}\chi_{\alpha s}\left[\phi_i^{\alpha}(1)+\phi_i^{\alpha}(L)\right]$$

(2.22)

The first term of the equation represents the free energy in the non-interaction system, while the second term represents the internal energy in the non-interaction system. When subtracted from each other, this provides the value of the -TS in the non-interacting, and hence also in the interacting system. The third, fourth and fifth terms in the equation are the value of the internal energy in the interacting system. They arise from the mutual interactions between different monomers that comprise the protein chains, the electrostatic interaction of charged monomers with other charged residues and finally, the interaction between monomers belonging to chains, the solvent and ions with the surfaces. The difference *F(L)-F(∞)* between the free energy when the surfaces are a distance L apart and when they are isolated (i.e. far apart) is precisely the interaction potential induced between the two surfaces due to the presence of the adsorbed polymers onto them.

The above calculations address the free energy between two planes. However, in colloid science, we are normally concerned with particle-particle interactions. If required, it is possible to convert the results obtained for two flat surfaces to a pair of spherical particles using the well-known Deryagin approximation (O'Melia, 2006).

In Deryagin approximation, the tiny section of arc generated by the spherical surface can be roughly described as a plane when the distance r between the two neighbouring particles is close and significantly smaller than the radius R of the particles themselves (see Figure 2.10).

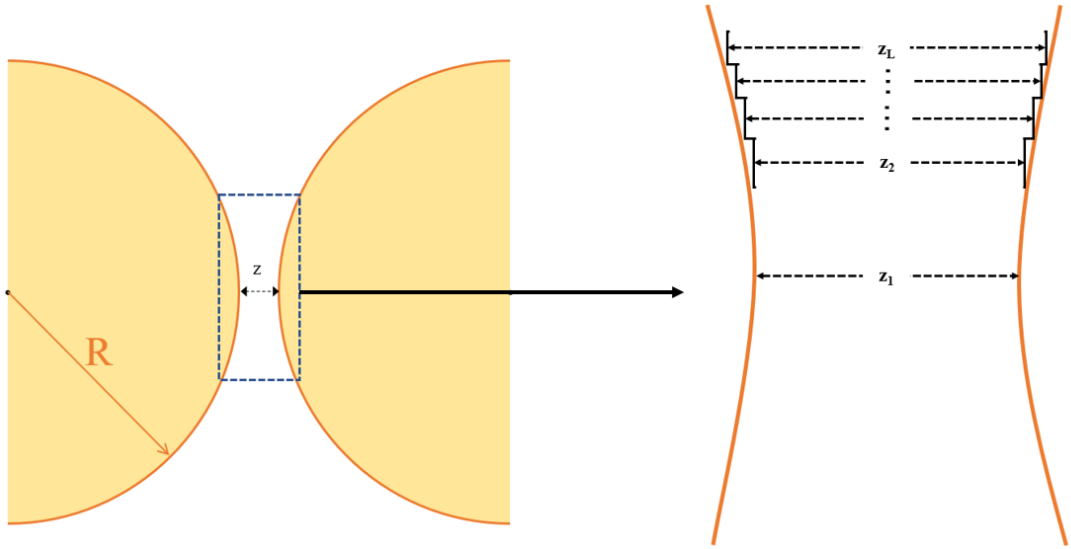


Figure 2.10 An illustration of the Deryagin approximation, which considers the surface of spheres as a collection of flat surfaces with varying surface-to-surface separation distances

The interaction force between two particles at various distances can now be estimated using the equation below derived on the basis of this approximation：

$$f_{particle} = \pi R\left[F(z) - F(\infty)\right] \tag{2.23}$$

The force $f_{particle}$ is simply the (-ve) of the differential of the required interaction potential between particles. Therefore the interaction potential is given by the definite integral with respect to the particle-particle separation distance *z*:

$$\mathrm{V}(z) = \pi R\int_{z}^{\infty}\left[F(z') - F(\infty)\right]dz' \tag{2.24}$$

In our calculations, the above integral is obtained by multiplying values $f_{particle}$ at various distances by the differential *dz* (in this work *dz* = $2a_0$). The interaction potential between the two particles in this work is expressed (in the unit of k_BT. In the above equation, R is the radius of the droplets, z is

the distance between the drops, and for the validity of the Deryagin approximation, we require that the separation distances of interest z<<R.

In order to make the computations more accurate, van der Waals forces are also taken into consideration in this work. The following equation is used to compute the van der Waals forces (Dickinson and Stainsby, 1982):

$$V_{vw} = -\frac{A_H R}{12z} \tag{2.25}$$

where A_H is the composite Hamaker constant for the dispersed phase in the continuous medium with a typical value of 1 k_BT for edible oils dispersed in water, as before, R is the radius of the two spherical droplets, and z is the separation distance between the droplets.

2.3.3.6 Calculation of charges, separation distance and amount of total adsorption

The third, fourth, and fifth classes of amino acids in Table 2.1 are the three classes of monomer residues that carry a charge, as reflected in SCF calculations. From their pKa values, the following equation is used to determine their charges:

$$q_\alpha = \frac{1}{1+10^{pKa-pH}} \tag{2.26}$$

The following table shows the pKa values from the previous work (Akinshina et al., 2008) used in this thesis:

Table 2.3 pKa values of three different types of amino acid residues

Types of monomers in Table 1	3	4	5
pKa	10	6.75	4.5

In particular, we have placed histidine in a group of its own (group 4) due to its very different pKa to other positively (at pH=7) charged groups. The average distance of the nth monomer of chain i from the surface, when this chain is adsorbed at the interface, can be calculated from the calculated density profile data as follows:

$$di_i^n = \sum_{z=1}^{L/2} \left(z \cdot \frac{\phi_i^n(z)}{\sum_z^{L/2} \phi_i^n(z)} \right) \tag{2.27}$$

where $\phi_i^n(z)$ is the volume fraction of the nth monomer of polymer type i on layer z, and is obtained from the composition law, Equation 2.17, as previously discussed:

$$\phi_i^n(z) = \frac{\Phi_i}{Ni} \cdot \frac{G_{Tn=\alpha}^f(n,z) G_{Tn=\alpha}^b(Ni-n+1,z)}{e^{-\psi_{Tn}(z)}} \tag{2.28}$$

Note that the summation in z in Equation 2.27 is only taken over half of the gap from layer 1 to layer *L*/2. This explores the fact that the density profile in the gap will be symmetric if the two interfaces (i.e. the surface of two neighbouring particles) are identical.

The amount of total adsorption (in the unit of chain/a_0^2) of the polymer can also be calculated from density profiles as follows:

$$\Pi_i = \frac{1}{Ni} \sum_{z=1}^{L/2\alpha} \sum_{\alpha} \left[\phi_i^{\alpha}(z) - \Phi_i^{\alpha} \right] \tag{2.29}$$

As seen from the equation above, a positive indicates that the polymer of type i has been adsorbed on the surface, whereas a negative Π_i indicates that the polymer has been depleted from the interfacial region.

2.3.3.7 Conclusion of the SCF calculation

To summarise, the complete SCF calculation process can be seen in Figure 2.11.

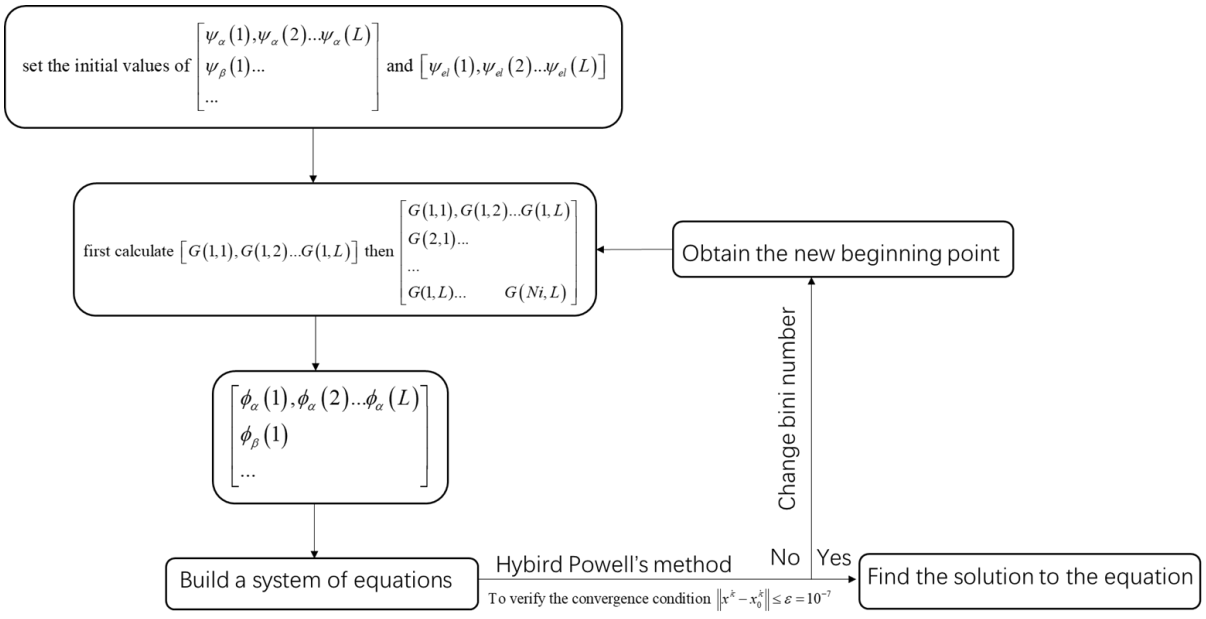


Figure 2.11 Schematic flow chart of SCF calculations as implemented in our program

In practice, for complex sequences, the use of the same initial guess starting solution does not allow the calculation to converge at all separation distances from z = 2~120 monomer sizes, as often used in this study. In such cases, we have to vary the initial bini number (used to generate the starting initial guess) to suit different distances. This inconvenience tends to increase the computational time considerably. This is why SCF calculations should be reserved for more detailed and accurate analyses of peptide sequences, reserved for those fragments that pass the quick screening by faster (but less detailed) methods described in sections 2.2.1 and 2.1.2

In this work, other methods, in addition to protein screening techniques, were also employed, such as calculating the relation between the degree of

hydrolysis and the relative molecular weight of the mixture. These will be described later in the following results in Chapter 5.

List of References

Akinshina, A., Ettelaie, R., Dickinson, E. and Smyth, G. 2008. Interactions between adsorbed layers of αS1-casein with covalently bound side chains: a self-consistent field study. Biomacromolecules. 9(11), pp.3188-3200.

Atkinson, P., Dickinson, E., Horne, D. S., Leermakers, F. and Richardson, R. 1996. Theoretical and experimental investigations of adsorbed protein structure at a fluid interface. Berichte der Bunsengesellschaft für physikalische Chemie. 100(6), pp.994-998.

Beavis, B. and Dobbs, I. 1990. Optimisation and stability theory for economic analysis. Cambridge university press.

Dickinson, E. 1997. Properties of emulsions stabilized with milk proteins: overview of some recent developments. Journal of Dairy Science. 80(10), pp.2607-2619.

Dickinson, E. 2001. Milk protein interfacial layers and the relationship to emulsion stability and rheology. Colloids and Surfaces B: Biointerfaces. 20(3), pp.197-210.

Dickinson, E. and Krishna, S. 2001. Aggregation in a concentrated model protein system: a mesoscopic simulation of β-casein self-assembly. Food hydrocolloids. 15(2), pp.107-115.

Dickinson, E., Semenova, M. G. and Antipova, A. S. 1998. Salt stability of casein emulsions. Food Hydrocolloids. 12(2), pp.227-235.

Dolan, A. and Edwards, W. F. 1975. The effect of excluded volume on polymer dispersant action. Proceedings of the Royal Society of London. A. Mathematical Physical Sciences. 343(1635), pp.427-442.

Dutka, J. 1991. The early history of the factorial function. Archive for history of exact sciences. pp.225-249.

Ettelaie, R., Akinshina, A. and Dickinson, E. 2008. Mixed protein–polysaccharide interfacial layers: a self consistent field calculation study. Faraday discussions. 139, pp.161-178.

Ettelaie, R., Holmes, M., Chen, J. and Farshchi, A. 2016. Steric stabilising properties of hydrophobically modified starch: Amylose vs. amylopectin. Food Hydrocolloids. 58, pp.364-377.

Ettelaie, R., Zengin, A. and Lee, H. 2014. Fragmented proteins as food emulsion stabilizers: A theoretical study. Biopolymers. 101(9), pp.945-958.

Euston, S. R. and Horne, D. S. 2005. Simulating the self-association of caseins. Food Hydrocolloids. 19(3), pp.379-386.

Henry, W. 1803. III. Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. Philosophical Transactions of the Royal Society of London. (93), pp.29-274.

Hoffmann, L. D., Bradley, G. L. and Rosen, K. H. 1989. Calculus for business, economics, and the social and life sciences. McGraw-Hill.

Horne, D. S. 2002. Casein structure, self-assembly and gelation. Current Opinion in Colloid Interface Science. 7(5-6), pp.456-461.

Horne, D. S. 2017. A balanced view of casein interactions. Current Opinion in Colloid Interface Science. 28, pp.74-86.

Leermakers, F. A., Atkinson, P. J., Dickinson, E. and Horne, D. S. 1996. Self-consistent-field modeling of adsorbed β-casein: effects of pH and ionic strength on surface coverage and density profile. Journal of Colloid Interface Science. 178(2), pp.681-693.

Moré, J. J., Garbow, B. S. and Hillstrom, K. E. 1980. User guide for MINPACK-1. CM-P00068642.

Mu, M., Karthik, P., Chen, J., Holmes, M. and Ettelaie, R. 2021. Effect of amylose and amylopectin content on the colloidal behaviour of emulsions stabilised by OSA-Modified starch. Food Hydrocolloids. 111, p106363.

O'Melia, C. R. 2006. Chapter 18 - Fundamentals of particle stability. In: Newcombe, G. and Dixon, D. eds. Interface Science and Technology. Elsevier, pp.317-362.

Pinfield, V., Horne, D. S. and Leermakers, F. M. 1997. Self-consistent-field modelling of adsorbed casein interaction between two protein-coated surfaces. Journal of the Chemical Society, Faraday Transactions. 93(9), pp.1785-1790.

Powell, M. J. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. The computer journal. 7(2), pp.155-162.

Scheutjens, J. M. H. M. and Fleer, G. J. 1979. Statistical theory of the adsorption of interacting chain molecules. 1. Partition function, segment density distribution, and adsorption isotherms. The Journal of Physical Chemistry. 83(12), pp.1619-1635.

Chapter 3
Application of Fast Screening Methods

3.1 Introduction

As was mentioned at the end of chapter 2, we initially collected suitable candidate protein fragments by a rapid screening procedure before conducting more detailed SCF calculations. This allows for a considerable saving in the computational time. Following the procedure in Figure 2.1, we eventually discovered and focused on a fragment of β-conglycinin-α-subunit with a di-block structure comparable to that of β-casein. This was obtained after numerous attempts to screen and identify various protein fragments. In Chapter 4, it is also demonstrated that this fragment has good emulsification qualities. This chapter will focus on the application of the quick screening procedure we devised and used to screen protein fragments, presenting the results that led to the choice of the above-mentioned protein fragment for further study in chapter 4.

3.2 Application of the moving average method

We begin by performing a briefer examination of the β-conglycinin-α-subunit protein using the method described in section 2.2, which results in the diagram shown below:

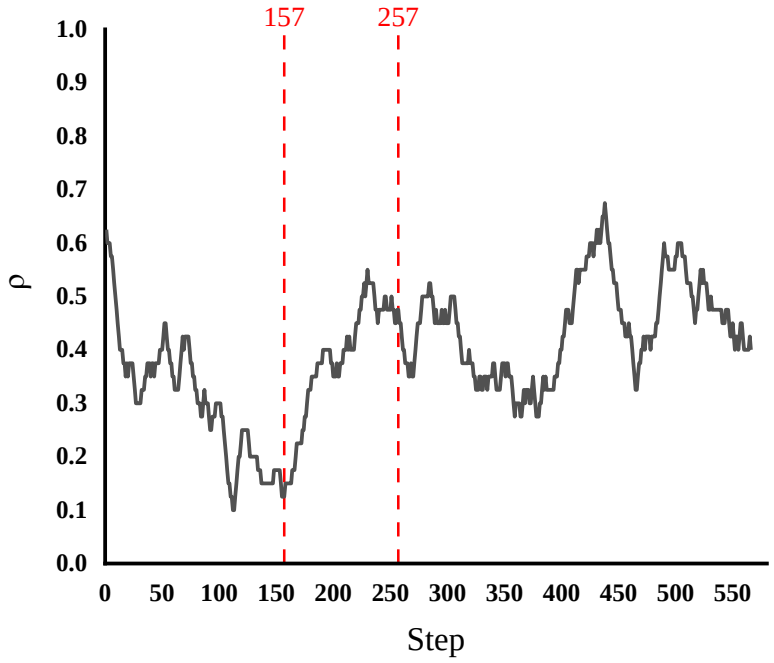
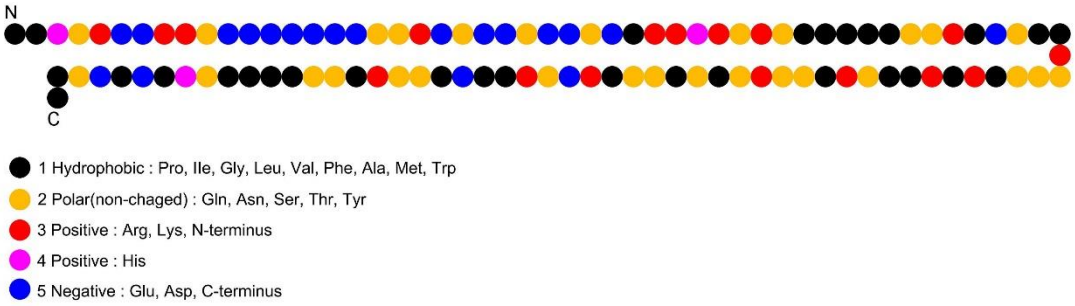


Figure 3.1 Screening of candidate proteins from the β-conglycinin-α-subunit using the moving average method (the averaged block length S=40). The x-axis represents the number of steps of the moving average method starting from the N-terminus of the protein and moving towards the C-terminus. The y-axis represents the density of the hydrophobic residues in a fragment with 40 residues

Figure 3.1 shows many peaks formed by the dense distribution of hydrophobic amino acids in a given region (as mentioned above, S = 40), indicating that, like the multi-block structure in Figure 2.2, hydrophobic and hydrophilic blocks are alternatively distributed inside the natural β-conglycinin-α-subunit. Therefore, based on our findings in Figure 2.3, proteins with this structure are also more likely to exhibit bridging flocculation, lie flatter on the surface and, thus, are less likely to create thick adsorption layers. It was evident that the intact β-conglycinin-α-subunit protein cannot exert enough steric repulsion. The question is then whether a fragment of this protein can behave in a more desired way as an emulsifier than the original intact molecule.

The protein fragment we selected here to study has an isoelectric point of 6.1, with a size of 100 amino acid residues, located between amino acid residues 157 and 257 (counting from the C-terminus). Figure 3.2 schematically illustrates its structure.



PPHQKEERKQEEDEDEEQQRESEESEDSELRRHKNKNPFLFGSNRFETLFK 1 - 51
NQYGRIRVLQRFNQRSPQLQNLRDYRILEFNSKPNTLLLPNHADADYLI 52 - 100

Figure 3.2 The schematic primary structure sequence of the di-block-like fragment of the β-conglycinin-α-subunit (The legend corresponds to the classification of the different amino acids in Table 2.1)

This fragment was chosen for two main reasons: firstly, it has fewer peaks above the moving average curve, suggesting that larger continuous trains of hydrophobic and hydrophilic blocks are present on the fragment; and secondly, the density of hydrophobic residues ρ from one end of the fragment to the other end exhibits an overall increasing trend from 0.15 to 0.48. The middle of the curve shows a slight reduction that a few distinct amino acid residues may have caused, but this does not change our overall assessment that the fragment has a close to a di-block-like structure, resembling one like β-casein. As can be seen more clearly in Figure 3.2, the hydrophilic component of this fragment is primarily located towards its N-terminus, while the hydrophobic group is primarily located near the C-terminus side (8 hydrophobic residues in 13 residues counting from the C-terminus).

To produce a more ideal di-block fragment, we could move the line indicating the endpoint in Figure 3.1 to the left to amino acid residue at position 230 or one at 195, but doing so would result in a shorter fragment, which might then alter how well such fragment will behave as an emulsifies (the relationship between fragment length and emulsification performance will be discussed in Chapter 6 in more detail). However, if we leave the fragment as it is, we end up with a few extra hydrophobic residues, which may affect the fragment's ability to emulsify (see the later discussion in Chapter 5). So overall, we would like a longer peptide with more hydrophobic groups as a prospective protein fragment in this study.

The moving average method can also identify the fragments with low potential for emulsification (see Figure 3.3).

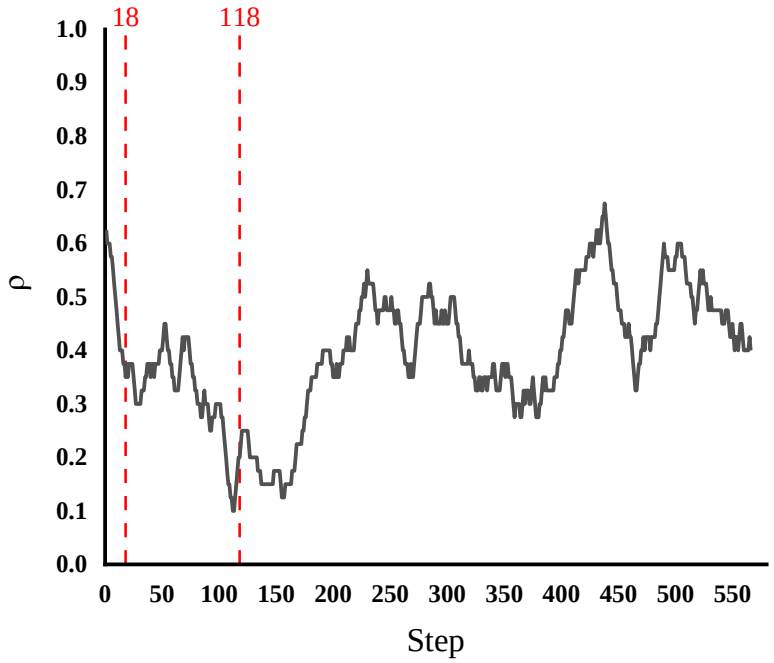


Figure 3.3 The simple moving average graph of hydrophobicity showing a fragment from amino-acid in position 18 to that in position 118, with predicted poor emulsification potential, derived from β-conglycinin-α-subunit, obtained using the moving average method (S=40)

This 100-amino-acid sized fragment, with a calculated pI of 5.3 (The calculation of pI is based on Equation 2.23, which is solved using the dichotomy method. The relevant Python source code can be found in Appendix A), was chosen between amino acids 18 and 118. It was selected because the curve in Figure 3.3, corresponding to this fragment, indicates a protein fragment with a multi-block-type structure. At least three clearly defined peaks can be identified on the SMAH curve for this fragment, pointing to its multi-block-type nature. The reason why a fragment with poor emulsification properties was also chosen here was for comparison with more favourable protein fragments in the SCF calculations, as presented in Chapter 6. Such a comparison provides a test of how well the quick-moving average method of Chapter 2 can filter out unsuitable fragments.

3.3 Application of the coarse-graining method

In the preceding section, we could use the moving average method to pinpoint a fragment with a di-block structure derived from the β-conglycinin-α subunit. This section uses the alternative coarse-graining method to analyse and highlight di-block-like fragments further. The application of the method is shown in Figure 3.4, alongside the outcomes of the coarse-graining approach for β-conglycinin.

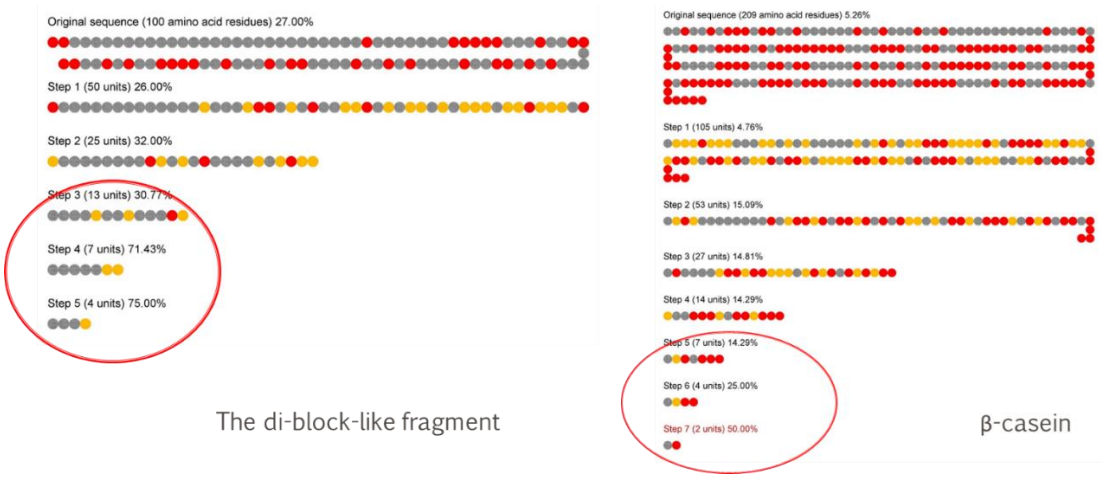


Figure 3.4 Comparison of the results obtained through the application of multiple steps of the coarse-graining method to our likely di-block-like fragment derived from β-conglycinin, with those obtained for β-casein

It can be seen that both the protein fragment from β-conglycinin and β-casein show a clear di-block-like structure after the fifth coarse-graining step. They are relatively hydrophilic near the C-terminus and appear neutral near to hydrophobic closer to their N-terminus side. The main difference with β-casein is that the hydrophobic amino acid density in the hydrophobic block near the N-terminal end of this fragment is not as high as that of β-casein. When comparing only the results of the last step of the coarse-graining method, the candidate fragment (step 5) has three-quarters of the hydrophilic coarse-grained units, while β-casein (step 7) has only one-half of the hydrophobic coarse-grained units. In contrast, when comparing the complete protein, the proportion of hydrophilic amino acids in the candidate fragment is 0.70, while the proportion of hydrophilic amino acids in the β-casein is 0.46. As a result, the fragment that resembles a di-block has a shorter hydrophilic block and a longer hydrophobic block in comparison to β-casein.

Similarly, we again applied the coarse-graining method to the fragment that was considered to have poor emulsification capacity in the moving average method (see Figure 3.5). It can be observed that this fragment is too uniformly distributed between hydrophobic and hydrophilic residues to allow di-block or tri-block to be found in the final coarse-graining results (step 5). This conclusion is consistent with that found in the moving average method.

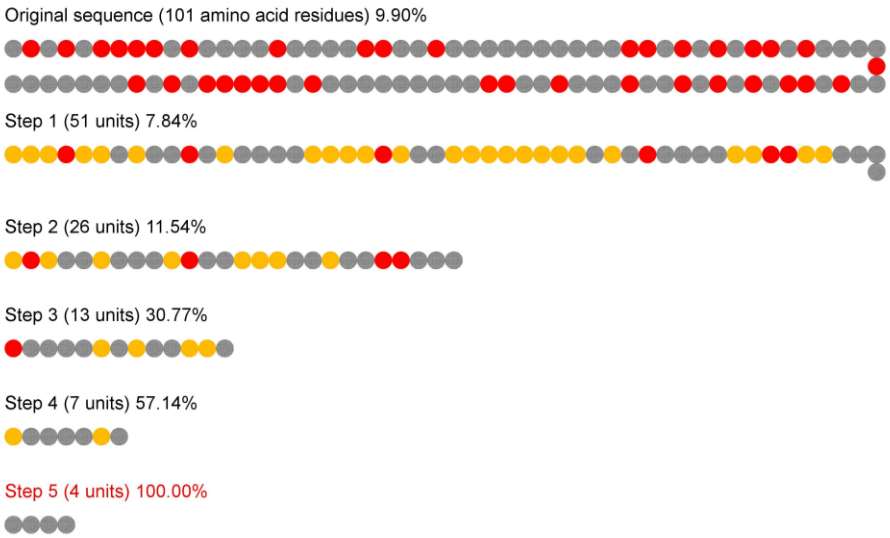


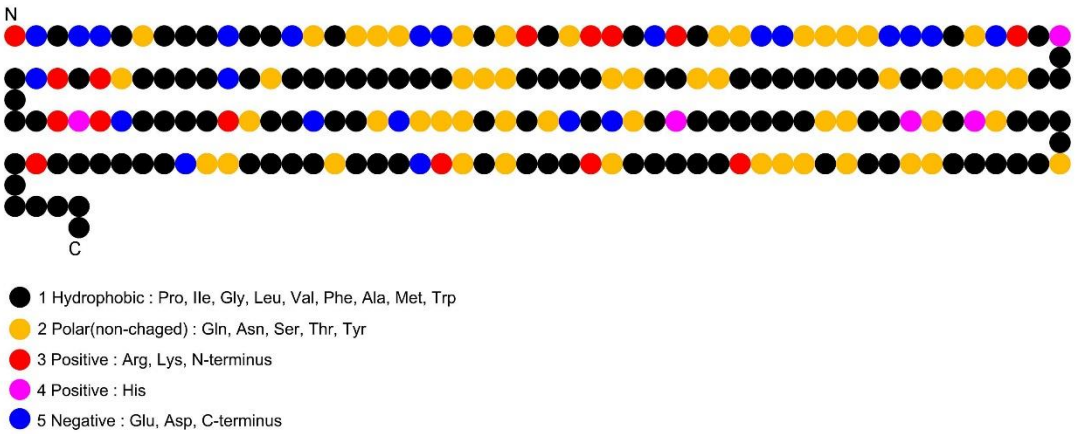
Figure 3.5 The result obtained through the application of multiple steps of the coarse-graining method to our multi-block-like fragment with predicted poor emulsification capacity derived from β-conglycinin

In the following chapter, we will continue to examine the various colloidal characteristics of this candidate fragment using SCF calculations and contrast it with other fragments.

3.4 Conclusion and discussion

In this chapter, we applied two rapid screening methods we discussed in Chapter 2 to select a fragment with an approximate di-block structure (reference) from β-conglycinin, a protein that, in its intact form, is not a naturally suitable emulsifier. This process demonstrates two significant advantages of the fast screening methods: (1) It allows for a more visual distribution of hydrophobic and hydrophilic amino acids. For example, it is difficult to determine the structure from the sequence of a protein (a segment of abbreviations of amino acids obtained from databases) without the help of fast screening methods. However, using a quick screening method such as

those we devised here, the distribution of different classes of amino acids can be transformed into more logical data to identify the presence of large, overwhelmingly hydrophobic and hydrophilic fragments. Moreover, the moving average graph (see Figure 3.2) is flexible in choosing the length of the candidate protein fragments according to practical requirements (for example, a consideration of where realistic cleavage of bonds can occur with a given enzyme).



RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHP	1 - 51
FAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAM	52 - 102
APKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPP	103 - 153
TVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGP	154 - 204
FPIIV	205 - 209

Figure 3.6 The primary structure of β-casein with the constituent amino acids grouped into five different groups as indicated

Secondly, the screening method introduced is much faster than the detailed SCF calculations, as employed in the following chapter. The entire procedure takes less than a second, whether it is creating a curve for the simple moving average of hydrophobicity (SMAH) or for the coarse-graining approach. Therefore, fast screening methods are suitable for large-scale data analyses that need to be built up from scratch, involving the examination of many proteins and many fragment possibilities for each protein. In this way, the more detailed SCF calculations can be reserved only for fragments identified as suitable by these faster techniques.

However, fast screening methods have their limitations. On the one hand, the fast screening method is based on the distribution of different amino acid

residues in the primary structure of the protein chain. In nature, protein chains are usually structured with more than one level of structure, having secondary and even tertiary structures that can limit the exposure of certain residues. At the same time, proteins are affected by the charge of the amino acid residues when they are used to stabilise emulsions. For example, a sequence considered by the fast screening methods to be of poor potential (because it has a multi-block structure) may also act as a suitable electrostatic emulsion stabiliser because of the charge it carries. To address this issue in the future, one can take into account the charge of the protein chain by integrating it into the fast screening methods. Another shortcoming of the fast screening method is that it is still empirically driven. For example, there is a lack of quantitative criteria when intercepting the appropriate fragment in a complete protein chain according to the moving average method. In future studies, we should try to specify the basis for selecting protein fragments more carefully, say, for example, by defining ten consecutive identical amino acid fragments as a block. In this way, the fast screening methods can be conducted directly and entirely automatically by the program, thus further increasing the speed of screening for suitable protein fragments. Therefore, while still quite useful, the fast screening method discussed does not entirely eliminate the need for more sophisticated (but computationally expensive) calculations, such as Self-Consistent-Field (SCF) calculations, to be discussed in the following chapter.

Chapter 4
Application of the Self-consistent Field Calculation to Identify the Fragments with the Potential to be Suitable Emulsifiers

We have discovered a potentially suitable emulsifier protein fragment (the di-block-like fragment) in Chapter 3, derived from β-conglycinin, with a structure resembling a di-block. This was identified by applying our two quick screening techniques discussed in Chapter 3. In this chapter, we will continue to analyse the various surface-related properties of this fragment using the more detailed SCF calculation method. As mentioned in chapter 2, we consider two approaching surfaces in a solution of this polypeptide and study various properties of the fragment, including its adsorption on the surfaces, the forces induced between the surfaces and other similar interfacial behaviour. The program converges more easily at shorter surface-surface separation distances (z = 2 ~ 80). In comparison, at larger distances (z = 80 ~ 120), we often need to switch the initial guess solution more frequently to ensure convergence, with this taking a longer computational time (usually 5 – 30 mins) to complete as compared to the more crude but rapid screening techniques of the previous chapter.

4.1 Amount of polymers adsorption

4.1.1 Effect of pH and bulk volume fraction of polymers

Equation 2.29 was applied to calculate the quantity of polymer adsorbed at the interface. In this section, we calculated the adsorption volume fraction profiles when the two surfaces are $120a_0$ (z=120) apart. This ensures that the adsorption of fragments onto one surface is not influenced by the presence of the other surface, that is to say, that the two surfaces can be considered isolated. This is necessary since the amount of adsorption varies at different gap distances between the two interfaces, with one surface affecting the adsorption onto the other one at close separations. Subsequent calculations for volume fraction profiles and configuration and interaction potential are also performed in the following sections.

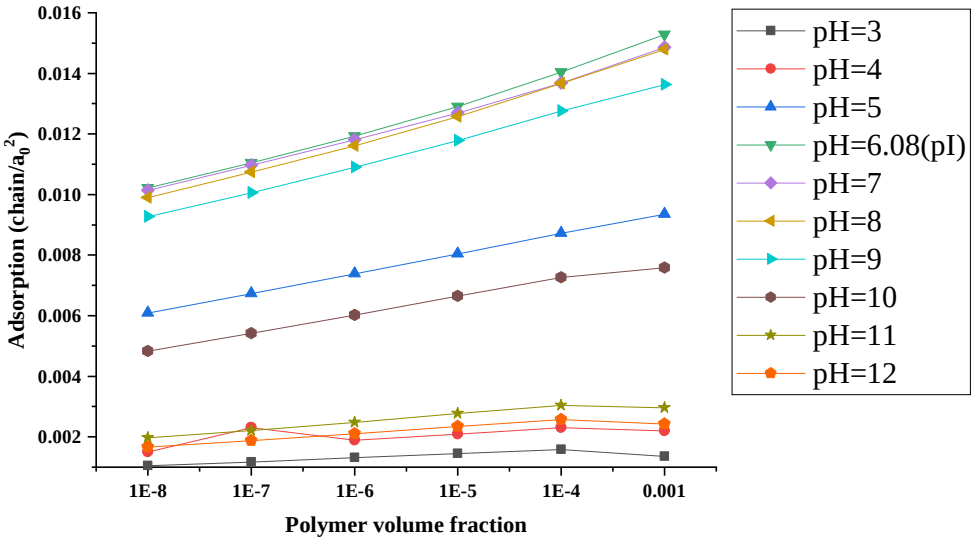


Figure 4.1 Amount of adsorption of the identified di-block-like fragment at various solution bulk volume fractions obtained at various pH values (bulk volume fraction of the ions was 0.001, roughly corresponding to 0.01 mol/litre)

Figure 4.1 illustrates how the adsorption of this di-block-like fragment varies with pH and polymer bulk volume fraction at a constant salt volume fraction. It can be visually seen that pH significantly affects the amount of the adsorbed protein fragment. In the figure, we can group the calculated curves into three different groups. The first is at the iso-electric point, which for this polypeptide is at pH = 6.08. At its pI, we see that the protein fragment attains the maximum adsorption. This is closely followed by curves at pH=7, pH=8 and pH=9, which are similar for all four curves. The second set of adsorption curves is at pH=5 and pH=10, where the adsorbed amounts are seen to be much less than those at the four pH values in the first group. The final group is the pH values far away from the iso-electric point, pH=11, pH=12, pH=3 and pH=4, which are again seen to have exhibited even smaller adsorbed coverage than that at pH values in the second group.

The above phenomenon is related to the net charges carried by the protein. When the pH is equal to or very close to the iso-electric point (i.e., the first group), the net charge of the protein fragment is very small, and therefore the electrostatic repulsion between adjacent chains adsorbed on the surface is almost non-existent. Polymers adsorb as much as possible to the surface until the available space on the particle surface is completely occupied, and the crowding effect due to the strong overlap of a chain on the surface limits

any further adsorption. As the pH increases (or decreases) away from pI., the magnitude of the net charge on the protein chain backbone also increases. The existence of electrostatic repulsion between chains now partially counteracts the affinity of chains to adsorb on the surface resulting in a reduction in adsorption (the second group). As the pH increases further until moving further away from the iso-electric point (the third group), the magnitude of net charges carried by the protein chains become very significant and provide a strong electrostatic repulsive force hindering adsorption of many such mutually repelling chains onto the surface. Thus, the highly charged polymers could only adsorb in tiny amounts to the interface under these conditions. The figure below displays the magnitude of the net charges of our protein fragment at different pH values (bulk volume fraction of polymers = 10^{-3}).

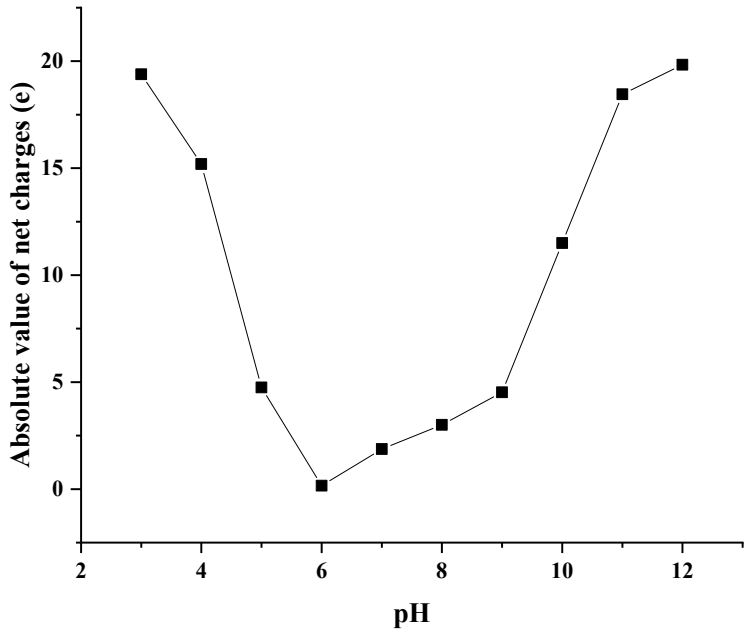


Figure 4.2 Adsolute values of net charge carried by our di-block-like fragment at different pH values

We can also look again at the effect of polymer volume fraction on the amount of adsorption. As shown in Figure 4.1, the adsorption increases slightly as the polymer bulk volume fraction increases at pH values that do not deviate too much from the iso-electric point (the first and second groups). This is because, at higher polymer volume fractions, more of the polymer is available to adsorb at the interface, resulting in a rather moderate increase

in polymer coverage of the surface. When the pH is far from the iso-electric point (the third group), the polymer bulk volume fraction has almost no effect on the amount of adsorption, which remains small at all volume fractions. Therefore, the electrostatic repulsion plays a dominant role at this point and completely counteracts the effect of increasing the polymer bulk volume fraction on the adsorption. The result also shows that even at very low bulk volume fractions, the fragment, in common with proteins more generally, almost saturates the surface and attains its maximum possible coverage.

From the above description, it is clear that the effect of pH on the amount of polymer adsorbed is very significant and far more so than the bulk volume fraction of the protein. For example, the adsorbed amount at a polypeptide bulk volume fraction of 10^{-3} and a pH value of 12 is 0.00243 chains per monomer unit area, whereas the pH=pI is 0.01529 chains per monomer unit area. This is a 6.29-fold increase. In contrast, at pH=pI, the coverage only increases from 0.01022 chains to 0.01529 chains per monomer unit area when the bulk volume fraction of the polypeptide is increased by four orders of magnitude from 10^{-7} to 10^{-3}, not even doubling in value. Thus, the polymer volume fraction has a relatively small effect on the amount of polymer adsorbed at the interface, particularly in comparison to the impact of pH. Moreover, in the food industry, it is impossible to increase emulsifiers' volume fraction indefinitely to increase the amount of adsorption, as this would not be cost-effective. Based on this conclusion, the bulk volume fraction of the polymer is 10^{-7} in all the following calculations.

4.1.2 Effect of salt volume fraction

Figure 4.3 shows the variation of the total amount of polymer adsorbed at the interface as a function of pH and salt volume fraction. We can divide the curves in this figure into two groups: four curves representing low salt volume fractions (salt volume fraction = 0.0001, 0.001, 0.005 and 0.01) and two curves representing high salt volume fractions (salt volume fraction = 0.05 and 0.1). Salt ions play a significant role in the system. They tend to screen the electrostatic interactions. On the one hand, the counter-ions of the salt screen the charges of the charged hydrophilic amino acid residues of the protein chains. This tends to reduce the repulsion between the protein chains (or, in this case, protein fragments), thus making it more difficult for the hydrophilic parts of the chains to extend further into the solution. This results in less extended and thinner interfacial layers. However, on the other

hand, the screening effect makes the electrostatic repulsion between the polymers less, thus allowing for an increased amount of polymer chains to adsorb to the surface of emulsion droplets. In this figure, we can find that the screening has the latter effect dominantly.

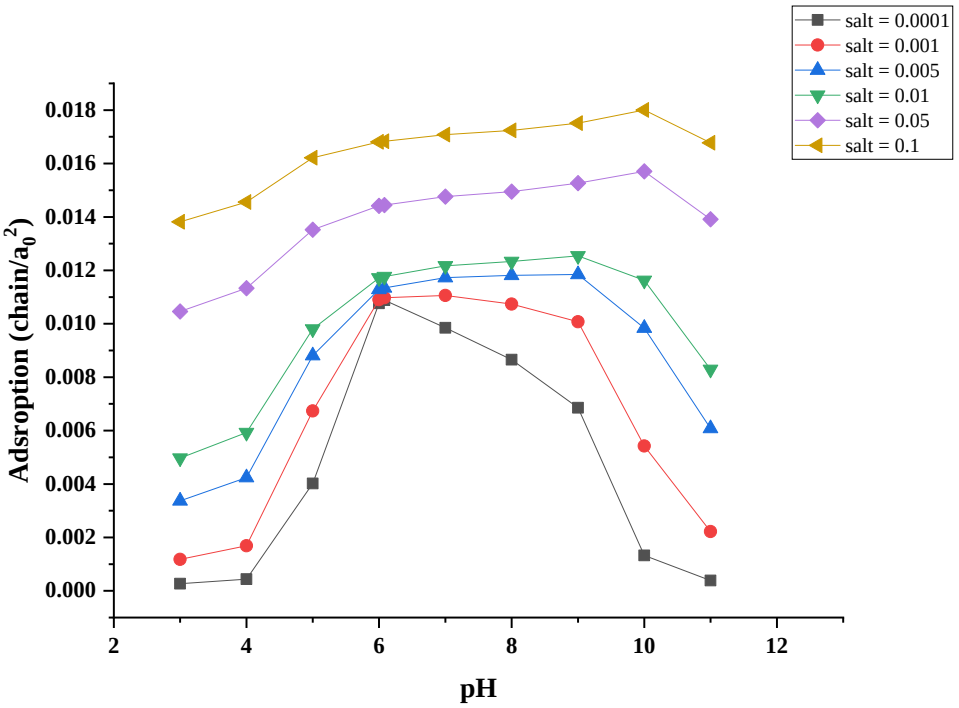


Figure 4.3 Polymer adsorption of our soya β-conglycinin derived di-block-like fragment at different salt volume fractions and pH values

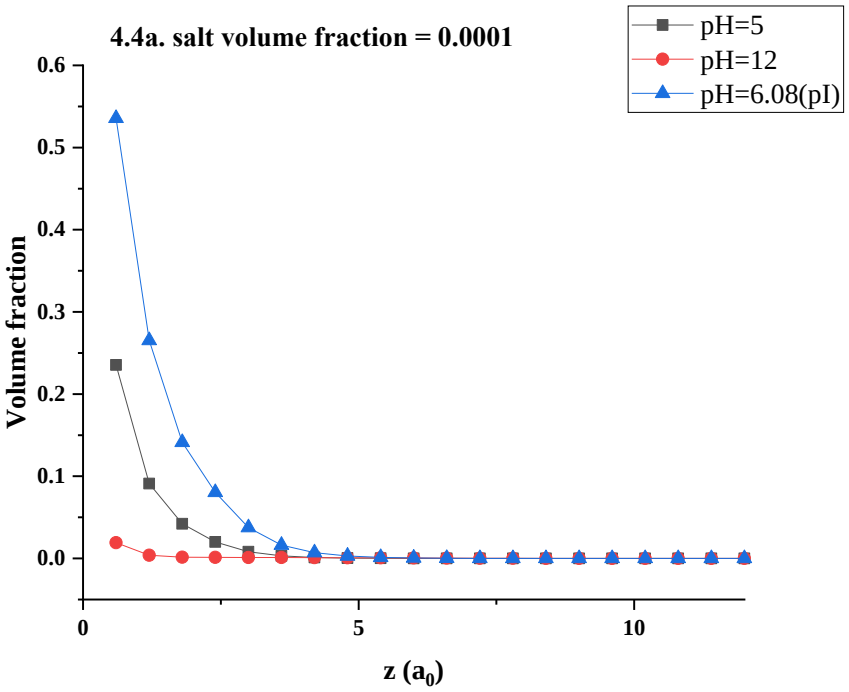
For the curves with low salt volume fractions, the overall trend in adsorption, at pH values away from the iso-electric point, is that the higher the salt volume fraction, the more pronounced the adsorbed amount. This can be understood by considering the screening effect of the salt on the surface charge at the interface due to adsorbed protein chains. The higher the salt volume fraction, the stronger the screening effect and the weaker the electrostatic repulsion between polymers. Near the iso-electric point, where the chains have no net charge, the four curves overlap, and there is no appreciable impact due to salt volume fraction. This is indeed expected since, at the iso-electric point, the electrostatic repulsion between the chains is already negligible, and any screening of this has minimal impact on the adsorption behaviour of the fragment. The differences between the four curves, representing the four lowest salt volume fractions, become progressively more pronounced as the pH deviates from pI. At pH values far from pI, the chains have substantial net charges and therefore altering salt,

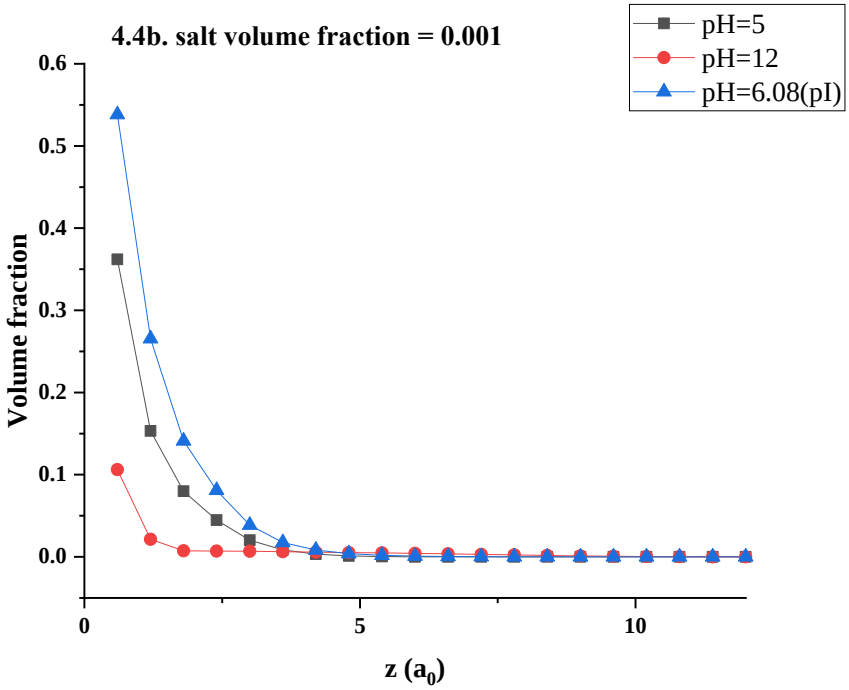
the electrostatic screening effect becomes the main factor limiting the amount of adsorption. Thus, the adsorption now becomes much more sensitive to the amount of salt present in the system. Furthermore, one can note that the shape of the four curves is not entirely symmetrical around the iso-electric pH. As can be seen in Figure 4.2, this is because the net charges on the protein chain do not increase uniformly with increasing pH. The increase in the magnitude of the net charge is greater at pH < pI than at pH > pI.

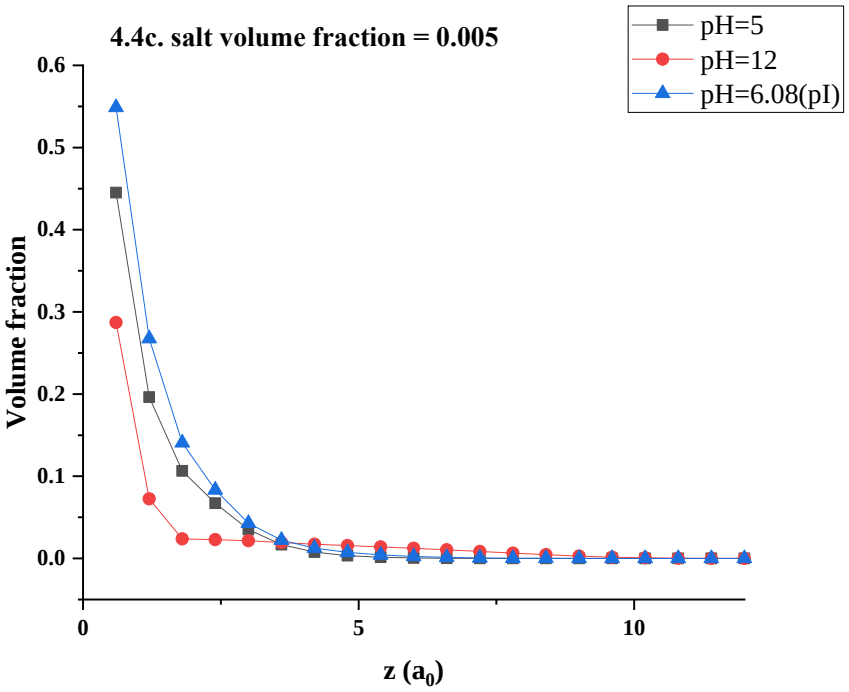
The curves for high salt volume fractions are very different from those for low salt volume fractions and do not seem to follow the above-described trend. They deserve some careful consideration. Firstly, at high salt volume fractions, the values of the adsorbed amount are roughly constant, remaining so whether one is at the iso-electric point or not. This may be explained by the fact that at very high ionic strengths, the charge of the protein is almost wholly screened at any pH so that pH can no longer influence the amount of polymer adsorbed. In addition, at the iso-electric point, the adsorption curves at high salt volume fractions do not overlap in the same manner as they do at low salt volume fractions. The amount of adsorption can be further increased when the salt volume fraction increases from 0.01 to 0.1. This is because a large amount of salt can lower the activity of water, which also changes the nature of the solution.

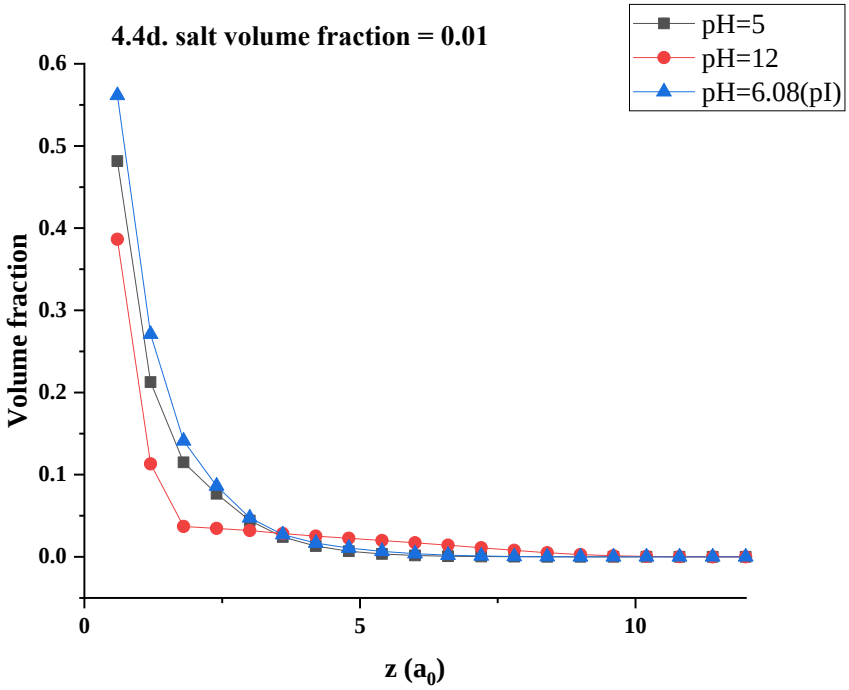
4.2 Volume fraction profile of the adsorbed protein fragment at the interface at different pH and salt volume fractions

As mentioned in Chapter 2, the amount of adsorbed protein fragments is calculated directly from the density profile variation of the protein distributed in the interfacial region, as obtained from Equation 2.17. It is constructive to look at this volume fraction distribution in more detail, as this provides some information on how extended (thick) the adsorbed polymeric interfacial layer may be. Therefore, the volume fraction of this di-block-like fragment was further analysed in combination with the results of the previous section. Figure 4.4 shows the variation of the polymer volume fraction plotted against distance away from the interface, calculated at six different salt volume fractions. We selected three different pH values from each of the three groups in Figure 4.1 in section 4.1.1 based on the adsorbed amount.


4.4a. salt volume fraction = 0.0001
pH=5
pH=12
pH=6.08(pI)
Volume fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
z (a0)


4.4b. salt volume fraction = 0.001
pH=5
pH=12
pH=6.08(pI)
Volume fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
z (a0)


4.4c. salt volume fraction = 0.005
pH=5
pH=12
pH=6.08(pI)
Volume fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
z (a0)


4.4d. salt volume fraction = 0.01
pH=5
pH=12
pH=6.08(pI)
Volume fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
z (a0)

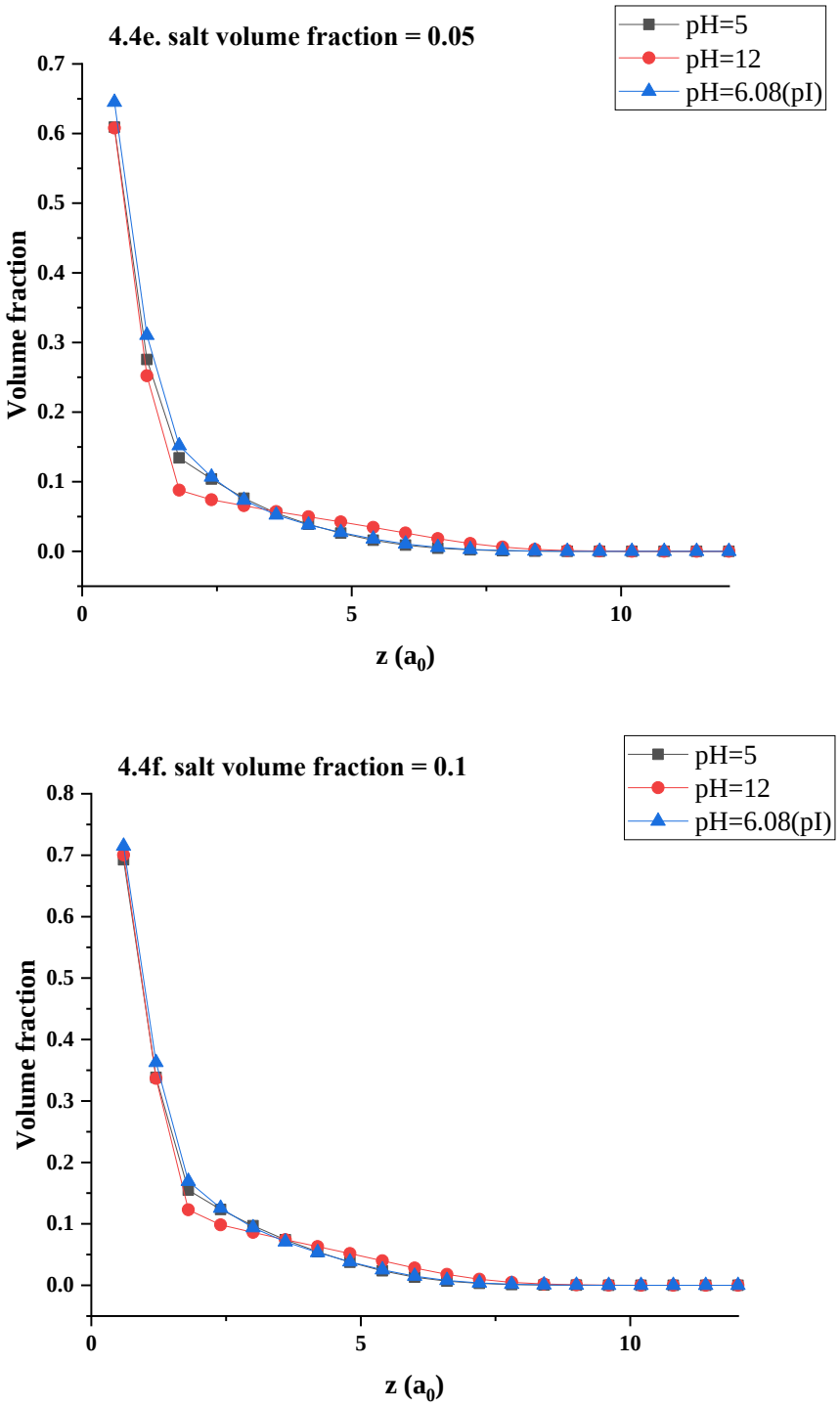


Figure 4.4 Variation of volume fractions of our chosen di-block-like fragment from β-conglycinin at a distance away from the surface when adsorbed at a hydrophobic-hydrophilic interface, as obtained under several different pHs and bulk electrolyte volume fractions

The most intuitive conclusion from the above six sets of curves is that the volume fraction of the fragment rapidly decreases with increasing distance away from the interface approaching its bulk value. These graphs indicate a

much higher volume fraction of polymer close to the interface, clearly reinforcing the result of the previous section that the chosen fragment has sufficient affinity for the surface and is adsorbed onto the interface. Though not in this case, in some other fragments, one may see that the fragment is actually depleted from the interfacial region, i.e. the polymer's volume fraction increases towards the bulk value as one moves away from the surface. These types of fragments are not likely to be very effective as emulsifiers.

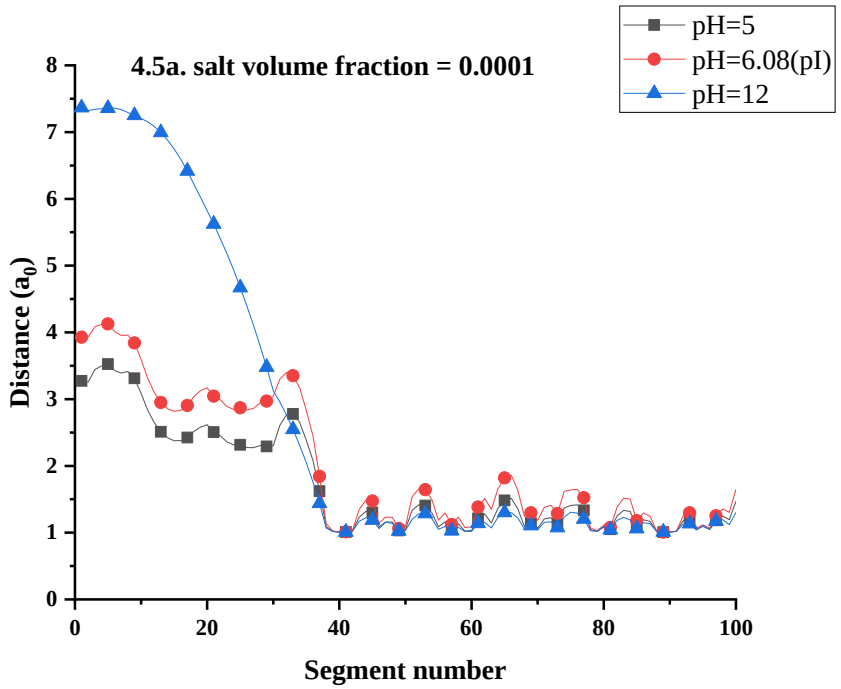
Moreover, we can also see from graphs in Figures 4.4e and 4.4f that the volume fraction of the polymer also tends to a constant maximum at the surface, regardless of the pH value, for systems with high bulk salt volume fractions. This is because, at these high salt volume fractions, the electrostatic repulsion is no longer the constraining factor for adsorption, but rather it is the packing and the space available to adsorbed chains on the surface of droplets that will limit the number of adsorbed chains. This maximum available space on the surface is not particularly altered by pH.

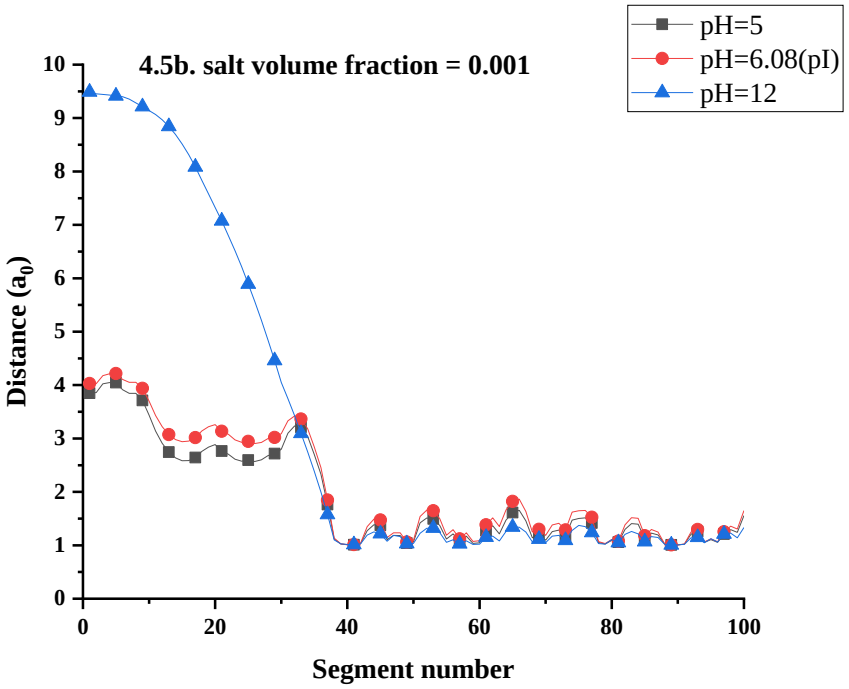
At low salt volume fractions, the effect of pH on volume fraction is significant. For example, at a salt volume fraction of 0.0001, the maximum volume fraction of the protein fragment in the layer next to the surface for pH=pI is 0.535, while for pH=12, it is only 0.019. However, when the salt volume fraction increases to 0.1, the maximum volume fraction of the fragment occurring in the layer immediately in contact with the surface, for pH=pI, is 0.715, while it is calculated at 0.700 when pH=12. The two values are seen not to be all that different. This situation is obviously related to and has a similar explanation as the variation in the total amount of adsorption, as was discussed in the previous section. Once again, these differences in the adsorption behaviour can be understood on the basis of where the electrostatic screening comes into play, being very effective at high salt volume fractions while not being so prominent at lower ones. The high volume fraction of salt screens the charge on the protein chain, allowing a higher volume fraction of adsorbed protein close to the interface. In addition, the polymer tends to take further to reach its bulk volume fraction at longer distances away from the interface when salt volume fractions are high. For example, at salt volume fraction = 0.0001, the curve obtained for pH=5 drops to the bulk volume fraction at a distance of ~ 7.8 nm, whereas at a salt volume fraction = 0.1, this distance increases to 12 nm away from the surface. This may be because the high salt volume fraction enriches the

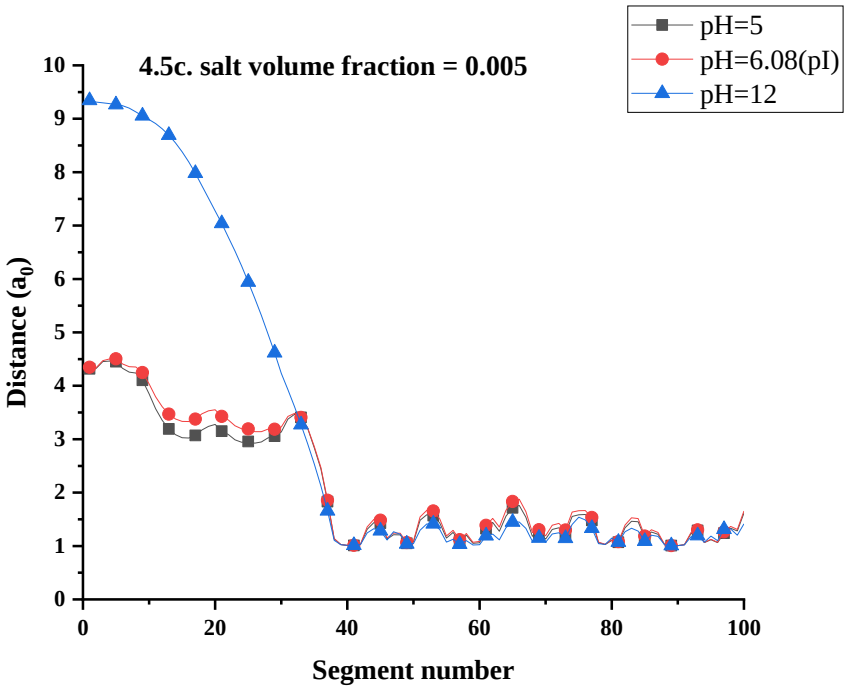
polymer on the surface of the droplets. The thickness of the interfacial layer is both a function of the number of adsorbed chains and the extent to which any individual chain protrudes away from the surface. The presence of a high amount of salt impacts these two factors differently. High background electrolyte in the system screens the repulsion between the chains, thus promoting a higher amount of adsorption, which in turn leads to thicker layers. But chains that repel each other more strongly do tend to extend further away from the surface when adsorbed. Thus, these two effects somewhat oppose each other. Our results show that the first of these, namely a larger degree of adsorption of chains, has a greater influence on the thickness of the interfacial layer formed by our protein fragment and can more than compensate for the second. The net result is that thicker surface layers are formed at higher salt volume fractions.

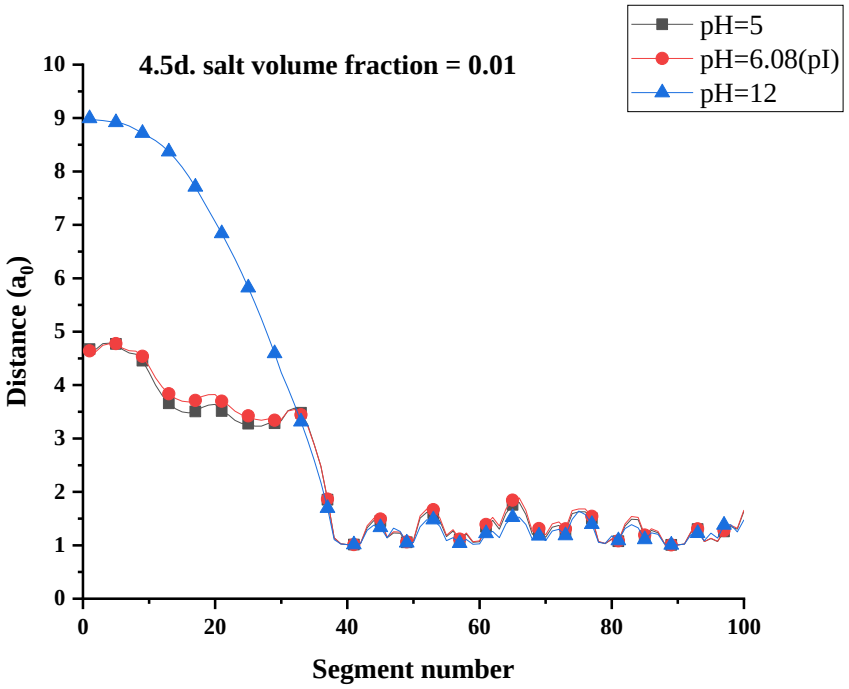
4.3 Configuration of the protein fragment adsorbed on the surface

The average distance of each amino acid residue of the protein fragment chain from the interface can provide further valuable information regarding the suitability of the fragment to be a good emulsifier. This information can once again be calculated using the SCF method and obtained via Equation 2.27. Figure 4.5 shows the effect of pH values on the conformation of the protein chain at different salt volume fractions. In the discussion part in Chapters 2 and 3 regarding the fast screening methods and the adsorption behaviour of different conformational peptides (Figure 2.3), we speculated that when this di-block-like fragment adsorbs to the interface, it should form a train of residues, seen as an adsorbed section, followed by a dangling end that extends away from the surface. The following six figures all confirm this idea. The adsorbed fragment shows a dangling end formed from the segment residue numbers from 1 - 40 (counting from the N-terminus end) and a train structure remaining very close to the surface from 40 - 100.


4.5a. salt volume fraction = 0.0001
pH=5
pH=6.08(pI)
pH=12
Distance (a_0)
Segment number
0
1
2
3
4
5
6
7
8
0
20
40
60
80
100


4.5b. salt volume fraction = 0.001
pH=5
pH=6.08(pI)
pH=12
Distance (a_0)
Segment number
0
1
2
3
4
5
6
7
8
9
10
0
20
40
60
80
100


4.5c. salt volume fraction = 0.005
pH=5
pH=6.08(pI)
pH=12
Distance (a_0)
Segment number
0
1
2
3
4
5
6
7
8
9
10
0
20
40
60
80
100


4.5d. salt volume fraction = 0.01
pH=5
pH=6.08(pI)
pH=12
Distance (a_0)
Segment number
0
1
2
3
4
5
6
7
8
9
10
0
20
40
60
80
100

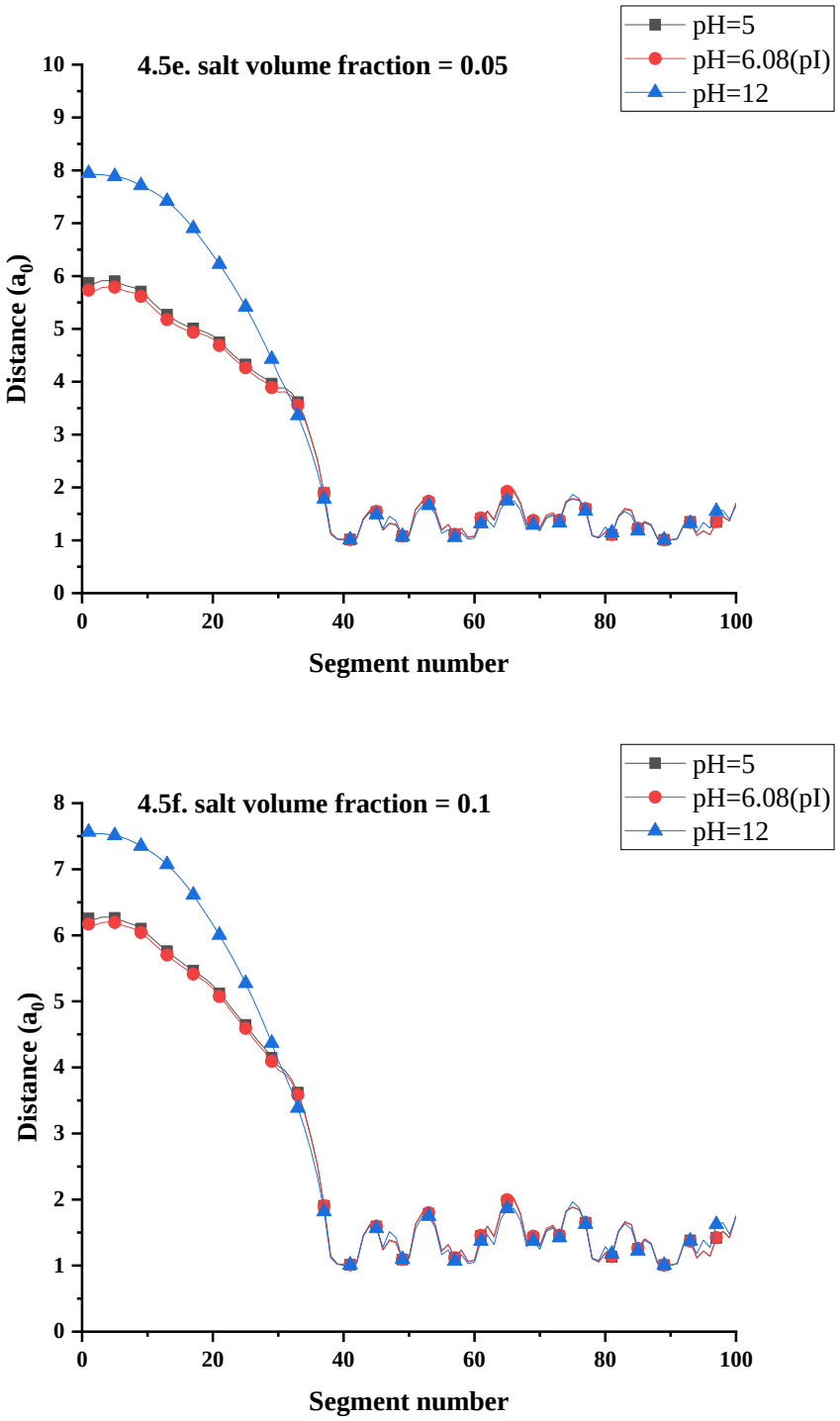


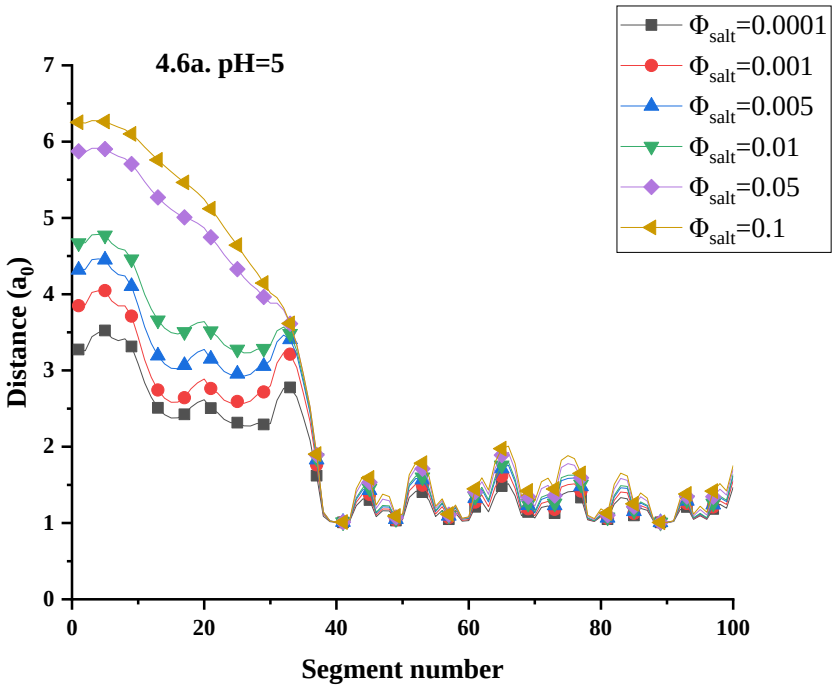
Figure 4.5 Average distance of each amino-acid monomer comprising our β-conglycinin-derived protein fragment from the interface. The amino-acid residues are numbered consecutively from 1 to 100, starting from the N-terminus side of the polypeptide fragment. The result for the adsorbed fragment on the surface is shown at different pH values and several salt volume fractions

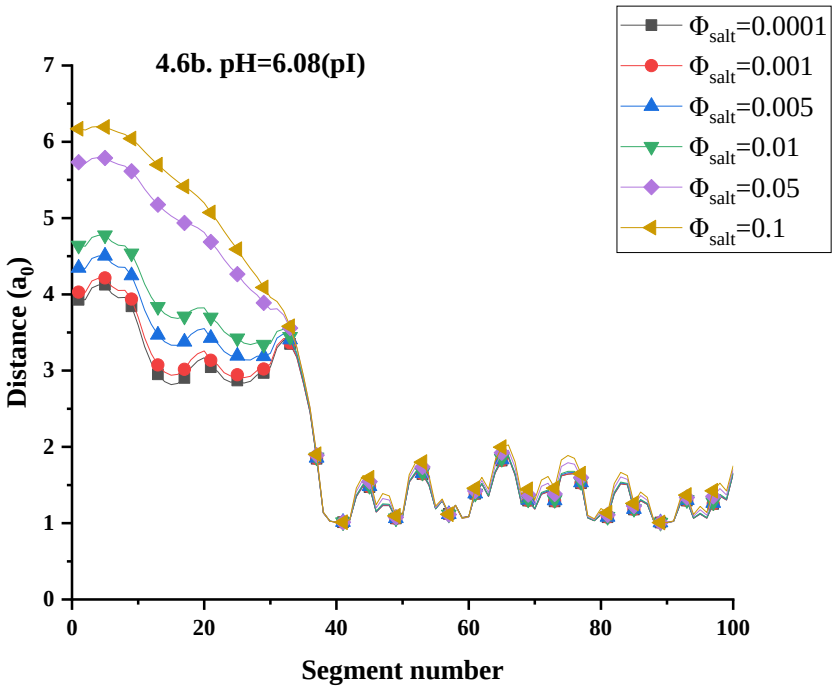
The figures demonstrate that the confirmation of the previously identified hydrophilic block (see chapter 3) is the part of the adsorbed fragment that is

most affected by pH. No matter what the salt volume fraction, the hydrophilic block extends significantly away from the surface when the pH is or above the iso-electric point (for example, at pH=12). The hydrophilic block protrudes to a lesser extent below the iso-electric point (i.e. pH=5) and at the iso-electric point (pH=6.08). However, the extension of the fragment below, i.e. and at the iso-electric point, remains very similar to each other at all salt volume fractions. We suggest that the self-charge on the hydrophilic block of the fragment may be the cause of this behaviour. The charged hydrophilic group would have a higher net charge and stronger hydrophilicity at pH values further away from the iso-electric point. As a result, fragments with higher hydrophilicity will adopt configurations that typically maintain their hydrophilic blocks away from the hydrophobic-hydrophilic interfaces, extending more into the solvent.

We saw that the results of section 4.1 highlighted that an increase in charge decreases the number of fragments adsorbed at the interface at low salt volume fractions. Thus, the hydrophilic residues at the intersection of the two blocks are more influenced by the presence of the hydrophobic part of the fragment. According to our calculations, it is evident that increasing the pH (away from the iso-electric point) has a beneficial impact on causing a larger protrusion of the hydrophilic block into the solution, away from the interface (i.e., helps the formation of thicker interfacial layers). However, as discussed in section 4.2., the higher charge also hinders the adsorption of a larger number of chains onto the interface. This later lead to a thinner layer. Result in section 4.2. indicate that the overall impact of a higher charge of the fragment (i.e., pH away from pI) is a likely decrease in the thickness of the adsorbed surface layer.

Figure 4.6 below highlights the impact of the background electrolyte volume fraction on the conformation of the adsorbed fragments.


4.6a. pH=5
Distance (a_0)
Segment number
Φ_{salt}=0.0001
Φ_{salt}=0.001
Φ_{salt}=0.005
Φ_{salt}=0.01
Φ_{salt}=0.05
Φ_{salt}=0.1


4.6b. pH=6.08(pI)
Distance (a_0)
Segment number
Φ_{salt}=0.0001
Φ_{salt}=0.001
Φ_{salt}=0.005
Φ_{salt}=0.01
Φ_{salt}=0.05
Φ_{salt}=0.1

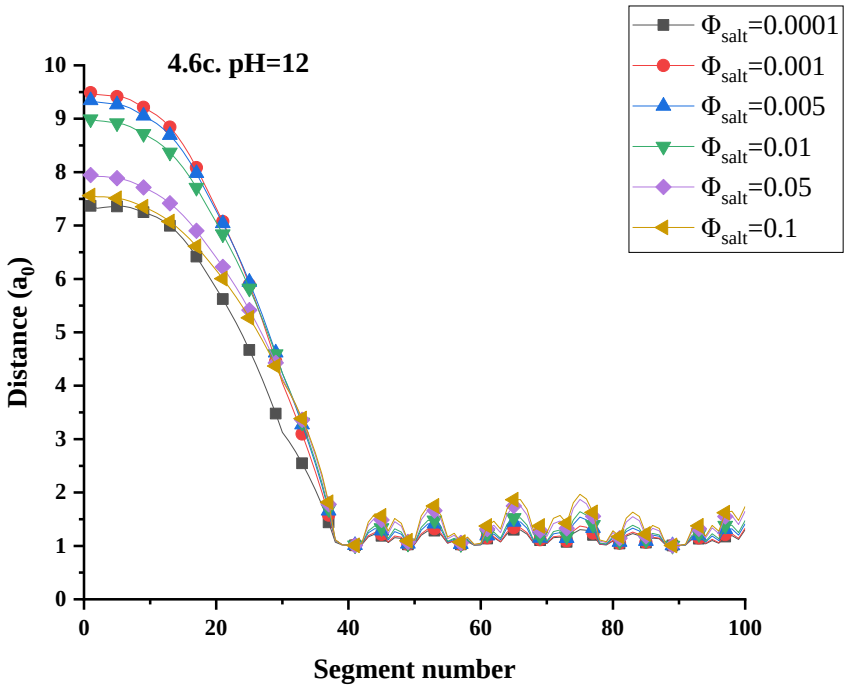


Figure 4.6 Confirmation adopted by our β-conglycinin derived protein di-block-like fragment adsorbed on the surface at different pH values and salt volume fractions. The graphs show the average position of each amino-acid residue of the fragment away from the hydrophobic-hydrophilic interface (in units of monomer size~0.3 nm), with monomers ranked from 1 to 100, starting from the N-terminus side of the polypeptide

Figures 4.6a and 4.6b show that the extension of the dangling tails is encouraged by an increase in the salt volume fraction close to the iso-electric point, but only at very high salt volume fractions. This makes sense because chains do not carry much charge when the fragment is close to its iso-electric point. Therefore, the hydrophilicity of the fragment will not be greatly affected by the screening of charge due to increasing salt volume fraction. This is clearly seen for the four lower salt volume fraction curves in figure 4.6 at pH=6.08, (i.e., at pI). Figure 4.3 in section 4.1 showed that the number of fragments adsorbed rises as salt volume fraction increases, with the most potent effect happening at high salt volume fractions ($\phi_{salt} = 0.05$ and $\phi_{salt} = 0.1$). This is corroborated by the fact that the adsorbed fragment for the four curves at lower salt volume fractions has somewhat less extended tails compared to those at $\phi_{salt} = 0.05$ and $\phi_{salt} = 0.1$. High salt volume fractions above 0.05 (~ 0.5 mole/litre) substantially alter the quality of the aqueous solvent, beginning to change the activity of the water molecules. This effect is only observed at these very high salt volume fractions. At lower volume fractions, the influence of added electrolyte is purely one of charge

screening (which becomes irrelevant at pI), with no impact on water activity. Reducing the quality of the solvent for the protein tends to cause more polypeptide fragments to adsorb at the interface. In turn, this increasing level of adsorption has a greater impact on the conformation adopted by chains when the pH is close to the iso-electric point, forcing the hydrophilic end of the di-block-like polypeptide fragment to extend more away from the interface (see $\phi_{salt} = 0.05$ and $\phi_{salt} = 0.1$ curves in figure 4.6b).

Contrary to the above behaviour, the extension of the tail end into the bulk solution at pH values away from the iso-electric point tends to be less as the salt volume fraction increases (see Figure 4.6c). This mostly depends on the balance between the hydrophilicity of the fragment, dictating the amount of adsorbed chains, and the screening effect from the salt ions. The image below illustrates the fundamental concepts and their impact on the conformation of the protein chain and, in particular, the extension of the hydrophilic side away from the interface. (This principle will also be utilised in the following section when discussing the interaction potentials mediated between droplets by adsorbed fragments on the surfaces).

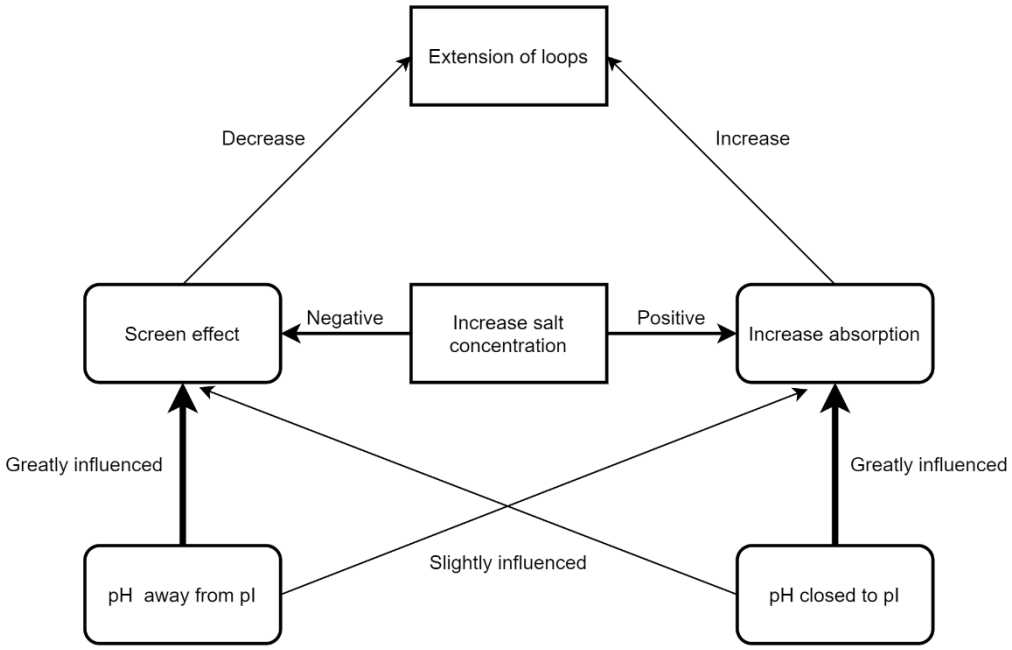


Figure 4.7 The interplay between the hydrophilicity of the fragment and the screening effect of added electrolyte on the amount of adsorbed chains and adopted conformations by the fragment

At a pH far from the iso-electric point, the fragments carry sufficient charges so that the screening effect of salt ions cannot be ignored. When the salt volume fraction is not yet very high (e.g. $\phi_{salt} = 0.0001 \rightarrow \phi_{salt} = 0.001$), an appropriate increase in salt volume fraction can increase the amount of fragment adsorption while still retaining some level of repulsive electrostatic interaction between them. This repulsion forces the dangling tails of the protein chains to stretch to the maximum extent possible away from the surface at this salt volume fraction ($\phi_{salt} = 0.001$). When the salt volume fraction is increased further, the gains from increased fragment adsorption do not offset the losses from reduced hydrophilicity and lower repulsion between the chains. Therefore, at higher salt volume fractions ($\phi_{salt} = 0.005$, $\phi_{salt} = 0.01$ and $\phi_{salt} = 0.1$), we observe a decrease in the outward extension of the tails with increasing salt volume fraction. In short, the impact of a higher number of adsorbed chains and repulsion between the protein fragments, both of which contribute positively to thicker layers, is oppositely affected by the increasing bulk electrolyte at pH values away from pI. This then leads to some optimum salt volume fraction at which the extension is maximum, but this is reduced with a higher or lower amount of electrolyte.

4.4 Interaction potential

In this section, we wish to consider the interactions that are mediated between two approaching droplets when their surfaces are covered by our adsorbed di-block-like, conglycinin-derived protein fragment. Apart from We also wish to include in our calculations the more direct van der Waals forces that operate between the droplets, irrespective of the presence or absence of the fragments. Such calculations provide a wealth of information on the suitability of our protein fragment to behave as a good emulsion stabiliser (stabilising emulsion for long-term storage). The van der Waals interactions between two droplets of radii *R* are given by the following equation (Dickinson and Stainsby, 1982)

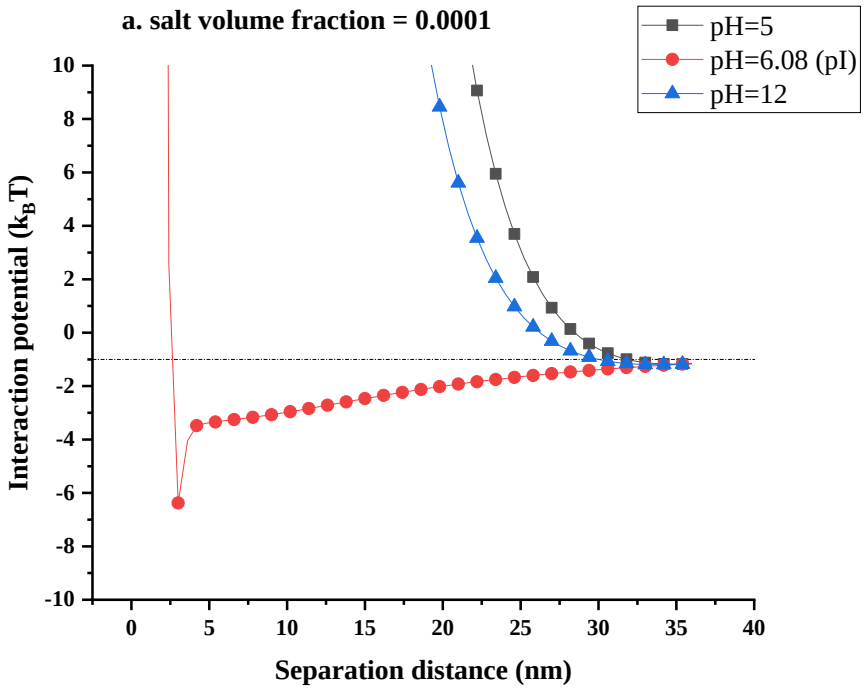
$$V_{vw} = -\frac{A_H R}{12r} \tag{4.1}$$

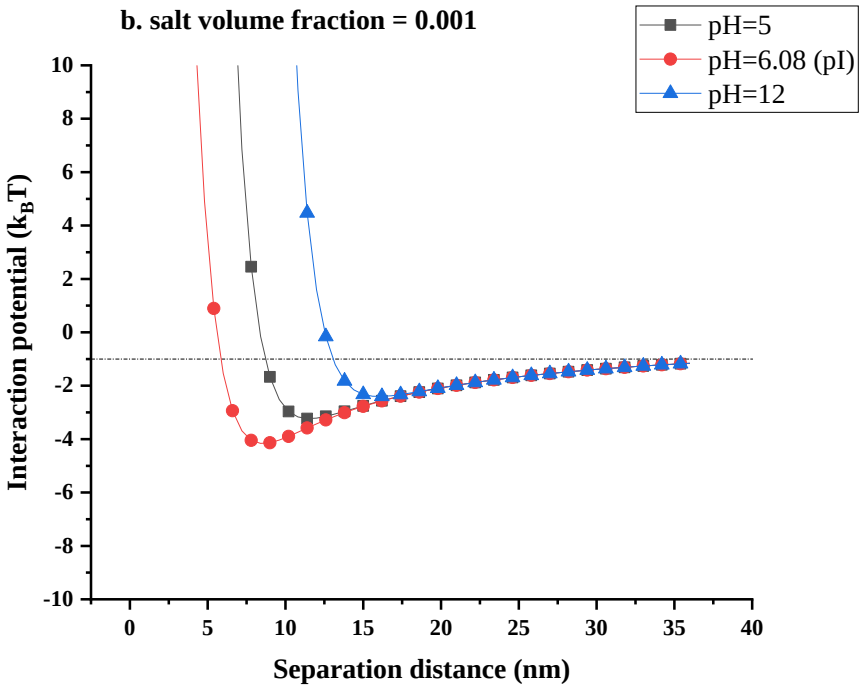
Where A_H is the composite Hamaker constant for the dispersed phase in the continuous medium, with a typical value of 1 k_BT for edible oils dispersed in water, and r is the separation distance between the droplets.

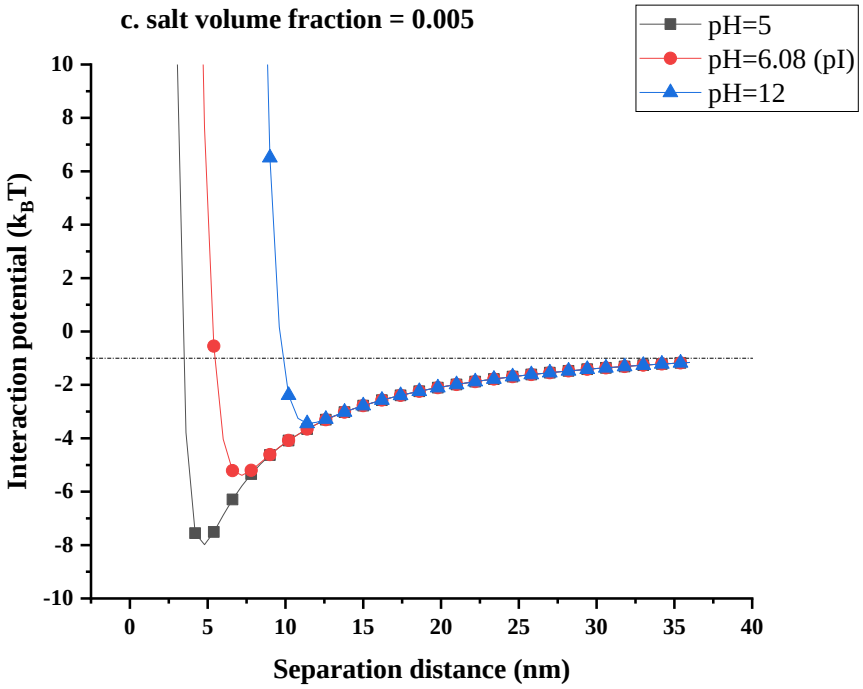
The Interaction potential mediated by the fragments is calculated by obtaining the free energy and the interaction force between two flat planes at different distances using Equations 2.22 and 2.23. This is then transformed into a particle-particle interaction by the use of Deryagin approximation (Equation 2.24). Finally, the direct van der Waals attraction, given by equation 4.1, is added to this result to obtain the final full interaction potential between the droplets. We will first investigate how pHs and salt volume fractions affect the interaction potential induced by adsorbed fragments. Following that, this interaction mediated by polypeptide fragments will be contrasted with those resulting from the adsorption of β-casein, as well as another fragment that we identified as having a rather poor emulsification property in Chapter 3.

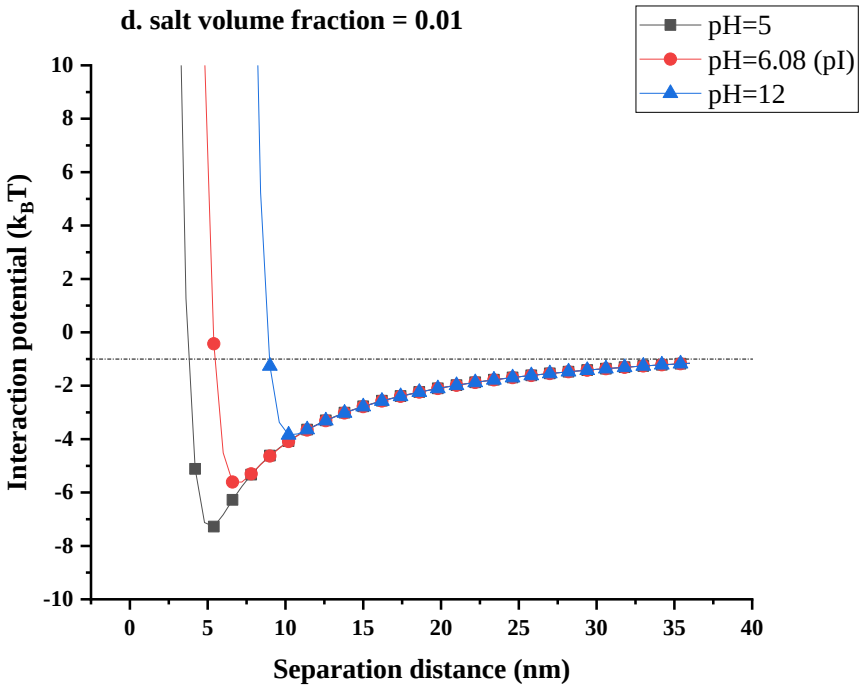
4.4.1 Effect of pH values and salt volume fractions

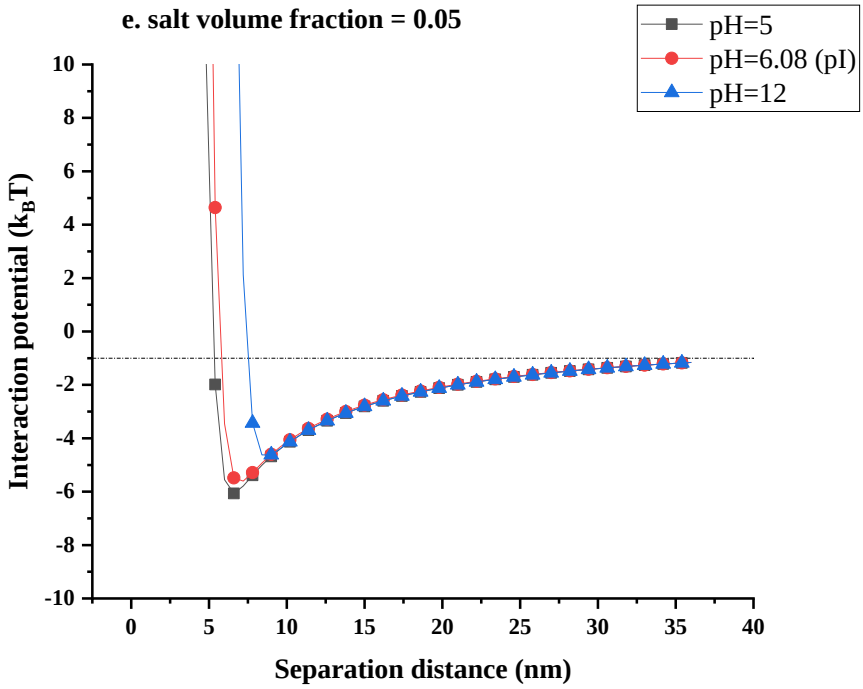
Figures 4.8 and 4.9 present the variation in the interaction potential for the three pH values where the amount of adsorption was found to be significantly different to each other (pH=5, pH=6.08 and pH=12) in Figure 4.1.




b. salt volume fraction = 0.001
pH=5
pH=6.08 (pI)
pH=12
Interaction potential (k_BT)
Separation distance (nm)
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
5
10
15
20
25
30
35
40


c. salt volume fraction = 0.005
pH=5
pH=6.08 (pI)
pH=12
Interaction potential (k_BT)
Separation distance (nm)
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
5
10
15
20
25
30
35
40


d. salt volume fraction = 0.01
pH=5
pH=6.08 (pI)
pH=12
Interaction potential (k_BT)
Separation distance (nm)
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
5
10
15
20
25
30
35
40


e. salt volume fraction = 0.05
pH=5
pH=6.08 (pI)
pH=12
Interaction potential (k_BT)
Separation distance (nm)
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
5
10
15
20
25
30
35
40

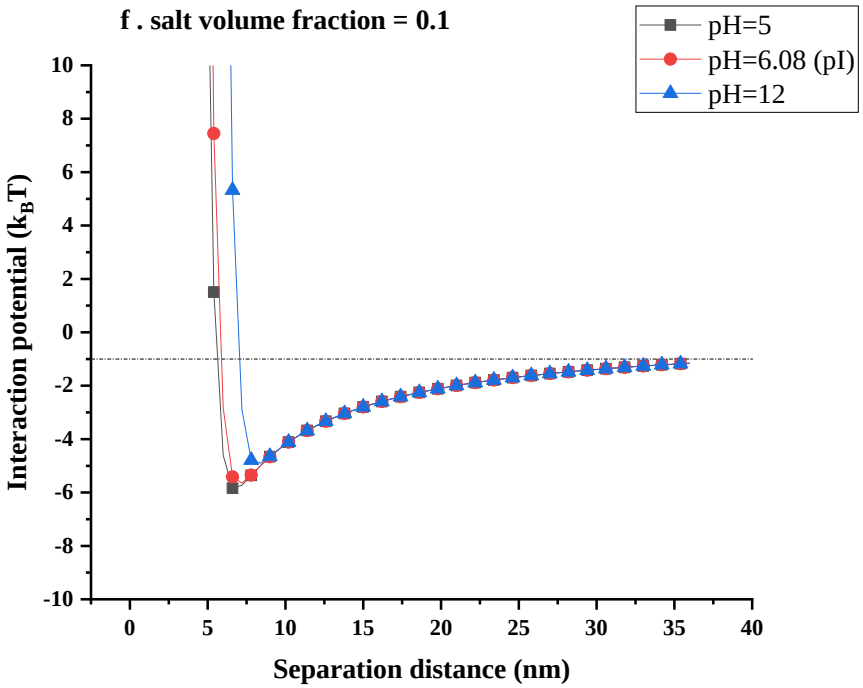
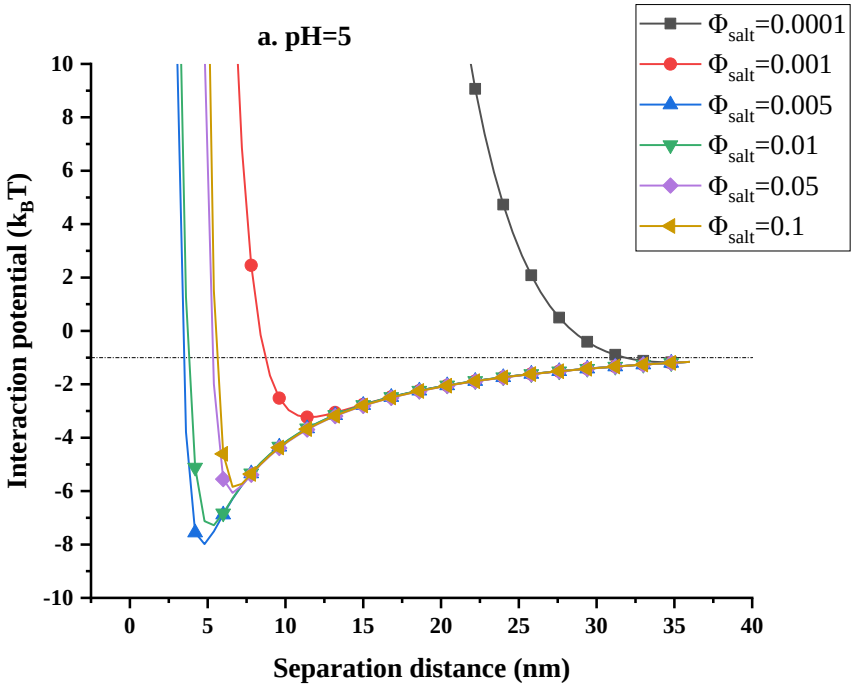


Figure 4.8 Interaction potential between two approaching droplets, induced by interfacial adsorbed conglycinin derived di-block-like protein fragment (see Figure 3.4 in Chapter 3), plotted against the separation distance between the droplets. Results are calculated for different salt volume fractions, at their pH values, and for droplets of size 1 μm



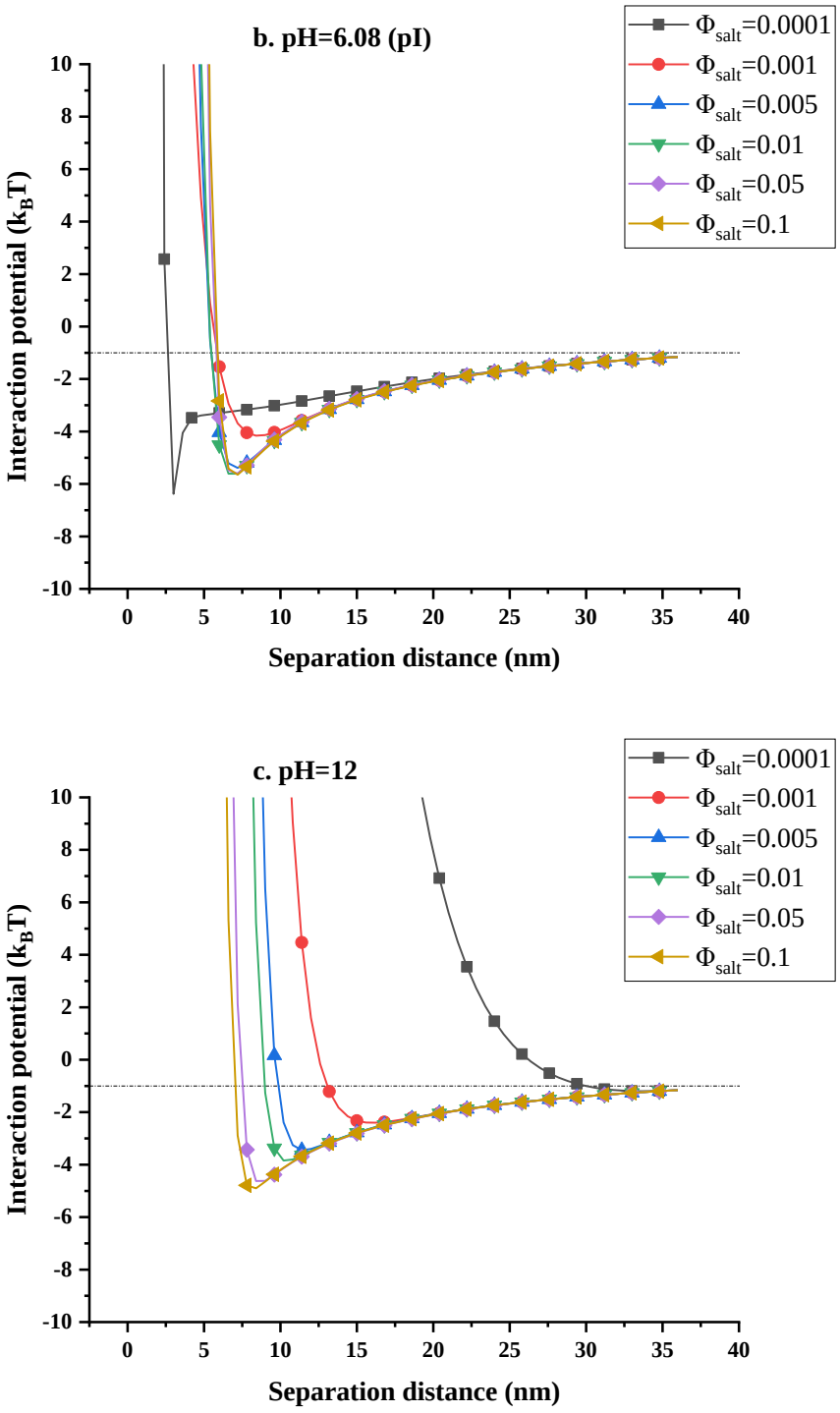


Figure 4.9 Interaction potential induced between two droplets by adsorbed interfacial layers of the di-block-like conglycinin-derived protein fragment (see Figure 3.4), plotted against separation distance between the droplets. The results are obtained for systems at various pH values, each at six different background electrolyte volume fractions. The size of droplets was taken as 1 μm

We begin our description with the shape of the curves. In the nine figures above (Figures 4.8 and 4.9), the value of the interaction potential starts to

fall from a negative value to a minimum value as the droplet-droplet separation is reduced from a far distance to near one. At very short separations, there is generally a sharp rise in the interaction potential, indicating a strong repulsion between the droplets at these close distances. In our calculations, three forces are involved: steric repulsion, electrostatic repulsion and van der Waals force. These three forces act at different distances; the steric repulsion and van der Waals force are more short-ranged, while the electrostatic repulsion is a relatively long-ranged interaction (depending, of course, on the amount of background electrolyte). In addition, the interactions indicated by interfacial protein layers can also become attractive, for example, through such effects as bridging. Though here, due to the deliberately chosen, more di-block nature of our polypeptide fragment, this is less likely. When two particles are very far apart (e.g. r = 120 a_0 in this study), steric repulsion does not exist. The range of action of the electrostatic repulsion depends mainly on the size of the Debye length (and hence background salt in the system) as discussed in electrostatic double layer theory in Chapter 1, so when the distance is large compared to the Debye length, the electrostatic repulsion is also almost non-existent. On the other hand, the equation for the van der Waals force, combined with Figure 4.9, shows that although the van der Waals force is small at longer distances, it tends to dominate, causing a slightly attractive (a negative interaction potential) at these larger separations. Therefore, the interaction potential always has a negative value at long droplet-droplet separation distances, though the magnitude of this asymptotically approaches zero as the distance *r* approaches infinity (see equation 4.1). As the distance decreases to values comparable to the Debye length, the electrostatic repulsion begins to appear and partially counteracts the attractive van der Waals force. The minimum value of the interaction potential occurs when the slope at a certain distance of van der Waals force + electrostatic repulsion is zero. In the absence of steric forces, this relationship between electrostatic repulsion and van der Waals forces is captured in the well-known DLVO theory (Dickinson, 1992b). As the separation distance between droplets decreases further, the steric repulsion also appears and, in combination with electrostatic interactions, opposes the van der Waals forces. This is seen as a sharp upward rise in the curve for the overall interaction potential (see Figures 4.8 and 4.9).

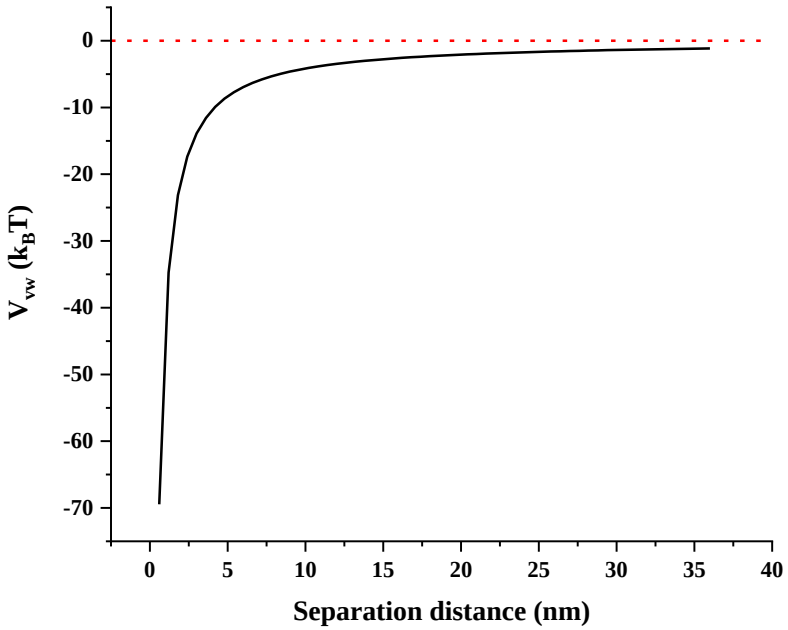


Figure 4.10 The attractive van der Waals interaction energy between a pair of droplets, plotted as a function of their separation distance

At low salt volume fractions (see Figure 4.8a), the minimum value of the interaction potential at the iso-electric point is -6.38 k_BT. In contrast, the minimum at two pH values away and on opposing sides of the iso-electric point (pH=5 and pH=12) are relatively similar to each other, at -1.17 and -1.20 k_BT, respectively. The magnitude of the minimum of the interaction potential at the iso-electric point is larger than that at the non-isoelectric points. This conclusion also applies to the case with $\phi_{salt} = 0.001$ (see Figure 4.7b). The order between the depth of the minima in the three curves and pH changes $pH = 6.08 < (pH = 5 \approx pH = 12)$ to $pH = 5 < pH = 6.08 < pH = 12$ when the salt volume fraction is further increased (see Figures 4.7c - 4.7f). This is mainly because, at low salt volume fractions ($\phi_{salt} = 0.0001$ and $\phi_{salt} = 0.001$), the ionic strength does not have a significant screening effect on electrostatic repulsion induced by the presence of charge fragments on the surface. At this point, both below and above the iso-electric point (pH=5 and pH=12), there is a sufficient surface charge to provide electrostatic repulsion, which happens to be the dominant repulsive interaction at these lower electrolyte volume fractions. The electrostatic repulsion substantially offsets the van der Waals forces and reduces the magnitude of the interaction potential minimum. The minimum also occurs at larger separation

distances (e.g., ~ 15 nm for pH=12 in figure 4.8b as opposed to only ~ 7.5 nm in Figure 4.8f).

In contrast, at the iso-electric point, where the net charge of the fragment is almost zero, very little electrostatic repulsion exists. The adsorption of fragmented protein can only counteract the van der Waals force by providing steric repulsion. However, steric repulsion is a shorter-range force, particularly for smaller fragments. The effect of the van der Waals force can only be fully opposed at closer distances where the interfacial adsorbed layers start to overlap. Thus, the magnitude of the interaction potential minimum will become less dependent on pH, as can be seen in figure 4.8f. The slight differences between the depths of minima are probably due to the slightly different amounts of adsorbed protein on the surfaces. As we saw from the results of section 4.1, the calculated adsorbed amount of the fragments was also positively correlated with salt volume fractions, especially at high salt volume fractions ($\phi_{salt} = 0.05$ and $\phi_{salt} = 0.1$). Thus at higher salt volume fractions, the increased adsorption on the surface seems to provide a slightly higher level of steric repulsion (due to a thicker adsorption layer). Nonetheless, this welcomed the additional increase in steric repulsion is more than compensated by the loss of electrostatic interactions due to screening. Overall then, at pH values away from the iso-electric point, the impact of higher salt is an undesirable increase in the depth of the minimum (figure 4.9c). Closer to the iso-electric point, where electrostatic interactions exist but are weaker, the opposing influence of salt on the two components of repulsion (namely steric and electrostatic) can lead to the somewhat counterintuitive situation where the depth of the minimum may slightly decrease with more background electrolyte in the system (see Figure 4.9a). For pH=6.08 (pI), where electrostatic repulsion no longer plays any role, the increase in steric repulsion due to the addition of more salt becomes clearer to see. Nonetheless, while the increase is desirable, it is also observed from the calculated data that steric repulsion still remains insufficient by itself to provide acceptable stability (i.e., the depth of minima is larger than that required for a good stable emulsion). This can be seen in Figure 4.9b. For the case far from the iso-electric point (pH=12), although the electrostatic repulsion is also affected by the screening effect, the net charges of the fragment at this point are far larger. Therefore, the impact of a higher salt volume fraction on any reduction in the electrostatic repulsion that these charges provide is more significant than any possible alteration in the steric interaction (Figure 4.9c).

It is seen from the results above that the effect of pH and salt volume fraction on the interaction potential induced by our conglycinin-derived protein fragments is multi-dimensional and depends on changes in both steric and electrostatic repulsion. These, in turn, depend on the specific primary structures of the fragments. Similar to the previous section, here we have summarised the effect of pH and salt volume fraction on the steric and electrostatic repulsion in Figure 4.11.

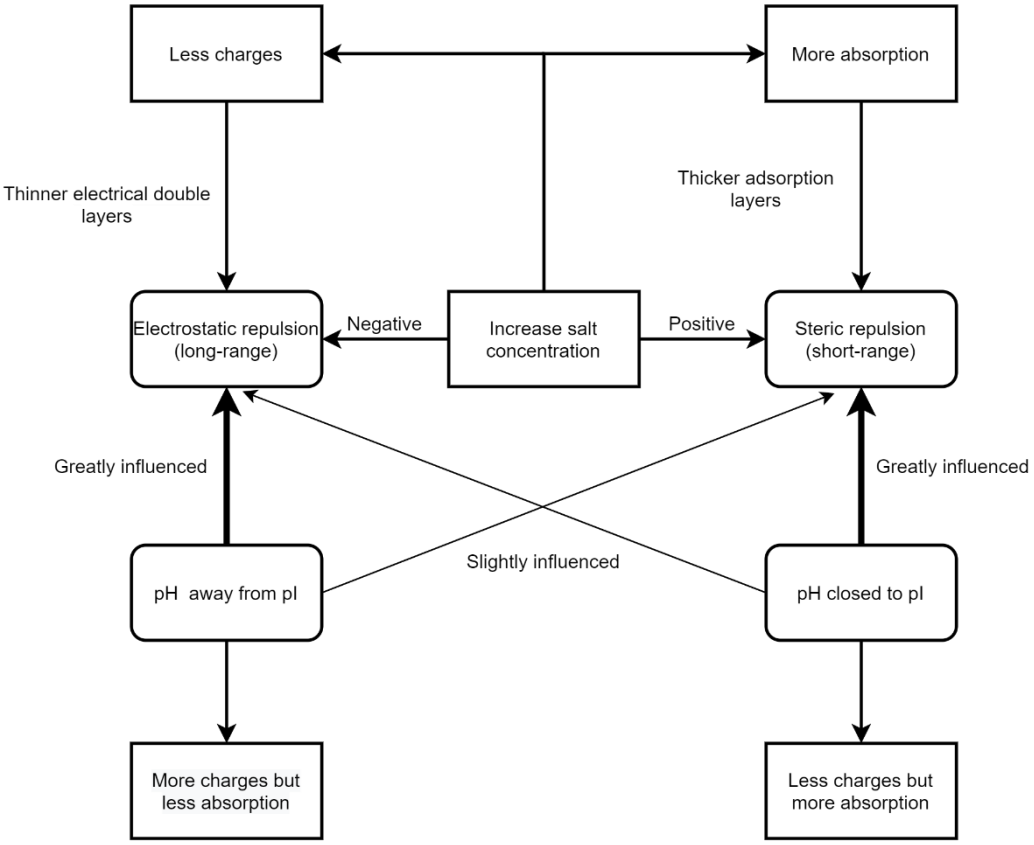


Figure 4.11 A summary of the effect of pH and salt volume fraction on the steric and electrostatic repulsion induced between a pair of droplets covered by fragmented polypeptide layers

4.4.2 Comparison of the di-block-like fragment with other proteins

In Chapter 3, upon the application of the fast screening methods, we selected a di-block-like fragment but also a multi-block fragment for comparison, both derived from the fragmentation of β-conglycinin. In particular, the multi-block fragment was chosen for comparison as an example of a polypeptide that was not considered to have the potential to be

a good emulsifier, given that it possessed too many small blocks. Figure 4.12 below illustrates the interaction potential mediated by the adsorbed layers of these two different fragments, as well as those calculated for α_{s1}-casein and β-casein, each at their corresponding iso-electric points. The results are at a salt volume fraction of 0.001.

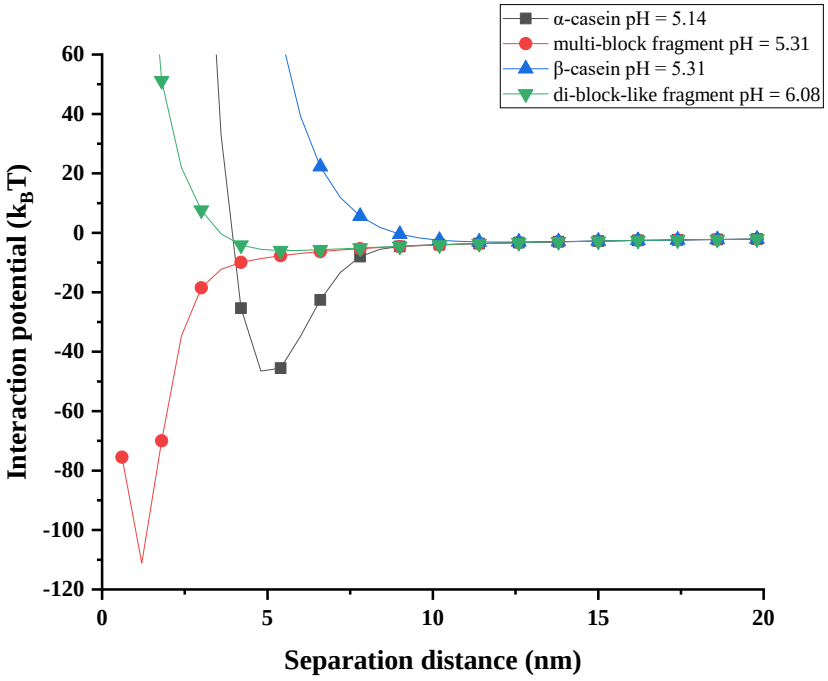


Figure 4.12 Comparison of the interaction potentials produced by the overlap of interfacial layers of the adsorbed di-block-like b-conglycinin derived fragments in comparison to several other proteins, each at their iso-electric points and with $\phi_{salt} = 0.001$

Figure 4.12 shows that, in terms of the potentials with the shallowest energy minima, the curves for the β-casein and the β-conglycinin-derived di-block-like fragment is similar. This may be the case because β-casein, like our chosen fragment, both have amino acid sequences that broadly resemble a di-block-like structure. As a result, the two fragments may be able to offer similar steric repulsions. In addition, the electrostatic repulsion does not influence the interaction potential too much at the iso-electric point for each protein, so the steric repulsion can act as the principal barrier for opposing the attractive van der Waals force between the droplets. Of course, the figure still demonstrates that β-casein, which produces a thicker interfacial layer, can create repulsive interactions at longer separation distances (8.4 nm for β-casein compared to 3 nm for the di-block-like fragment). As such, β-

casein remains a superior emulsion stabiliser than this chosen β-conglycinin-derived di-block-like fragment. The size of the β-casein in terms of its number of amino acids (N=209 compared to N=100 for the di-block-like fragment) may be a factor in explaining this outcome. The important role of the molecular weight of the polypeptide fragment in stabilising emulsions will be discussed further in Chapter 6.

Compared to β-casein and the di-block-like fragment, α_{S1}-casein possesses a larger minimum size of the interaction potential (-46.51 k_BT). This is also apparent: α_{S1}-casein has a tri-block-like structure, so α_{S1}-casein can only provide limited steric repulsion at the iso-electric point compared to di-block-like structures. This result is also consistent with the experimental as well as reported theoretical results that α_{S1}-casein does not make as good an emulsifier at its iso-electric point as β-casein (Cayot et al., 1991; Leermakers et al., 1996; Pinfield et al., 1997).

Finally, let us take a look at the results obtained for the multi-block protein fragment. In Figure 4.11, the value of the interaction potential for this selected fragment remains negative at all calculated distances. This means that the adsorbed layers of the multi-block protein fragment not only do not give rise to repulsive forces, but they induce attraction between droplets at all distances. As this fragment carries too many short blocks along its backbone, it was predicted not to be able to form sufficiently extended layers when it adsorbs to the surface of the droplets. Thus, the steric repulsive force it can provide would only arise at short-ranged separations, at distances where the van der Waals interactions are already strong. Furthermore, due to their multi-block nature, such fragments can simultaneously adsorb onto the surface of two nearby droplets. This leads to the possibility of undesirable bridging attraction between the droplets. This SCF calculation result somewhat validates the predictions of our fast screening methods. It confirms that, indeed, the fragments identified to have poor emulsification properties by both fast screening methods also do not seem capable of providing sufficient steric repulsion, according to the more detailed SCF calculations.

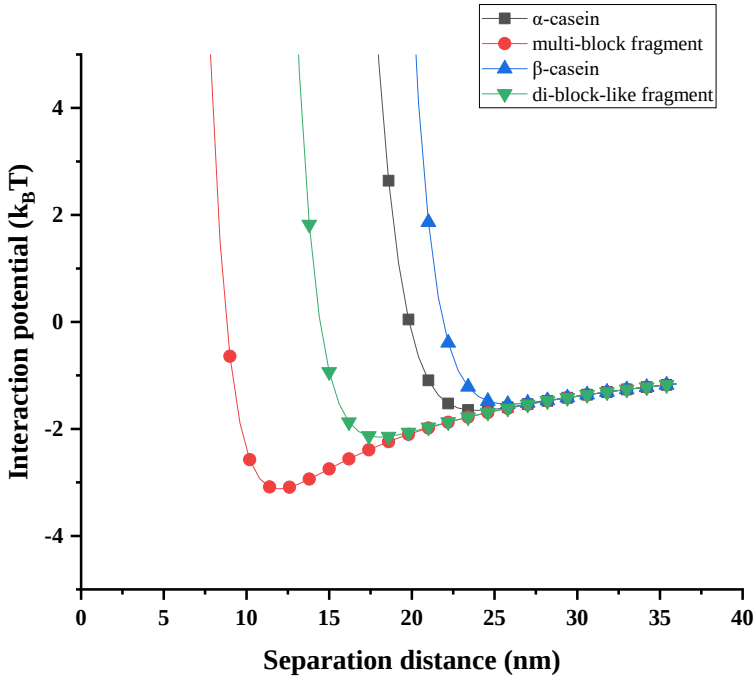


Figure 4.13 Comparison of the interaction potential produced by the overlap of the two conglycinin-derived polypeptides (one di-block-like and the other multi-block) with two other milk proteins at pH = 10 at a pH that is far from the iso-electric point of all the considered proteins. The background salt volume fraction in all cases was $\phi_{salt} = 0.001$

However, as shown in Figure 4.13, when we move the pH away from the iso-electric point of the proteins, all four curves, including the multi-block fragment, lead to droplet-droplet interaction potentials that have a minimum no deeper than 5 k_BT. This means that all four different fragments at this pH can stabilise the droplets and have reasonable emulsification performances. Nonetheless, it must be noted that the multi-block fragment still has the deepest minimum and, thus, the likely worst emulsification performance of all the four studied proteins. This result is a good illustration of the inadequacy of the fast screening methods which attempt to identify good steric stabilisers, but not necessarily electrostatic ones. The fast screening methods only focus on the possibility of the primary structure of fragments that can lead to emulsifiers with good provision of steric repulsion. They completely ignore any effect of electrostatic repulsion arising from the charge of the polypeptide on the emulsification performance of the fragments. Therefore, the use of more detailed SCF calculations post-fast screening remains crucial.

4.5 Conclusion and discussion

In this chapter, we employed SCF calculations to further analyse the various surface adsorption properties of the di-block-like polypeptide, obtained from the fragmentation of β-conglycinin and identified in chapter 3 to have the potential to be a good emulsifier by our fast screening methods. The performance of this fragment was also compared with other proteins at the end of this chapter. The SCF calculations show that this di-block-like fragment theoretically does have the potential to be a good emulsifier. One of the future extensions of this work could be to try experimentally to isolate and purify this particular fragment from the intact protein. While at present, the task of isolating one fragment from amongst many remains a challenging task, in Chapters 5 and 6, we will discuss the experimentally observed adsorption and emulsification properties of mixtures of various fragments separated along their molecular weight.

Furthermore, we have demonstrated that theoretical SCF calculation methodology can be a useful tool for identifying and choosing appropriate protein fragments as emulsifiers and stabilisers. This saves time and allows future experimental research to be focused on such identified fragments only.

It is important to also mention the possible limitations of SCF calculations. For example, while performing our calculations, we did not factor in a particular interaction between amino acids, such as the formation of covalent disulfide bonds or the secondary structures (polypeptides were assumed to be completely denatured during the fragmentation process). It is challenging to completely denature an intact protein, such as the soybean protein investigated in our study, having a native globular form, into an ideal denatured structure in practice. In order to make SCF calculations more accurate and closer to reality, we can try to further improve the method in future studies.

The SCF calculation considered in this chapter does, however, highlight some drawbacks of the fast screening methods. Before performing SCF calculations, the fast screening methods might save us much time, but at a cost: it may overstate or underestimate the potential of a certain class of fragments. Additionally, even with the aid of the quick screening procedure, we are still unable to exhaust all conceivable combinations due to a very

large number of combinations of fragments that may arise from limited fragmentation of an entire protein chain (see the predictions in Chapter 2). Even slight variations between fragments can produce drastically different outcomes in SCF calculations, particularly when the inclusion or absence of one or two charged amino acid residues is involved. For instance, we tried to remove the first two hydrophobic amino acid residues from our β–conglycinin-produced di-block fragment. This was to improve its resemblance to a di-block-like structure even more. However, the results were rather unsatisfactory compared to the initial di-block-like fragment we selected in the SCF calculation. As summarised in Figure 4.11, the emulsification performance of a fragment is greatly influenced by its structure, deviation of pH from its pI and background salt volume fraction. Before SCF calculations, it can be challenging to forecast how the interaction potential induced by a particular fragment would behave at various droplet-droplet separation distances. In the future, we might be able to try to create a quick screening approach based on SCF calculations utilising machine learning techniques in order to resolve this issue. One strategy that could prove useful is to build on methodologies demonstrated presented in this and previous chapters to go further and create a predictive model that aims for the minimum value of interaction potential. First, we need to perform SCF calculations on many random sequences used to build test and train datasets. We can then generalise the data features to the actual situation, which in this case may refer to pH, salt volume fraction, size and number of amino acid residues of the protein chain (or polypeptide fragments derived from proteins), size of the longest hydrophilic/hydrophobic blocks, and the proportion of hydrophobic amino acids. Finally, the data are fitted and modelled using standard machine learning algorithms (i.e. regression, decision trees and neural networks) to provide an artificially intelligent system trained to spot and screen the best possible polypeptide fragments as food-grade emulsifiers under the broadest range of conditions.

List of References

Cayot, P., Courthaudon, J. L. and Lorient, D. 1991. Emulsifying properties of pure and mixed .alpha.s1- and .beta.-casein fractions: effects of chemical glycosylation. Journal of Agricultural and Food Chemistry. 39(8), pp.1369-1373.

Dickinson, E. 1992. Introduction to food colloids. Oxford university press.

Leermakers, F. A., Atkinson, P. J., Dickinson, E. and Horne, D. S. 1996. Self-consistent-field modeling of adsorbed β-casein: effects of pH and ionic strength on surface coverage and density profile. Journal of Colloid Interface Science. 178(2), pp.681-693.

Pinfield, V., Horne, D. S. and Leermakers, F. M. 1997. Self-consistent-field modelling of adsorbed casein interaction between two protein-coated surfaces. Journal of the Chemical Society, Faraday Transactions. 93(9), pp.1785-1790.

Chapter 5
Preliminary Experiment: Measurements of Surface Tension

5.1 Introduction

We conducted a theoretical study in Chapters 3 and 4 to identify a protein fragment that may have advantageous emulsification characteristics. However, experimentally isolating and purifying a protein fragment with a particular primary sequence is challenging. We can explain this difficulty using the following theoretical calculations. Assume there are numerous enzymatic cleavage sites on the protein chain in Figure 5.1 below. The fragment we require is the sequence between the two blue enzyme cleavage sites in the figure. We want the enzyme to break only the blue peptide bonds while retaining the middle k red peptide bonds.

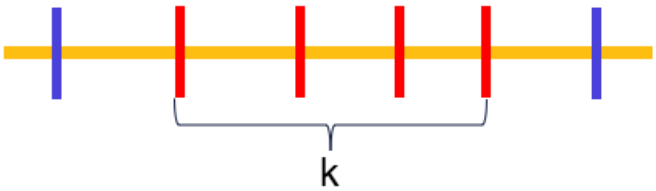


Figure 5.1 Different enzymatic cleavage sites on the peptide chain

Here it is assumed that the probability of the protease reacting with each cleavage site is p. The probability of breaking only the two blue peptide bonds and retaining the red peptide bond is $P = p^2(1-p)^k$. This probability can also be interpreted as the proportion of a particular fragment to the total hydrolysis product η, which is related to p, as shown in Figure 5.2 below.

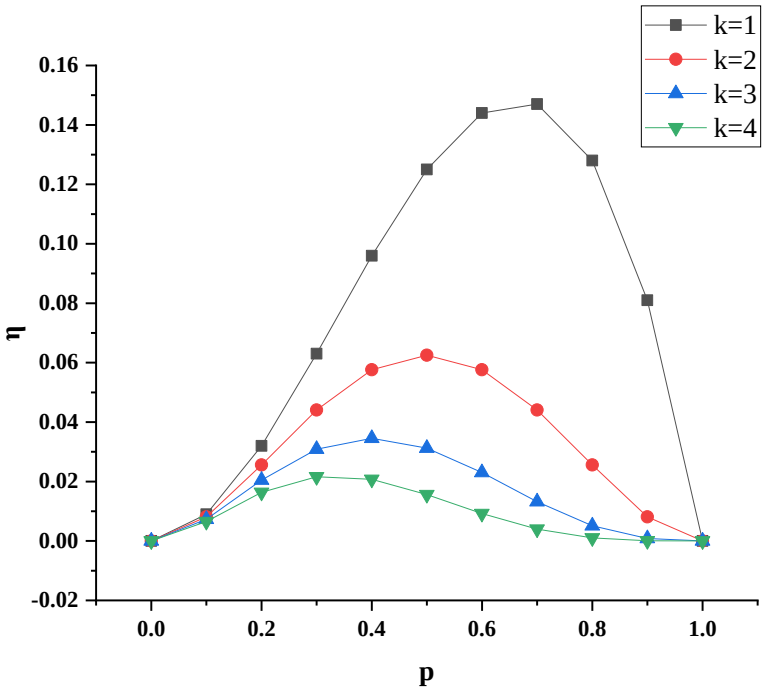


Figure 5.2 The relationship between the probability of each bond breaking p and for different values of k

Then, we can take $\frac{\partial P}{\partial p} = \frac{\partial p^2 * (1-p)^k}{\partial p} = 0 \rightarrow p = \frac{1}{1+\frac{k}{2}}$ to obtain the maximum value of P. Therefore, the optimal degree of hydrolysis can be calculated as $DH_{optimum} = \frac{p \cdot N}{N-1}$ (N is the total number of residues, N-1 is the number of peptide bonds the enzyme can react with), which is displayed in Figure 5.3.

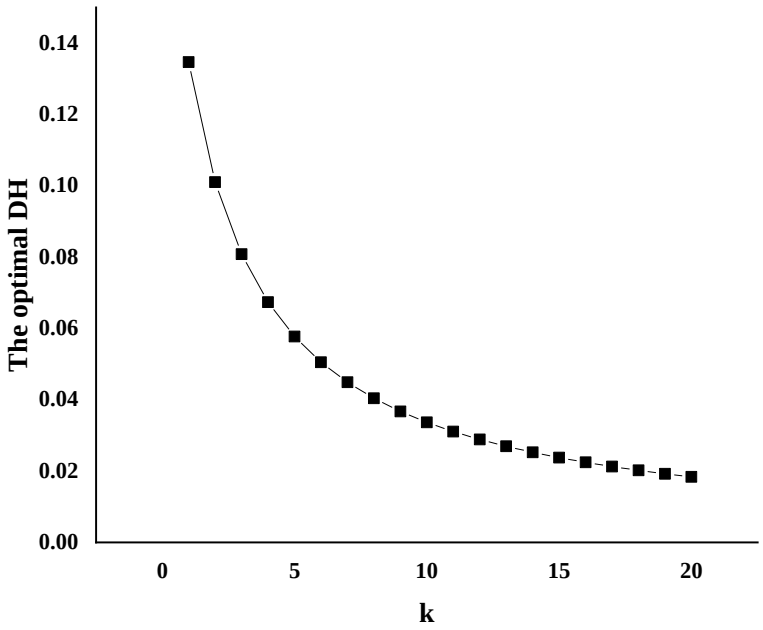


Figure 5.3 Variation of the optimal hydrolysis pH values with k

As can be seen from Figure 5.2, the proportion of specific fragments in the hydrolysis product is small (less than 0.15) even at k=2. In addition, Figure 5.2 shows that to achieve the maximum yield at k=2, we need to control the degree of hydrolysis to be around 13%, which is also a relatively low degree of hydrolysis. However, in reality, k is usually greater than 2. For example, suppose we assume we hydrolyse the β-conglycinin-α-subunit with trypsin to obtain the fragment we studied in the previous section. In that case, we can obtain the following figure.

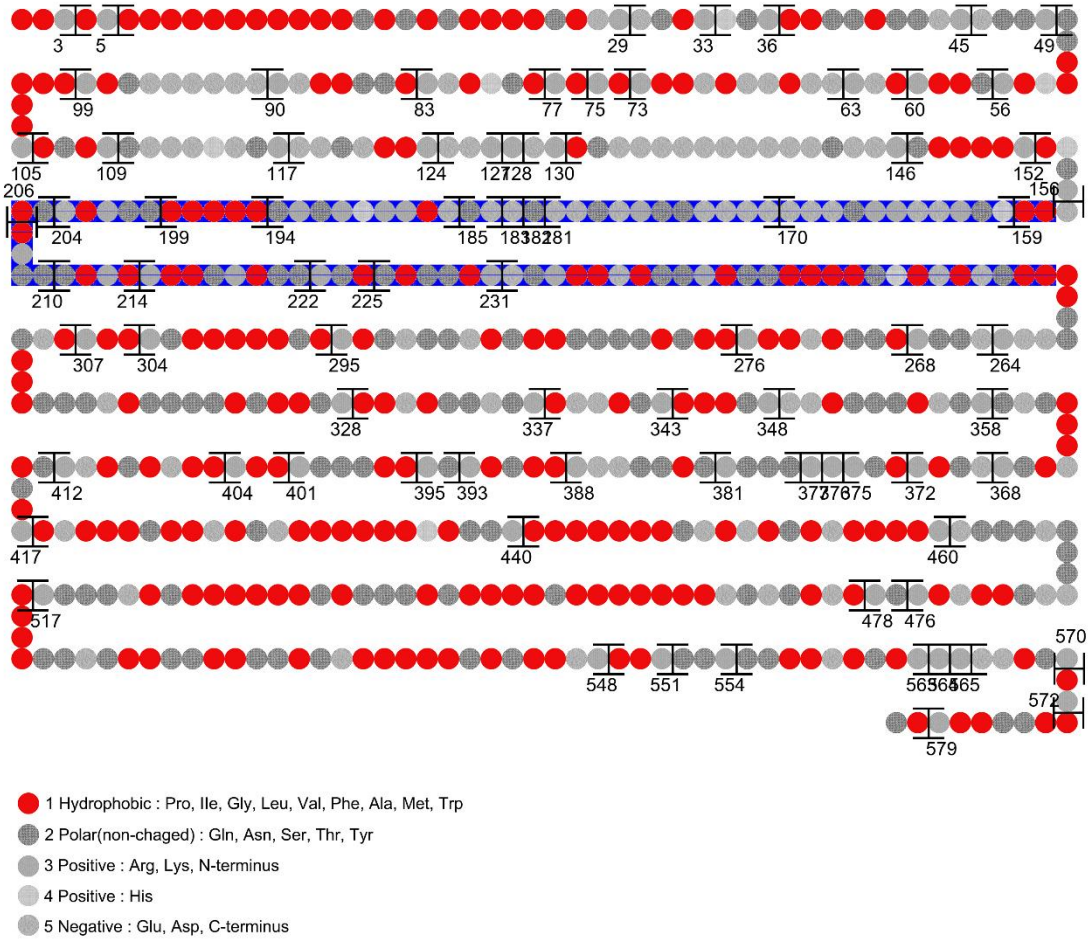


Figure 5.4 The enzymatic cleavage sites of trypsin on the β-conglycinin-α-subunit (the blue fragment represents the di-block-like fragment in Chapters 3 and 4)

In Figure 5.4, the di-block-like fragment corresponds to an enormous k value of 15. From the above calculations, it is clear that trying to obtain a single specific fragment directly in the experiment is not realistic. This is because it requires a low degree of hydrolysis, which is very difficult to achieve in experiments. Furthermore, even at the optimum hydrolysis degree, the proportion of specific fragments to the total hydrolysis product is still reasonably low, making isolation and purification considerably more problematic.

As a compromise, we will now attempt to create peptide-based emulsifiers and examine their varied colloidal characteristics experimentally. Chapter 5 can be seen as a pre-experiment for the experiments in Chapter 6 on the preparation of peptide-stabilised emulsions. This chapter examines the impact of peptide molecular weight on emulsification characteristics and the variations in surface tension among various protein hydrolysates.

Surface tension γ reflects the cohesive force between molecules and is responsible for the trend of a liquid surface shrinking to its minimum area when in contact with an incompatible phase. (Note the term surface tension is often used to refer to the tension at the surface of a liquid in contact with air, whereas interfacial tension is a more general term that tends to be used for the tension between the material and any other phase (solid, liquid or gas) but also sometimes the air-material interface). Surface tension is equivalent to the Gibbs free excess energy per unit area of the interface, but therefore also has units of force per unit length (N/m or mN/m) and measuring the tensile force per unit length in various ways is the principal way in which γ is actually measured. Anything which preferentially occupies the interface and so reduces the area of contact between the phases will lower γ and produce a thermodynamically lower energy state, and this is what surfactants do – surface active agents. Emulsifiers are, therefore, also surfactants – they adsorb and lower γ the oil-water interface. This makes it easier to form emulsions, but then the emulsifiers generally play a secondary role in maintaining the emulsified and higher area state by preventing flocculation and coalescence of the droplets by keeping them far enough apart via electrostatic and or steric interactions. An emulsifier that lowers γ more will also allow for forming more surfaces and, therefore, smaller droplets. In order to investigate the effect of enzymatic treatment in reducing the surface tension of protein solutions, this study measured the surface tension of different soya protein hydrolysates (SPH) samples and their corresponding original protein using the method highlighted in Chapter 2.

5.2 Methods

The surface tension γ of various *SPH* solutions (10^{-3} wt%) in 0.05 M phosphate buffer at pH 7 was measured via an Ez-Pi Plus Kibron tensiometer (Kibron Inc, Finland). Measurements were taken every 6 s at room temperature. The calibration reference was the γ of pure water at 20 °C 20 – taken as 72.8 mN m^{-1}. We collected the data until the γ became constant (± 0.3 mN m^{-1}), which was after no more than 10 min.

5.3 Effect of molecular weight on protein emulsification properties

5.3.1 Predicted average molecular weight change as a function of hydrolysis conditions

As this study investigates the emulsification properties of fragmented proteins, it is very revealing to understand the Mw change of each fragment in the hydrolysates with different *D* values. The degree of hydrolysis, *D*, can be defined as the number of bonds cleaved by the protease divided by the total number of bonds on the protein backbone.

$$D = \frac{N_{broken}}{N_{total}} \tag{5.1}$$

Therefore, for a given *D* value, we can figure out the relative average Mw of fragments compared to the non-hydrolysed protein:

$$\frac{\left\langle MW_{\text{fragment}} \right\rangle}{MW_{\text{intact}}} = \frac{1}{D \times (N_i - 1) + 1} \tag{5.2}$$

Where N_i is the total number of amino acid residues of the intact protein; $\left\langle MW_{\text{fragment}} \right\rangle$ is the average Mw of each small fragment in the hydrolysate; MW_{intact} is the Mw of the intact protein.

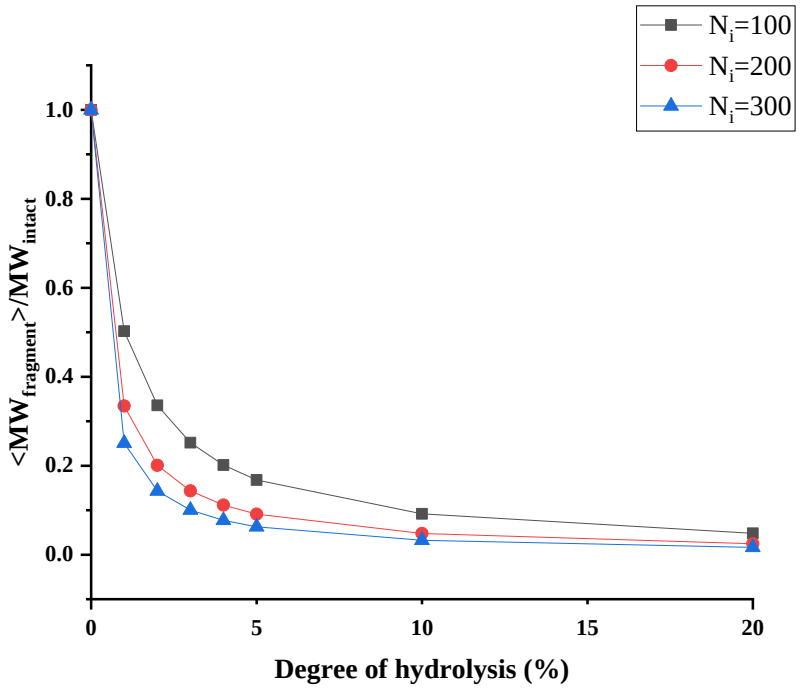


Figure 5.5 Calculated average Mw of hydrolysates relative to that of the intact polymers, plotted against the degree of hydrolysis (*DH*)

As hinted at in Figure 5.5, the average Mw declines rapidly with *DH* and roughly as 1/*DH* for chains with a large degree of polymerisation. This is most notable in the *DH* interval from 0% to 10%. Take N_i=100 as an example; when the *DH* value increases to 10% and a further 20%, the average Mw of the resulting fragments drops approximately to 0.09 and 0.04 of the original intact chains, respectively. This means that most of the fragments in such hydrolysate consist of short chains – with less than ten residues. Moreover, for larger protein chains, the average Mw will decline even more rapidly with *DH* (see the curves for N_i=200 and N_i=300 in Figure 5.5). On average, and irrespective of its size, the Mw of a chain drops by approximately a factor of 2 each time a bond on its backbone is broken. However, for a given value of *DH,* the number of broken bonds scales with the size of the chain. Therefore, it is clear that for larger chains, the drop will be more significant for the same DH. While this may seem an obvious point to make, it is emphasised here due to its practical importance. In our experiments, it is hoped that the fragmented proteins will provide enough steric and electrostatic repulsion to stabilise the O/W emulsion droplets. Such fragments must contain a relatively long hydrophilic block to form a thick enough protruding layer and a hydrophobic block long enough to adsorb strongly at the oil-water interface. On the other hand, electrostatic

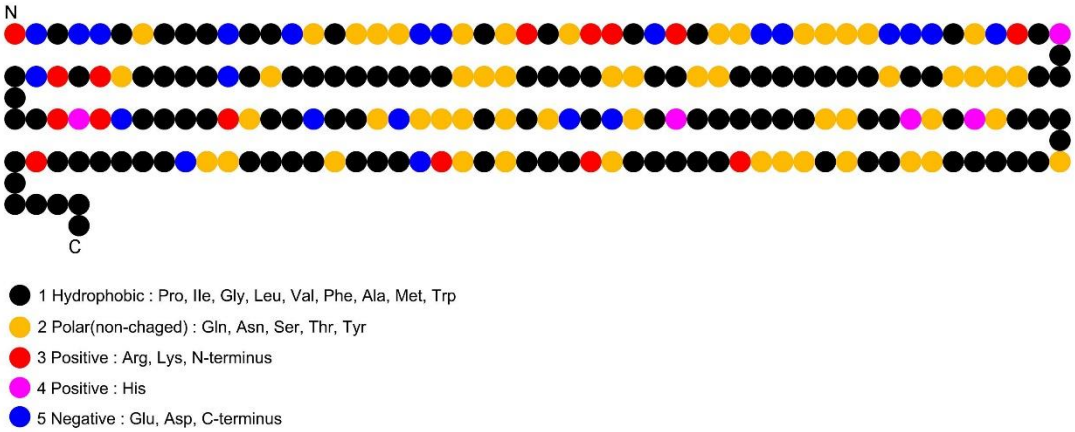
repulsion is partly dominated by the magnitude of charges close to the droplet surfaces, according to DLVO theory (Dickinson, 1992b). Nevertheless, it is clear that limited hydrolysis is preferred; otherwise, the vast majority of the peptides will be too short of providing sufficient electrostatic stabilisation. This is why we deliberately chose a range of pH values during hydrolysis that will have different efficiencies of hydrolysis and also used filtration to spate out a range of different M_w for each hydrolysis condition.

Our experiments considered three different pH values (1.3, 2.1 and 4.7) to hydrolyse soybean protein by pepsin because pepsin's specificity and activity are significantly different at these pH values. Pepsin is more specific but with relatively low activity at pH 1.3. The specificity tends to be lost when the pH is higher than 2. At pH 2.1, pepsin obtains the maximum activity with broad specificity. We also chose a sub-optimal value of pH 4.7, where we expect broader specificity but only moderate activity (Piper and Fenton, 1965) to generate more varied but also perhaps longer fragments.

5.3.2 Calculation of the interaction potential

Section 5.3.1 above stresses the importance of *DH* in determining the M_w of the fragments obtained. In this section, we use SCF calculation to illustrate the effect of M_w on colloidal interaction forces.

Much previous research has shown that the milk protein β-caseinate displays excellent emulsification properties in practice (Dickinson, 1997; Dickinson et al., 1998; Dickinson, 2001). Moreover, β-caseinate has been considered one of the few natural proteins with an approximately di-block-type structure (Dickinson, 2003). Figure 5.6 shows that β-caseinate consists of a predominantly hydrophilic N-terminus side and a mostly hydrophobic C-terminus end. Therefore, even without the contribution of electrostatic repulsion from charged amino acid residues, the di-block structure of β-caseinate is expected to induce a reasonable steric repulsion between droplets. Though probably not quite enough to stabilise the emulsions by itself, the provision of such steric repulsion can greatly enhance the good emulsification ability of this protein.



RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHP	1 - 51
FAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAM	52 - 102
APKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPP	103 - 153
TVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGP	154 - 204
FPIIV	205 - 209

Figure 5.6 The primary structure of β-caseinate with the constituent amino acids grouped into five different groups as indicated

However, it should also be noted that a di-block-like structure is not the only important criterion for the provision of strong repulsive forces and, thus, for colloidal stabilisation ability. As mentioned in section 5.3.1, the length of adsorbed fragments is another critical factor. This point can be illustrated more clearly using the theoretical calculation for the induced inter-droplet interaction potentials between β-caseinate-covered emulsions and those involving adsorbed layers of an ideal short di-block fragment (see Figure 5.7). The term ideal here is used to refer to a fragment that consists of only two consecutive purely hydrophilic and solely hydrophobic blocks (Murray et al., 2021).

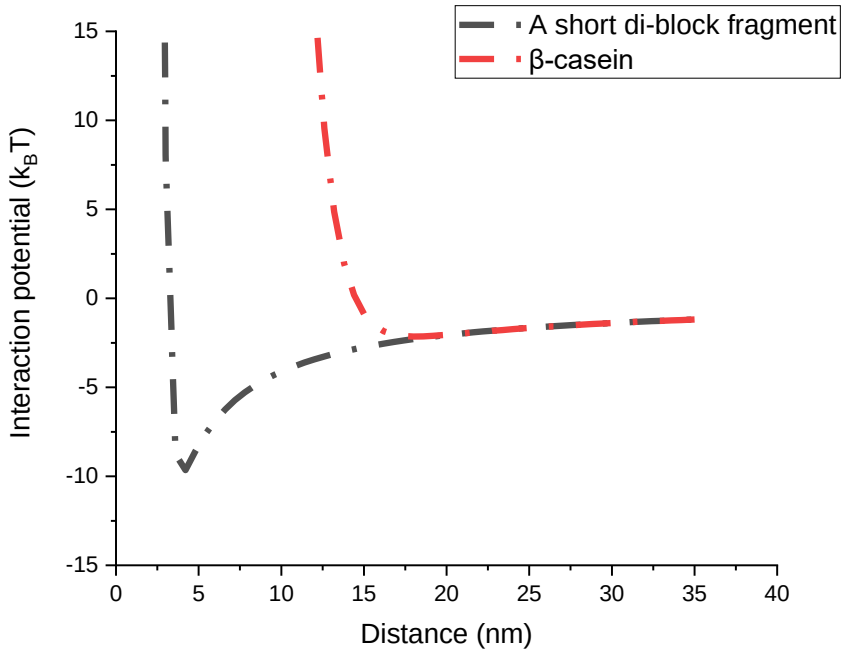


Figure 5.7 Comparison of theoretical interaction potentials plotted vs separation distance, as obtained from SCF calculations, between two droplets covered with β-caseinate and with a short di-block fragment. The results also include direct van der Waals interactions operating between the droplets. The graphs were obtained at pH=7 of β-caseinate, with the volume fraction of salt ions at 0.001 (~ 0.01 mol/l) and for droplets of size 1μm

At pH = 7, β-caseinate has a net charge of -6.14e. We make the charge density carried by our short fragment comparable with that for β-caseinate. This is done by setting the total charge q_{short} of the hypostatical fragment equal to

$$q_{short} = q_{\beta-casein}(N_{short} / N_{\beta-casein}) = (-6.14e)(15/209) = -0.44e\,.$$

Figure 5.7 shows the interaction potentials between two polymer-coated droplets, where in one case, the polymer in question is β-caseinate and, in the other, the ideal short-sized di-block fragment. Here interaction potential is expressed in the unit of k_BT. As mentioned in Chapter 2 (Deryagin approximation), its value between two droplets is obtained from the potential calculated between two flat parallel plates, converted to that for spheres with the aid of equation (2.24). Decreasing values of the interaction potential with increasing separation indicate a repulsion between two droplets. Generally, flocculation happens due to a weak attraction between droplets when the interaction potential is more negative than around $-5k_BT$. At less negative values, thermal agitation and Brownian motion are sufficient to separate two

weakly aggregated droplets. Nevertheless, as the attraction increases, flocculation will further develop and cause stronger irreversible aggregates that can lead to coalescence. In Figure 5.7, apart from forces mediated by polymer layers, we have also included the more direct van der Waals forces. These were given by Equation 2.25.

The primary structure of β-caseinate was taken from the previous experimental paper (Farrell et al., 2004). The short di-block fragment comprises a hydrophobic domain with five hydrophobic residues and a hydrophilic section with 15 non-charged polar residues. Though not entirely impossible, this fragment is not very likely to arise from the fragmentation of any common food proteins. However, our calculations aim to compare the interaction potential of these two di-block-like macromolecules with quite different sizes (where β-caseinate consists of 209 amino acid residues).

As shown in Figure 5.7, the minimum value of the interaction potential curve induced by adsorbed layers of β-caseinate is found to be around -2.0 k_BT. This occurs at a droplet separation distance of ~ 18 nm. The interaction potential curve drops rapidly in the range of 0 to 18 nm but with the interaction potential still higher than zero at r < 14.4nm. Therefore, according to these calculations, β-caseinate produces a repulsive effect at a suitably long range of separation distances. This explains why β-caseinate exhibits excellent emulsion stability properties, as also found in many previous theoretical studies (Dickinson, 1992a; Pinfield et al., 1997), especially at pH values away from its isoelectric point. As for the curve obtained for the short di-block fragment, we found that the magnitude of the minimum well in the inter-droplet interaction potential was 9.64 k_BT for the short fragment. This is much deeper than that obtained for β-caseinate. Moreover, the minimum value occurs at a substantially closer distance between the droplets, at r=4.2nm. In this case of stabilisation by short fragments, the van der Waals forces seem to dominate any repulsion produced by the overlap of polymer layers; thus, the emulsion prepared with our hypothetical ideal short di-block fragment is predicted not to be colloidally stable.

From the above calculation, we can compare the surface behaviour of two roughly similar “di-block” chains with different lengths. For β-caseinate, we found a marked improvement in the provision of induced repulsive interactions as compared to the short di-block fragment. This result

emphasises the fact that any peptide-based emulsifier should contain fragments comprising a sufficiently large number of amino acid residues, even when possessing the most favourable structure, i.e. a di-block-like fragment.

5.4 Measurement of surface tensions of caseinate and soybean protein hydrolysates

The degree of suppression of γ at the air-water (A–W) interface of a surfactant solution is also usually a reasonable indicator of its surface activity at an oil-water (O–W) interface. After hydrolysis and filtration etc., to obtain the various M_w fractions, the amounts of sample available were limited, and because of this, surface tension γ is far easier to measure reliably at the A–W rather than the O–W interface; we opted to measure γ at the A–W interface.

A summary of the final values of γ (i.e., after 10 min – see Methods section) is shown in Table 5.1, measured at 10^{-3} wt.% protein to hopefully accentuate any differences between the samples, rather than the higher concentration used to prepare the emulsions (see later), where values tend to converge more at the short adsorption times relevant to emulsion formation. The values of γ given are the average values of 20 measurements every 6 seconds after 10 min, by which time plateau values appeared to have been reached. Appendix 2 data gives further data on γ versus time, but since the *SCFC* cannot give kinetic, but only equilibrium data, only the final values of γ are relevant here to compare the surface activity of the various hydrolysates and the intact proteins from which they were derived.

Table 5.1 Average surface tensions γ at 10^{-3} wt.% protein and pH 7 of different *SPH* samples produced by 3 different enzyme treatments

	Caseinate hydrolystes			
	Molecular weight (kDa)			
	<3	3~10	>10	whole

Alcalase pH 8	63.00 ± 0.62	69.00 ± 1.36	70.00 ± 3.85	70.00 ± 2.38
Pepsin pH 1.3	59.00 ± 2.45	51.00 ± 3.07	61.00 ± 1.09	62.00 ± 2.77
Pepsin pH 2.1	72.00 ± 1.07	54.00 ± 3.95	68.0 0± 0.68	67.00 ± 1.02

	SPI hydrolystes			
	Molecular weight (kDa)			
	<3	3~10	>10	whole
Alcalase pH 8	65.00 ± 4.24	59.00 ± 5.72	56.00 ± 1.05	60.00 ± 2.33
Pepsin pH 1.3	49.00 ± 2.11	50.00 ± 1.34	58.00 ± 1.49	60.00 ± 1.16
Pepsin pH 2.1	69.00 ± 2.75	54.00 ± 2.03	66.00 ± 2.00	67.00 ± 1.64
Pepsin pH 4.7			52.00 ± 0.70	

The ‘whole’ values refer to the measurements on whole hydrolysates, i.e., before fractionating in the 3 different Mw ranges. The corresponding for non-hydrolysed soy protein and sodium caseinate were 71±2.80 and 55±0.39 mN m-1, respectively.

As observed in Table 5.1, none of the caseinate hydrolysates produced via alcalase- or pepsin-treatment gave a lower γ than the original (non-hydrolysed) caseinate solution (γ = 55 mN m^{-1}). Except for the pepsin-treated samples with an M_w of 3-10 kDa, which had a slightly reduced γ, the caseinate-based hydrolysates produced under the other conditions had a higher γ than the non-hydrolysed protein, i.e., were *less* surface active. This is probably because the hydrolysis destroyed the already ideal primary structure of the two main components of sodium caseinate: β- and α_{s1}-casein (ideal in terms of block copolymer structure and surface activity). In other words, there is probably little advantage in hydrolysing the sodium caseinate to smaller protein fragments since caseinate is already an excellent emulsifier. In theory, fragments of the constituent proteins might be slightly better, but to produce them would require much more selective hydrolysis than used here and finding enzymes that could do this is probably quite

challenging. Even if some suitable small fragments did exist in the samples (such in the 3-10 kDa fraction), the presence of higher concentrations of other less favourable fragments might prevent the more surface active fragments from adsorbing.

In contrast to the caseinate, all the soybean protein hydrolysates had a lower γ (i.e., were more surface active) than the non-hydrolysed *SPI* (γ = 71 mN m^{-1}). Indeed, under the conditions of measurement used, the highly aggregated and relatively insoluble *SPI* seemed to be hard surface active at all. The increase in surface activity on hydrolysis must therefore be attributed to increased solubilisation and/or the production of smaller soy protein polypeptides that have a greater tendency for adsorption, which therefore might be better emulsifiers. The γ results suggest that a higher M_w fraction (M_w > 10 kDa) produced by pepsin treatment at pH 2.1 might be the best emulsifier since this has the lowest γ (43 mN m-1) of all the *SPH*, lower even than γ for caseinate under these conditions. The use of pH 2.1 is significant because this is the optimum pH for the action of this enzyme (Piper and Fenton, 1965; Keil, 1992). The lowest γ at pH 2.1 was closely followed by low (M_w < 3 KDa) and medium (M_w 3 – 10 kDa) *SPH* produced by pepsin at pH 1.3, which gave similar γ (49 and 50 mN m^{-1}, respectively), again lower than the value for non-hydrolysed caseinate. Although these low γ values at this bulk concentration indicate strong adsorption, as explained earlier, a low γ does not necessarily guarantee good colloidal stabilisation. As well as being strongly adsorbed to the interface, the material must provide sufficient steric and/or electrostatic stabilisation via sufficiently long chains produced in the aqueous phase and/or chains that provide a sufficiently high charge density at the oil droplet surface. The more surface active peptides might be too in M_w to provide either, so further experiments were required to test this, as reported in Chapter 6.

5.5 Conclusion and discussion

In this chapter, the theoretical calculations show that the molecular weight of a protein can significantly affect its emulsification properties. Even if a fragment has a relatively perfect di-block structure, it will not be a suitable emulsifier if the protein chain length is insufficient. This is mainly because the steric repulsion that a protein chain can provide depends not only on the

structure of the fragment but also on the number of amino acid residues that make up the peptide chain.

On the other hand, measurements of the surface tension of different protein hydrolysis products showed that pepsin-hydrolysed soy protein could significantly reduce the surface tension at the A-W interface compared to intact soy protein. As a result, the pepsin hydrolysate soy product may contain certain components with effective emulsification capabilities. The different colloidal characteristics of emulsions stabilised by soy protein hydrolysates will be further discussed in Chapter 6.

List of References

Dickinson, E. 1992a. Faraday research article. Structure and composition of adsorbed protein layers and the relationship to emulsion stability. 88(20), pp.2973-2983.

Dickinson, E. 1992b. Introduction to food colloids. Oxford university press.

Dickinson, E. 1997. Properties of emulsions stabilized with milk proteins: overview of some recent developments. Journal of Dairy Science. 80(10), pp.2607-2619.

Dickinson, E. 2001. Milk protein interfacial layers and the relationship to emulsion stability and rheology. Colloids and Surfaces B: Biointerfaces. 20(3), pp.197-210.

Dickinson, E. 2003. Interfacial, Emulsifying and Foaming Properties of Milk Proteins. In: Fox, P.F. and McSweeney, P.L.H. eds. Advanced Dairy Chemistry—1 Proteins: Part A / Part B. Boston, MA: Springer US, pp.1229-1260.

Dickinson, E., Semenova, M. G. and Antipova, A. S. 1998. Salt stability of casein emulsions. Food Hydrocolloids. 12(2), pp.227-235.

Farrell, H. M., Jimenez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., Hicks, C. L., Hollar, C. M., Ng-Kwai-Hang, K. F. and Swaisgood, H. E. 2004. Nomenclature of the Proteins of Cows' Milk—Sixth Revision. Journal of Dairy Science. 87(6), pp.1641-1674.

Keil, B. 1992. Essential substrate residues for action of endopeptidases. Specificity of proteolysis. Springer, pp.43-228.

Murray, B. S., Ettelaie, R., Sarkar, A., Mackie, A. R. and Dickinson, E. 2021. The perfect hydrocolloid stabilizer: Imagination versus reality. Food Hydrocolloids. 117, p106696.

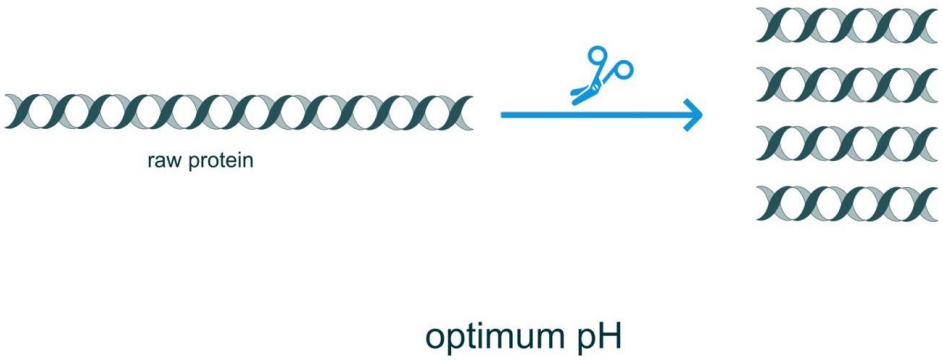
Pinfield, V., Horne, D. S. and Leermakers, F. M. 1997. Self-consistent-field modelling of adsorbed casein interaction between two protein-coated surfaces. Journal of the Chemical Society, Faraday Transactions. 93(9), pp.1785-1790.

Piper, D. and Fenton, B. H. 1965. pH stability and activity curves of pepsin with special reference to their clinical importance. Gut. 6(5), p506.

Chapter 6
Investigation of the Colloidal Properties of the Emulsion Stabilised by SPH

6.1 Introduction

Previous studies usually focus on the cleavage of proteins at the optimum pH of the enzyme (weak specificity and high enzymatic activity) with limited degree of hydrolysis. These approaches can also generate some fragments with suitable emulsification capacity (Agyare et al., 2009; Yu et al., 2018; Jin et al., 2020). However, a shortcoming of these approaches is that proteins are quickly and uniformly cleaved into similarly sized fragments. This means that the number of different permutations of polypeptides inside the hydrolysate is small. In addition, homogeneous cleavage tends to favour the generation of shorter fragments rather than longer ones. For example, assuming a protein chain has 100 (evenly distributed) enzymatic sites for cleavage, 10 peptide bonds will be opened at a 10% degree of hydrolysis. In this case, the most likely composition of the hydrolysis product we would obtain would be a set of fragments containing ten enzymatic cleavage sites (see Figure 6.1).



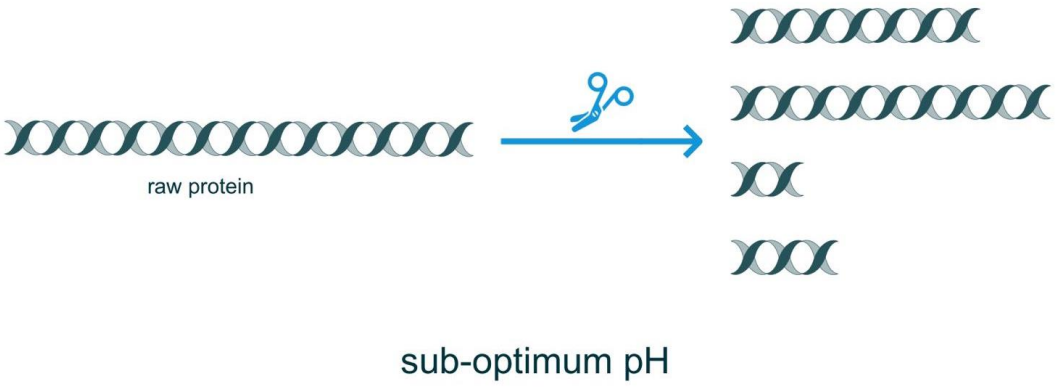


Figure 6.1 Comparison of the composition of protein hydrolysis products predicted to be obtained at optimal and sub-optimal hydrolysis pH

As can be seen in Figure 6.1, even at the same degree of hydrolysis, the optimum and sub-optimum pH values produce different fragment compositions; at the sub-optimal hydrolysis pH, the enzyme cleavage is more selective and tends to produce size polarised fragments. In this chapter, we first investigate the emulsification and stabilisation abilities of emulsions stabilised by SPHs. One of the SPH samples considered a suitable emulsifier was then sequenced and further analysed by similar SCF calculations as in Chapter 4.

6.2 Materials and methods

6.2.1 Materials

Commercial soybean protein isolate (*SPI*) powder with 90% purity was purchased from Pulsin, Gloucester, United Kingdom. Pepsin from porcine gastric mucosa ($\geqslant$250 units/mg), Micellar Casein with 83% purity was purchased from Bulk, London, United Kingdom. Alcalase (P4860), n-tetradecane, Ultrafiltration Discs (3 kDa and 10 kDa NMW), Picrylsulfonic acid solution 5 wt% (2,4,6-Trinitrobenzenesulfonic acid solution, TNBS) and all other chemicals were purchased from Sigma-Aldrich, Gillingham, United Kingdom.

6.2.2 Methods

6.2.2.1 Preparation of soybean protein hydrolysates (*SPH*)

The preparation method for soybean protein hydrolysates followed the method of (Han et al., 2021) with slight modifications. Commercial soybean protein powder (5wt%) was dissolved in Milli-Q water, and the pH was adjusted to 1.3, 2.1 and 4.7. Pepsin, at an enzyme-to-substrate ratio of 1: 25, was added to the solution and stirred for 15 min at room temperature. Then, the mixtures in a tube were immersed in a 37℃ water bath for 6 hours, followed by heating in a boiling water bath for a further 15 minutes to denature and deactivate the pepsin. When the mixtures were cooled back down to room temperature, any precipitates were removed by centrifugation (x4000g, for 20 min, at 4°C), with pH adjusted to 7. After centrifugation, the mixtures were filtered through 10 or 3kDa ultrafiltration membranes. Thus, a series of *SPH* samples with three different molecular weight (M_w) ranges: $\leqslant$ 3 kDa, 3 to 10 kDa and $\geqslant$ 10 kDa, were obtained (Yields for samples 2.1 and 4.7 (MW > 10kDa) were 1.45g/16g and 1.65g/16g respectively). Finally, the *SPH* samples were freeze-dried and stored at 4°C in a fridge for further use.

6.2.2.2 Preparation of emulsions stabilised by *SPH*

SPH-stabilized emulsions were made with 10% n-tetradecane and 90% aqueous solution containing 1% protein/peptides. The two phases were blended at room temperature (20 to 25 °C) using an IKA homogeniser (12,000 rpm, 10 min) to form coarse droplets, followed by 3 passes through the Leeds Jet Homogenizer (Burgaud et al., 1990) at 500 bar. The emulsions also had 0.02wt% sodium azide added as an antimicrobial agent.

6.2.2.3 Droplet size and zeta potential measurements

The particle size distribution (*PSD*) of protein solutions prior to emulsification and for the emulsions stabilised by the same protein solutions were measured via a Mastersizer 3000 (Malvern Panalytical Instrument, USA). All emulsions were shaken by hand to ensure homogenous sampling before the size measurements. The refractive index for soybean protein was taken as 1.45 (for the PSDs of the protein dispersions), whilst the refractive index of the oil was taken as 1.43. All zeta potentials were measured using a Zetasizer Nano ZS (Malvern Panalytical Instrument, USA) and calculated using Henry's equation with Smolukowski approximation for polar media.

The results reported are the averages of 3 separate measurements ± the standard deviation.

6.2.2.4 Degree of hydrolysis measurements

The degree of hydrolysis (*DH*) was estimated by a modified TNBS method (Han et al., 2021). The principle behind this method is the reaction between the TNBS reagent and N-terminal amino acid groups. In brief, concentrated TNBS reagent (5 wt.%) was diluted 25 times in 0.1M $NaHCO_3$ buffer at pH 8.5. Leucine solutions of a range of concentrations were used as the standard solutions for the calibration curve. In this study, 40 μl diluted TNBS was added to 80 μl protein solutions or the standard solutions at 37C for 2 h. Then, 80μl of 1 mol dm^{-3} HCl and 40 μl of 10 wt.% SDS buffer solution was added to the samples to terminate the hydrolysis reaction. Finally, the absorbance of each sample was measured at 330 nm with water as a reference with water as a reference. The *DH* value of the samples was calculated from:

$$DH = 100 \times (C_H - C_0)/C_T \quad (6.1)$$

where C_H is the equivalent leucine concentration (as obtained from the TNBS method above) of the hydrolysed sample; C_0 is the equivalent leucine concentration of the non-hydrolysed sample; C_T is the *theoretical* equivalent leucine concentration if the protein was completely hydrolysed to its constituent amino acids, calculated from the total amine nitrogen groups in *SPI*.

It should be noted that the values of *DH* calculated in this way will probably always be < 100% because it will be almost impossible to achieve 100% hydrolysis of the soy protein in practice by pepsin under any conditions so that the denominator in Equation 6.1 is artificially high compared to the true value when practical pepsin action may be said to be 'complete'.

6.2.2.5 Protein Identification and PTM mapping

Some *SPH* samples were sent to the institute (Mass Spec Facility, Faculty of Biological Science, University of Leeds) to confirm their protein identity and determine the primary structures. Here is the detail of the procedures.

In-gel digest

Gel bands were excised and chopped into ~ 1 mm^3 pieces. Bands were detained by covering bands with 30 % ethanol and incubating them at 70 °C for 30 min with shaking. The supernatant was then discarded. This was repeated until all stain was removed. The gel pieces were washed by covering them with 25 mM ammonium bicarbonate/50% acetonitrile and incubated for ten min with shaking. The supernatant was discarded. To reduce cysteines, 100 µL 10 mM DTT solution was added and incubated at 57 °C for 1 hr with shaking. The supernatant was discarded, and the gel pieces were allowed to reequilibrate to room temperature. Cysteines were alkylated by the addition of 100 µL 55 mM iodoacetic acid and incubated at room temperature in the dark for 45 min with shaking. The supernatant was discarded. Gel pieces were dehydrated by the addition of 100% acetonitrile and incubation for 5 mins on the bench. The acetonitrile was removed, and the gel pieces were left to dry in a laminar flow hood for 60 min. Once dry, the gel slices were cooled on ice and then covered by the addition of ice-cold trypsin solution (20 ng µL-1 in 25 mM ammonium bicarbonate) and left on ice for 10 mins to rehydrate. Excess trypsin solution was removed and discarded, and the gel slices were covered with a minimal amount of 25 mM ammonium bicarbonate. After briefly vortexing and centrifuging, the gel slices were incubated at 37 °C with shaking for 18 hours. The resulting digest was vortexed and centrifuged. The supernatant containing the peptides was recovered and added to an eppendorf containing 5 µL acetonitrile/ water/ formic acid (60/35/5; v/v) to quench protease activity. 50 µL acetonitrile/ water/ formic acid (60/35/5; v/v) was added to the gel slices and vortexed for an additional 10 mins. The supernatant was pooled with the previous wash, and one additional wash of the gel slices was performed. The pool of three washes was dried by vacuum centrifugation. The peptides were reconstituted in 20 µL 0.1% aqueous trifluoroacetic acid.

LC-MS-MS of peptides

The 3 µL samples (approx. 0.6 ug of protein) were injected into an in-house-packed 20cm capillary column (inner diameter 75µm, 3.5µm Kromasil C18 media). An Ultimate 3000 nano liquid chromatography system was used to apply a gradient of 2–30% ACN in 0.1% formic acid over 30 min at a flow rate of 300 nl/min. Total acquisition time was 60 min including column wash and re-equilibration. Peptides were eluted from the column and into an

Orbitrap Exploris 240 Mass Spectrometer (ThermoFisher Scientific, Hemel Hempstead, UK) via a nanospray flex ion source using a capillary voltage of 2.7 kV. Precursor ion scans were acquired in the Orbitrap with a resolution of 60000. EASY-IC internal calibration was used for precursor ion scans. Up to 20 ions per precursor scan (charge state 2+ and higher) were selected for HCD fragmentation using a normalised collision energy of 30%. Fragments were measured in the Orbitrap at a resolution of 15000. Dynamic exclusion of 30 s was used.

Data analysis

Peptide MS/MS data were processed with PEAKS Studio XPro (Bioinformatic Solutions Inc, Waterloo, Ontario, Canada) and searched against the soy database. Carbamidomethylation was selected as a fixed modification, and variable modifications were set for the oxidation of methionine and deamidation of glutamine and asparagine. MS mass tolerance was 20 ppm, and fragment ion mass tolerance was 0.05 Da. The peptide false discovery rate was set to 1%.

6.3 Characterisation of selected SPHs and O/W emulsions stabilised by them

Emulsions were prepared using 9 different SPHs, as in Table 5.1, except the alcalase treatment was replaced by pepsin treatment at pH 4.7. This pH is further away from the optimum pH for pepsin action than pH 1.3 and was used to see if this might generate at least equally surface active fragments but perhaps even better stabilising material, i.e., higher M_w and fewer types of fragments. In fact, out of these 9 samples, only the *SPH* produced by both the pH values 2.1 and 4.7 were able to stabilise emulsions with any measurable stability at all – and both in the M_w faction > 10 kDa. Therefore, in the remaining characterisation of the SPHs and the corresponding emulsions, we have focused on only those samples stabilised by these two SPHs, denoted as P2 and P4, for simplicity.

DH for the P2 and P4 SPHs was calculated as 17.0 ± 4.6 and 11.1 ± 0.4%, respectively, substantiating the expectation that hydrolysis is less efficient at the higher pH, away from the optimum. Both values are still much lower than the theoretical degree of hydrolysis that can be achieved by pepsin

(maximum 25 %) if one only considers the peptide bonds that pepsin is traditionally considered capable of breaking. Similar conclusions were reached in a similar recent experimental study (Han et al., 2021). In fact, there are at least two factors that can lead to lower *DH* than expected : (a) the continuous hydrolysis of the protein by the enzyme leads to an increase in the pH of the solution, which then changes both the activity and the specificity of the enzyme; (b) the native compact globular form of the soy protein means that some potential cleavage sites remain inaccessible to the enzyme.

Figure 6.2 shows a comparison in appearance between 1 wt.% dispersion of the P2 and P4 samples and the non-hydrolysed *SPI*. The *SPI* is quite turbid, whilst the P4 sample preparation is almost completely transparent, and the turbidity of the P2 solution is somewhere between. This in itself suggested that the solubility of both hydrolyzates was higher than that of the raw protein but particularly for the pH 4.7 treatment with lower *DH*.

Figure 6.2 Visual images (from left to right) of 1 wt% non-hydrolysed SPI, P2 and P4 protein suspensions

This was substantiated by Mastersizer measurement so these dispersions, showing that the D_{32} of the *SPI* was reduced from 128.00 ± 21.00 to 4.40 ± 0.30 and 2.90 ± 0.24 μm for P2 and P4, respectively.

Measurements of ζ-potential were carried out for P2 and P4 solutions using section 2.2.4, with a background electrolyte concentration of 10 mM NaCl and a pH of 7. Both samples exhibited a negative charge, but the ζ-potential of P4 (–40.3 ±6.1 mV) was significantly higher in magnitude than that of P2 (–15.6 mV ± 2.4 mV). It must be emphasised that both samples will contain many different fragments, having varying charges, so these values are just average values of all these fragments. Nonetheless, the differences between the two samples are significant and suggest that all other things being equal,

adsorption of all the P4 material might lead to a higher emulsion droplet surface charge than adsorption of all the P2 material. Of course, preferential adsorption by certain fractions within each material could alter this prediction entirely.

Emulsions were prepared using 1 wt.% P2 and P4 as stabilisers according to the procedure described in section 2.2.2 and emulsion stability was monitored for up to 30 days. Figure 6.3 shows the visual appearance of the emulsions, Figure 6.4 shows the changes in the D_{32} value measured over the first 30 days, and Figure 6.5 shows the particle size distributions of the P2 and P4 in 20 days.

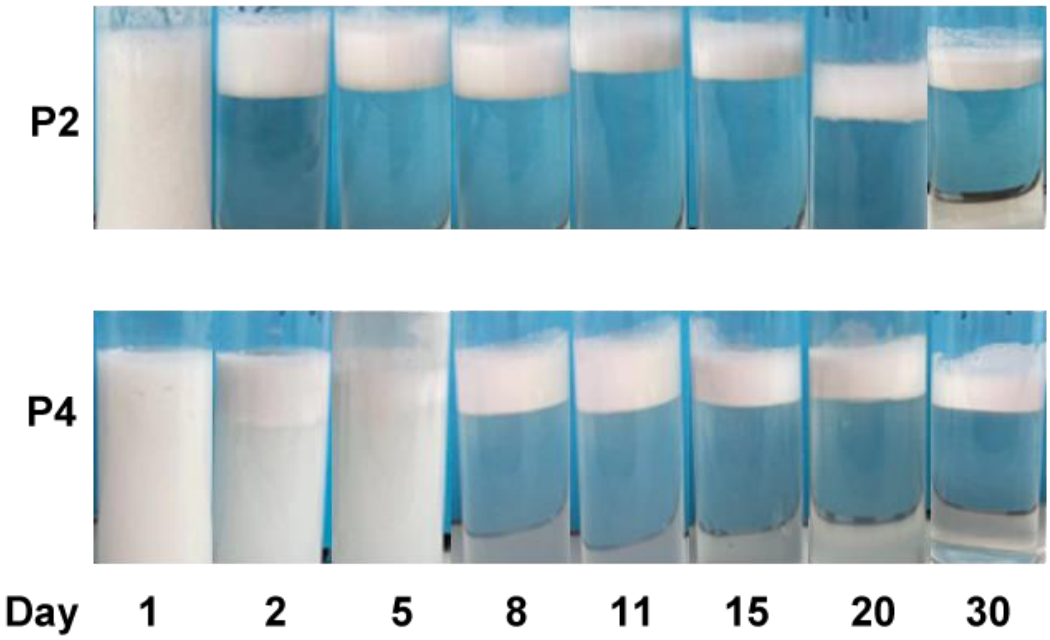


Figure 6.3 The visual appearance of emulsions prepared using 1 wt% P2 or P4

Considering the freshly prepared (i.e., day 1) emulsions, it is clear that the emulsion stabilised by P4 has a more uniform milky-white appearance than that stabilised by P2. After two days, the visual difference was even more marked: the P2 emulsion formed distinctly separate cream and aqueous layers, whilst the P4 emulsion became slightly less cloudy at the bottom of the tube and with a thicker cream layer just starting to become visible. At longer times (e.g. > 1 week, the appearances started to converge, but the earlier observations did suggest slower creaming of the P4 emulsions and, therefore, probably smaller mean droplet sizes. It is also worth stating that if

the P4 emulsion was hand shaken at any time, it easily reverted back to the appearance of the fresh sample, i.e., there was no strong irreversible droplet aggregation. From day 8 to day 60 (data not shown), the appearance of both emulsions remained largely unchanged, with no appreciable increase in the thickness of the cream layer.

That the P4 sample is better able to stabilise smaller droplets is confirmed in the *PSD* measurements shown in Figure 6.5. The D_{32} value of the fresh P4 emulsion is lower (0.4700 ± 0.0013 μm) than that of the P2 emulsion (11.92 ± 0.23 μm), but moreover, after a slight increase in the first 1 or 2 days (to 1.620 ± 0.051 μm). D_{32} was essentially constant until a slight increase was again observed between day 15 and day 20 (to 3.07 ± 0.036 μm). This might indicate the onset of some coalescence or flocculation, or the latter possibly leading to slight sampling errors if the samples are not shaken well enough before taking an aliquot for size measurements. In contrast, D_{32} of the P2 emulsion increases more or less continuously after day 5, reaching a value of 40.64 ± 3.43 μm after 20 days. Note that for these droplet sizes in a low viscosity medium, such as in this case (essentially water), creaming at this droplet volume faction (1.0 wt.%) and droplet size would be expected to be relatively rapid if the droplets were not strongly flocculated. (Rapid creaming and/or flocculation may, in fact, explain the apparent slight decrease in D_{32} for the P2 emulsion between day 1 and 4(from 11.92 ± 0.23 μm to 11.56 ± 0.39 μm), which may be due to a slight sampling error if the larger droplets cream out of the measurement path during the Mastersizer measurements).

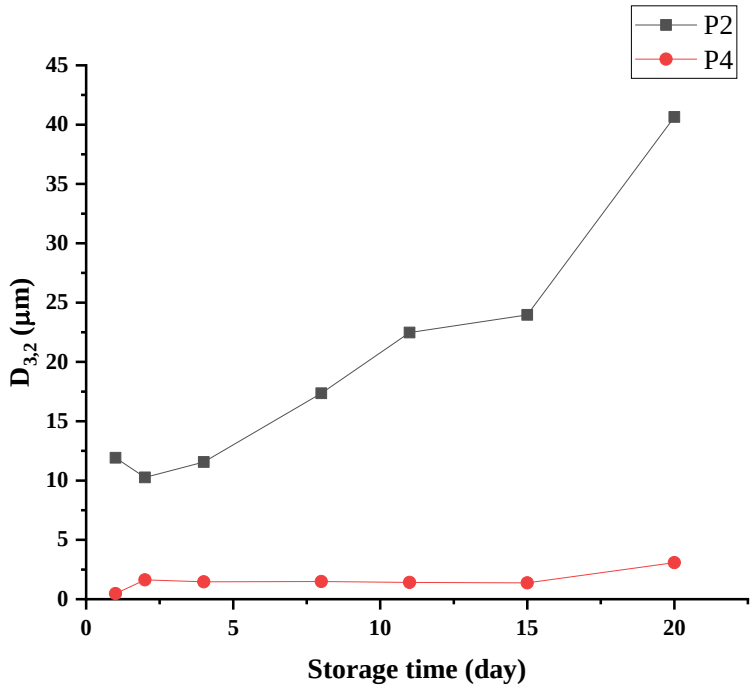
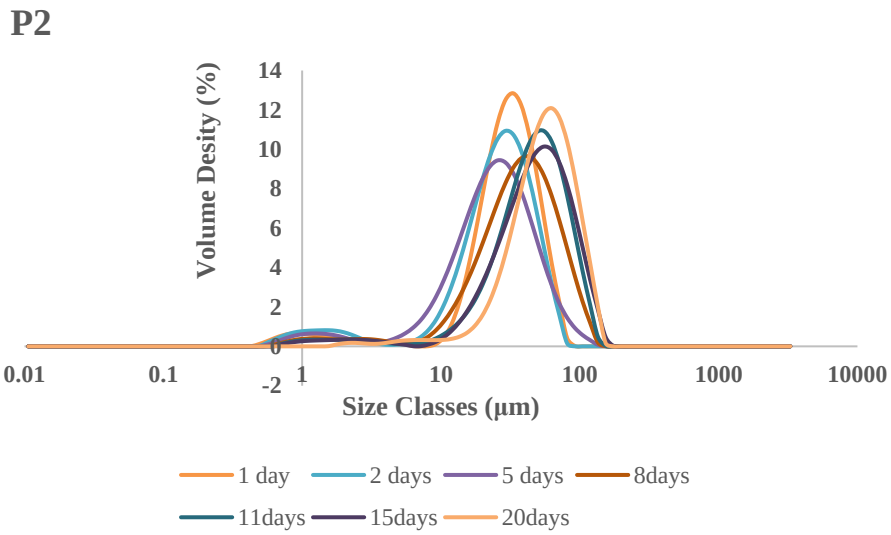


Figure 6.4 The change of droplet size $D_{3,2}$ of O/W emulsions, stabilised by P2 and P4 hydrolysates, with storage time. The samples were stored at pH7 and a temperature of 4℃



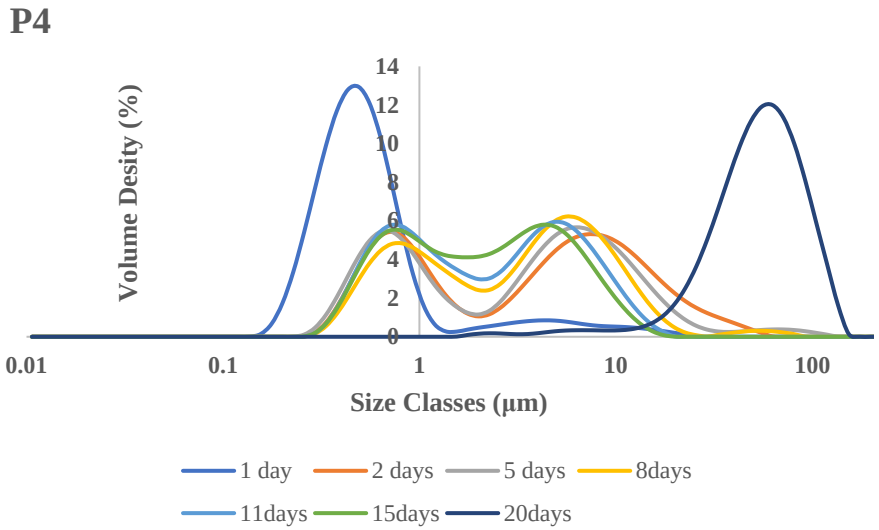


Figure 6.5 Particle size distributions of the P2 and P4 in 20 days (The x-axis is logarithmic with a base of 10)

It can be seen that the P2-emulsion forms a single peak when first prepared, and this peak gradually shifts flat to the right with time. This means that the rate of aggregation of the P2 stabilised emulsion is relatively uniform. In contrast to P2, the P4 emulsion also formed a single peak on day 1, but it started to form two peaks on day 2 until day 15. Finally, on day 20, the two peaks changed back to one peak. This result can be explained by comparison with Figure 6.4, where the P4-emulsion shows a significant aggregation on day 2. Then the aggregation process tends to slow down until the rate of aggregation increases again on day 20.

In summary, the P4 soy protein hydrolysates, obtained by pepsin treatment at the sub-optimal hydrolysis pH of 4.7 for the enzyme, exhibited significantly better emulsification and stability properties than those P2 obtained at optimum pepsin hydrolysis pH (2.1).

6.4 Protein sequencing results

The emulsion studies suggested significant superiority of the P4 sample and that this might be due to the presence of fewer types and/or higher M_w peptides with more suitable amino acid sequences for emulsion stabilisation. Therefore, it was worthwhile investigating this fraction in more detail to see if amino acid sequence information could be obtained, and this was challenging. Gel electrophoresis results (Figure 6.6) showed that the P4

sample was still very complex in composition, with a wide range of small fragments. Thus bands F, G and H are faint and broad, but there was one band (E) at 25 kDa that was relatively clear and more intense. Moreover, this M_w is in a range capable of imparting reasonable steric stabilisation – provided it adsorbs in the appropriate configuration. At the same time, the P4 sample was prepared by filtration through a 10 kDa ultrafiltration membrane so that only a small number of fragments with an Mw of less than 10kDa should be present. Sections 3.1 and 3.2 demonstrated the influence of the size (Mw) of the peptide on its emulsification properties, and the emulsion results seem to agree with the predictions that there is a lower M_w size limit below which the peptides will not be effective stabilisers. All in all, the dominant 25 kDa band, therefore, seemed a strong contender for the dominant species imparting the good stabilising characteristics of the P4 hydrolyzate and full sequencing of the peptides in this band was attempted.

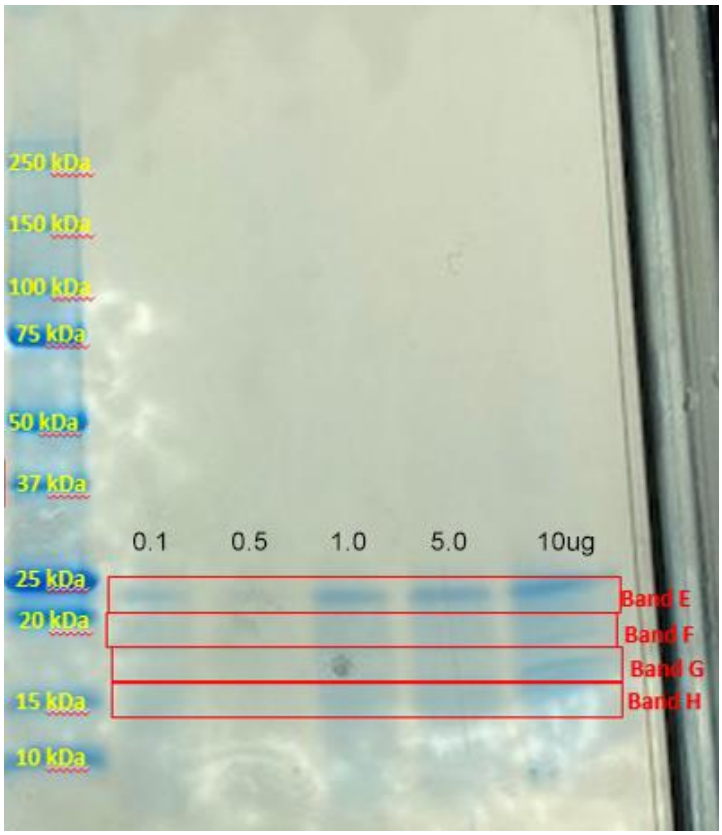


Figure 6.6 SDS-PAGE pattern of the P4 hydrolysate

Unfortunately, the sequencing instrument/technique used can only measure fragments of lower size. Therefore, the already hydrolysed P4 sample was further broken down into shorter sequences via selective proteolysis, with the hope that the original sequence(s) could be re-constructed from the

measured sequences, their overlap and the known amino acid sequences of the constituent soy proteins. Trypsin, chymotrypsin and glutamic proteases were used.

In the final analysis, the protein sequencing identified a total of 13 protein fragments from about six different proteins, including eight fragments from the β-conglycinin-α-subunit and a small number of fragments from glycinin (G4 and G5), β-conglycinin α' subunit, trypsin inhibitor (A and B) and Kunitz trypsin inhibitor. Thus, the conclusion was that most of the detected fragments (see the blue lines in Figure 6.7) must have originated from the β-conglycinin-α-subunit, which constitutes one of the main components of the soy protein. As mentioned above, it is difficult to measure the complete structure of all components in a band due to the limitations of the sequencing method. Predicted sequence 1 (PS1, the red line in Figure 6.7) contains all of the detected fragments and has an Mw = 12 kDa. This is obviously somewhat smaller than the supposed M_w of band E (25 kDa), but it should be remembered that there will be some error on this 25 kDa M_w due to the M_w markers in the standard not necessarily aligning perfectly with the sample gel.

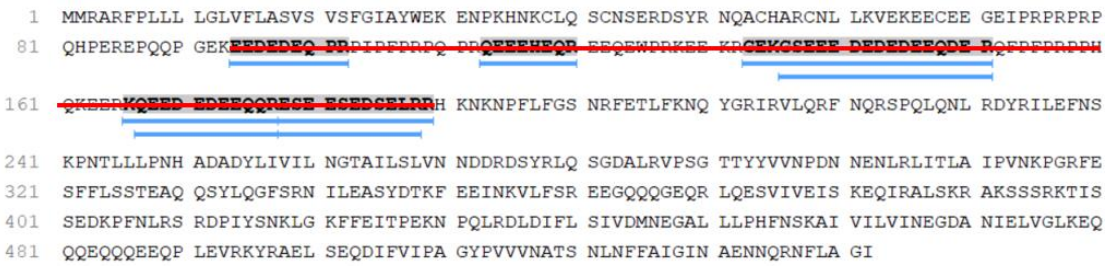


Figure 6.7 Sequence results of the fragments detected in the P4 sample for β-conglycinin-α-subunit, involving the hydrolysates in the 25 kDa band. The blue lines represent the sequences found in the sample that match this part of the subunit sequence; the red line represents the predicted sequence 1 (PS1) that contains all of the above but also the intervening parts of the sequence

Based on PS1, we can obtain predicted sequences that have the more 'correct' Mw of 25 kDa by including appropriate extensions on either side of the PS1 sequence. Obviously, there are a large number of possibilities, depending on what fraction of the 'missing' mass is allocated to the amino or carboxyl end of the PS1 chain. However, as a first attempt, we chose the three possibilities indicated in Figure 9, denoted PS2, PS3 and PS4.

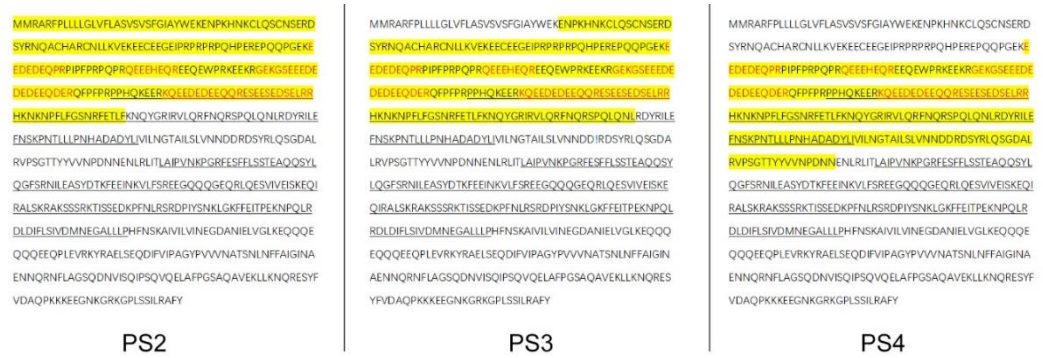


Figure 6.8 Three predicted protein fragments that may be derived from β-conglycinin and present in Band E of the P4 hydrolysate

From the 3 possible sequences PS2, PS3 and PS4 one can now start to calculate a variety of surface properties. Some of their parameters can be found in Table 6.1

Table 6.1 The pI and Mw values of the four predicted peptides

	PS1	PS2	PS3	PS4
pI	4.4	4.9	5.0	4.8
Mw (kDa)	12.0	25.0	24.5	25.0

Using the same SCF calculations as illustrated in Chapter 4, one can now also predict the adsorbed configurations and colloidal interactions induced by these three different peptides adsorbed layers of these fragments (see Figure 6.9). All the calculations were carried out at pH = 7 to match the conditions in the experimental section.

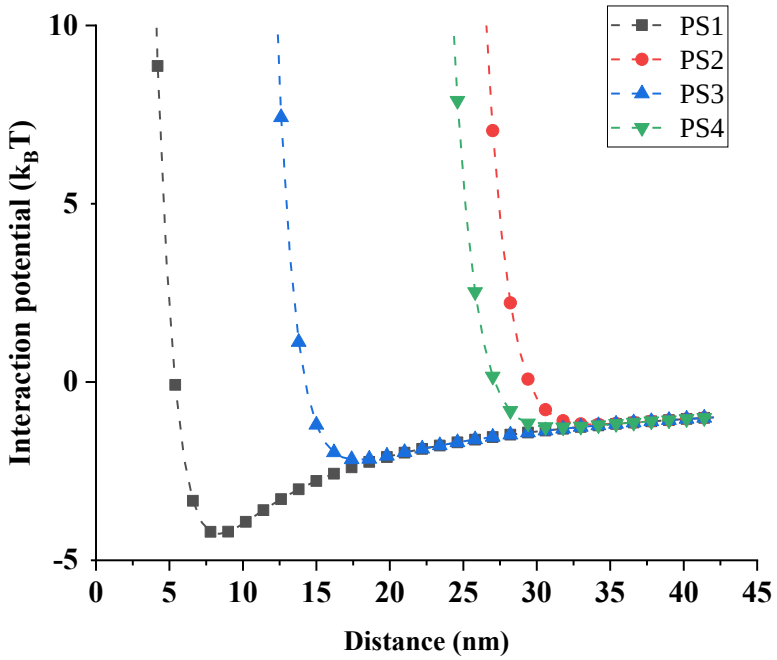


Figure 6.9 Calculated colloidal interaction potentials induced between two droplets of size 1 μm when covered by an adsorbed layer of each of the four possible predicted fragments likely to be present in the P4 hydrolyzate sample. All results also included van der Waals interactions and were calculated for a pH = 7. The volume fraction of the background electrolyte was set at 0.001 (~ 0.01mol/l)

As is seen in Figure 6.9, the depth of the interaction potential minimum mediated by the PS1 fragment is predicted to be 4.25 kBT. This occurs at a droplet separation distance of 8.4 nm. This is marginally deeper than the potential well obtained for the interaction potential produced by caseinate layers discussed in Section 3.2 (~ 2.14 kBT). The interaction potential induced by the PS2, PS3, and PS4 fragments, which all contain PS1 as part of their structure, were found to be broadly similar to PS1. Since pH 7 is far from the isoelectric point of the predicted sequences, all the systems can provide electrostatic, as well as potentially steric, repulsion. It can be seen that all the fragments are predicted to have fairly good emulsification performance (high repulsive energy at short surface separations) away from their isoelectric point. Nonetheless, the higher Mw peptides, PS2, PS3 and PS4, are seen to provide longer range repulsion than the shorter PS1, whilst in terms of increasing order of repulsion PS2 > PS4 > PS3 > PS1. We suppose that the reason for this ordering may be related to the absorbed configuration of four peptides (see Figure 6.10). As shown in Figure 11, both PS2 and PS4, exhibiting a relatively perfect di-block-like structure, tend to behave as such when adsorbed to the interface. Compared to the tri-block

structure of PS3, the polypeptides PS2 and PS4 are seen to protrude further away from the surface. As such they are able to provide longer-ranged steric repulsion. In addition, the hydrophilic tail of the PS2 seems to constitute a larger portion of the chain and extends marginally further outwards away from the interface than the PS4. Again, this allows the PS2 to form a thicker adsorbed interfacial layer than the PS4, so we may expect and indeed find that PS2 has an even better provision of steric interactions than PS4. As for PS1, although it also has a similar roughly di-block-like structure, its molecular weight is too small to provide sufficiently long ranged steric forces. Thus, its emulsion stabilizing capacity is predicted not to be as good as that of PS3, which has a larger molecular weight but a tri-block-like structure. This result is again in agreement with the discussion in Chapter 5.

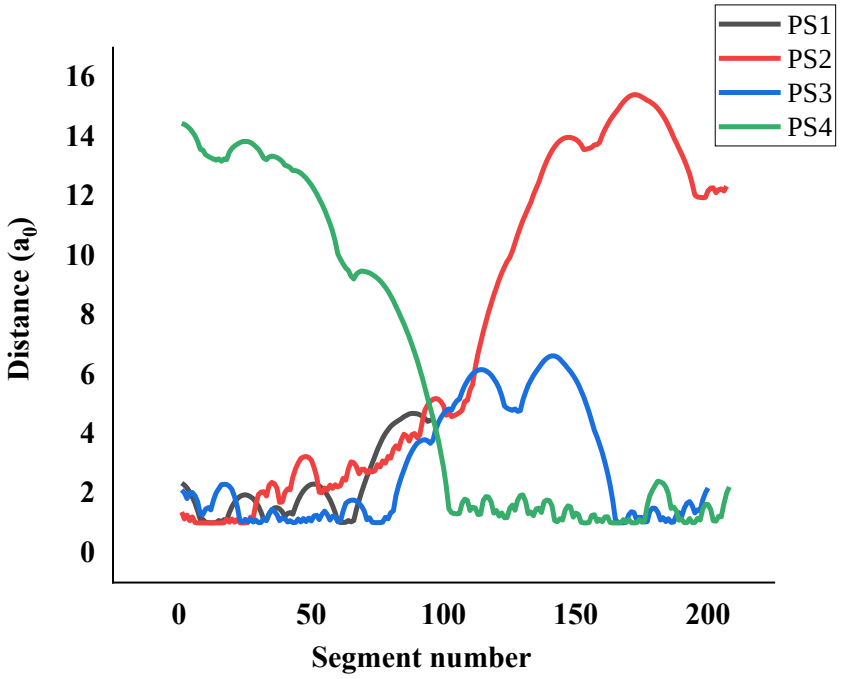


Figure 6.10 Predicted configuration of four peptides at pH 7 upon adsorption, calculated by SCFC. The graphs show the average distance of each amino acid in the sequence of the polypeptide away from the surface. The volume fraction of the background electrolyte was set to 0.001 (~ 0.01mol/l) in all cases

All in all, the agreement between the behaviour of these predicted fragments and that of the real hydrolyzates is remarkable, even taking into account the discrepancy in the apparent M_w of band E and the M_w of the only fragments identified.

6.5 Conclusions

We have shown that SCF calculation predicts that even (theoretical) selective hydrolysis of β-casein to what might seem a more ideal, simple block-copolymer structure does not result in any advantages in terms of stabilising capacity. In fact, the opposite. Thus β-casein naturally seems to have the optimal structure, chain length and charge distribution to give maximum electrostatic stabilisation, a least at pH 7. Highly selective proteolysis of other proteins to produce fragments that closely enough match the β-casein sequence will probably be very difficult to achieve. Rather, proteolytic enzymes, under their optimal conditions of action, are more likely to produce a range of fragments that are too low in M_w to be of any use as stabilisers (even though they may be surface active), even when the degree of hydrolysis (% of peptides binds cleaved relatively low (< 10%)). On the other hand, we have shown that deliberately moving away from optimal pepsin hydrolysis of soy protein isolate leads to a range of higher M_w fragments that apparently have much superior surface activity and O/W emulsions stabilising properties compared to the native non-hydrolysed protein. This was explained by the M_w fraction > 10 kDa having peptide sequences that *SCFC* indeed predicted to have the right combination of hydrophobic and hydrophilic segments to give strong adsorption but also a strong inter-droplet repulsion force. Therefore this strategy of using enzymes that are commonly used to help solubilise plant proteins but under non-optimal conditions might be exploited as a general method for producing longer, more functional proteins (or other biopolymers) from a range of plant materials that also can be explained and predicted by theory.

List of References

Agyare, K. K., Addo, K. and Xiong, Y. L. 2009. Emulsifying and foaming properties of transglutaminase-treated wheat gluten hydrolysate as influenced by pH, temperature and salt. Food Hydrocolloids. 23(1), pp.72-81.

Burgaud, I., Dickinson, E. and Nelson, P. 1990. An improved high‐pressure homogenizer for making fine emulsions on a small scale. International journal of food science technology. 25(1), pp.39-46.

Han, R., Hernández Álvarez, A. J., Maycock, J., Murray, B. S. and Boesch, C. 2021. Comparison of alcalase- and pepsin-treated oilseed protein hydrolysates – Experimental validation of predicted antioxidant, antihypertensive and antidiabetic properties. Current Research in Food Science. 4, pp.141-149.

Jin, F., Wang, Y., Tang, H., Regenstein, J. M. and Wang, F. 2020. Limited hydrolysis of dehulled walnut (Juglans regia L.) proteins using trypsin: Functional properties and structural characteristics. LWT. 133, p110035.

Yu, J., Wang, G., Wang, X., Xu, Y., Chen, S., Wang, X. and Jiang, L. 2018. Improving the freeze-thaw stability of soy protein emulsions via combing limited hydrolysis and Maillard-induced glycation. LWT. 91, pp.63-69.

Chapter 7
General Conclusion and Discussion

7.1 Introduction

Due to their distinct safety profile, fragmented proteins can be regarded as one of the potentially useful food emulsifiers in the modern food business. This is particularly true for plant-based proteins, where often the full intact proteins do not show acceptable emulsification properties. The theoretical and practical work conducted as part of this PhD project is focused on finding soy protein fragments with good emulsification capabilities. The approach taken has produced some promising findings. The critical results of the study are summarised in the following part, and this is followed with recommendations for future work that we might be able to improve even further on present results.

7.2 Fast screening methods

Except for casein, few proteins found in nature, employed directly in the food industry, have anywhere close to an ideal emulsifier structure. However, casein, as a relatively expensive animal-based protein, has some limitations when used as an emulsifier, as was discussed in Chapter 1. In this work, we applied two fast screening methods to look for a fragment on the β-conglycinin-α-subunit, a component of soy protein. The aim was to identify a polypeptide fragment that resembles a di-block-like structure comparable to that of β-casein. We believe that under conditions when charge repulsion is also reasonably sufficient, the identified fragment, which has a close to a di-block-type structure (Murray et al., 2021), produces significant steric repulsion and thus has the potential of being a very good food emulsifier. The significance of the fast screening methods is that they can assist us in narrowing our search for suitable fragments and finding fragments with specific desirable structures. This saves considerable time for subsequent, more in-depth SCF calculations.

7.3 Self-consistent field calculations

SCF calculations were carried out on the di-block-like fragment selected by the fast screening methods to confirm whether the fragment was indeed a suitable emulsifier in the amount of polymer adsorption, volume fraction, polymer conformation and interaction potential. These calculations allow us to obtain the inter-particle (inter-droplet) interaction potential mediated by adsorbed layers of the polypeptide fragment on the surface of the particles. The potential surface separation graphs allow one to assist if the adsorbed protein is able to keep the droplets apart and prevent aggregation.

7.3.1 Amount of polymers adsorption and volume fractions

We have studied the adsorption of fragments in terms of pH, polymer volume fraction and salt volume fraction. SCF calculations show that pH strongly influences the amount of adsorption. The closer one is to the isoelectric point (pH = 6.08) of the fragment, the greater the amount of adsorbed fragment on the interface. This is mainly because the net charge of the fragment is close to zero at the isoelectric point, and thus, with the repulsion between chains switched off, more chains can be accommodated on the interface. Furthermore, the bulk protein volume fraction does not significantly affect the amount of adsorption. When the bulk polymer volume fraction is increased from 10^{-8} to 10^{-3} (i.e. by five orders of magnitude), the amount of adsorbed polymer on the surface does not even double. This shows that surfaces become saturated with protein even at very low bulk protein volume fractions. The effect of salt volume fraction on polymer adsorption is divided into two prominent cases: low salt volume fraction ($\phi_{\text{salt}} = 0.0001$, $\phi_{\text{salt}} = 0.001$, $\phi_{\text{salt}} = 0.005$ and $\phi_{\text{salt}} = 0.01$) and high salt case ($\phi_{\text{salt}} = 0.05$ and $\phi_{\text{salt}} = 0.1$). At low salt volume fractions, the adsorption of all polymers was relatively similar at the isoelectric point. In contrast, at non-isoelectric points, the adsorption increased with rising salt volume fraction because of the screening effect of the electrolyte, reducing electrostatic repulsion between the chains. At high salt volume fractions, the amount of adsorbed protein did not differ significantly at different pH values and was significantly higher than that at low salt volume fractions.

The volume fraction profile of the polymer in the vicinity of the surface exhibits a maximum near the isoelectric point, similar to adsorption. It then

starts to drop as the net charge of the fragment rises (i.e. as we move away from the isoelectric point). However, when salt volume fraction increases, a screening effect reduces the difference between the volume fractions corresponding to various pH values. At the highest salt volume fraction ($\phi_{salt} = 0.1$) utilised in the calculations, the volume fraction of the adsorbed protein is almost identical at various pH levels since the electrical charge of the chains plays a diminishing role due to its screening by salt ions.

7.3.2 Polypeptide fragment conformation

The effect of pH and salt volume fraction on the protein conformation is mainly displayed by the outward extension of the dangling tail formed by the hydrophilic blocks of our diblock-like fragment. Increasing salt volume fraction can screen the charge on the protein chains and also increase the amount of adsorption. When the pH is close to the isoelectric point, the protein chain does not carry much net charge. Thus it is less affected by the screening effect of the salt ions, while increasing the amount of adsorption can significantly increase the extent of stretching the polymer tail away from the interface into the bulk. In contrast, when the pH is far from the isoelectric point, the protein chain has a higher net charge. Therefore it is more strongly affected by the screening effect of electrolytes at increasing volume fractions. At this point, the screening effect tends to reduce the degree of the stretching of the dangling tail. Therefore, there is an optimum interplay between pH and salt volume fraction, seen in the calculations, that maximises the extent of stretching of the loop (in this study, the optimum solution is pH = 12 and $\phi_{salt} = 0.001$), giving rise to the thickest (most extended) interfacial adsorbed protein films.

7.3.3 Interaction potential

Interaction potential mediated by adsorbed layers is an important indicator in describing whether a fragment can be a suitable emulsifier. In this study, the minimum depth of the interaction potential induced by adsorbed polypeptide fragments is always the smallest when the calculation is performed at the pH values far from the isoelectric point. In the vicinity of the isoelectric point, the magnitude of the interaction potential minimum fluctuated with salt volume fractions, mainly due to the balance between steric repulsion, electrostatic repulsion and van der Waals forces. In a subsequent comparison with α_{S1}-

casein and β-casein, at their isoelectric point, the depth of the energy minimum in the interaction potential of our chosen di-block-like fragment, derived from soya protein, is similar to that of β-casein and much smaller than that of α_{S1}-casein. The magnitude of the minimum value of the interaction potential is similar to that of both β-casein and α_{S1}-casein at a pH far from the isoelectric point. Therefore, we believe that this soya protein-derived di-block-like fragment could theoretically be a very suitable food emulsifier derived entirely from plant-based materials. Furthermore, as the isoelectric point of this fragment (pI = 6.08) is different from that of both β-casein (pI = 5.14) and α_{S1}-casein (pI = 5.31), it may be used in the future in the food industry as a substitute for casein in acidic pH situations where casein is not such a good emulsifier.

7.4 *SPH* prepared at a sub-optimal pH as a novel food-grade emulsifier

In measuring the surface tension of protein hydrolysates obtained under different conditions, we found that pepsin treatment significantly reduced the surface tension of the soy protein. Therefore, three pH values of 1.3, 2.1 and 4.7 were used during the hydrolysing of the soya proteins and the emulsification properties of the resulting hydrolysates obtained under each pH regime were investigated. The results demonstrated that the best emulsification performance was obtained for *SPH* with Mw > 10 kDa prepared at the pepsin's sub-optimal hydrolysis pH of 4.7. Compared to the *SPH* prepared at the optimum hydrolysis pH, the *SPH* prepared at the sub-optimal hydrolysis formed smaller droplets and showed better stability over the 30-day observation. Based on a subsequent sequencing of *SPH-4.7*, we attempted to identify the resulting fragments and analyse the emulsification properties of these possible fragments using SCF calculations. This showed that all four predicted likely fragments exhibited excellent emulsification ability (the depth of the minimum in the interaction potential was less than 5 k_BT). Thus, the results of this theoretical calculation confirm the existence of fragments with good emulsification properties in the hydrolysates obtained at the sub-optimal pH.

Indiscriminate fragmentation, even at a moderate degree of hydrolysis (say 10%), leads to the production of many small polypeptides. Very few larger

and more desirable chains remain in such cases. The small polypeptide chains are shown here not to be very helpful (or even counterproductive) for stabilising the emulsions. This was predicted to be the case even in the most favourable situation, assuming that the fragments possessed the ideal diblock-type structure (Murray et al., 2021). To improve this situation, two possible strategies can be envisaged. The first is to design very specific enzymes with exact cleavage sites that can produce very precise desirable fragments that have been theoretically identified to act as good emulsifiers. However, even here, there may be a need to filter out other less desirable hydrolysate by products that may compete with the required fragments for adsorption, as was shown by Ettelaie et al. (Ettelaie et al., 2014). This strategy is both expensive and difficult to implement.

The second approach is to employ the widely available enzymes but to use them at pH values where they may not be very active. As such, they will also tend to be more selective in peptide bonds that they can cleave. Thus, the hope is that in doing so, there will be a large train of consecutive bonds which are immune to hydrolysis. This, in turn, will lead to the formation of some larger and more suitable fragments. The results presented here, though by no means fully optimised, do seem to provide some support for the promising potential of this second approach.

7.5 Conclusions and future work

This PhD project has demonstrated theoretically and experimentally the possibility of using fragmented proteins as novel food-grade emulsifiers. This is particularly of interest in relation to otherwise generally less soluble and globular plant proteins. By selecting the appropriate enzyme (enzymatic cleavage site) and controlling the suitable degree of hydrolysis, we can obtain peptide fragments with a reasonable emulsification capacity from intact proteins that by themselves do not naturally have good emulsification properties.

In theoretical chapters 3 and 4, we have succeeded in finding a fragment with a di-block structure from the β-conglycinin-α-subunit with excellent emulsification properties by analysing its primary structure. We speculate

that such similar fragments are not unique to the same protein. For example, in this study, the di-block-like fragment is located in the middle of amino acid residues 18 - 118 on the intact protein backbone. Similarly, one could argue that fragments [19 – 118] or [20-118] are likewise di-block-like fragments. However, SCF calculations suggest that any addition/deletion of hydrophobic/hydrophilic amino acid residues may dramatically affect the emulsification properties of the fragment we selected. Therefore, in order to investigate the emulsification performance of all fragments that may have the potential to be good emulsifiers, we needed a faster screening method in the first instance. It may be possible to specify more detailed criteria for evaluating the merits of protein fragments (i.e. a numerical mark to describe the possibility of a fragment that could be a suitable emulsifier) for such a fast screening method before performing the more accurate but time-consuming SCF calculations. Specific criteria could be based on the distribution of hydrophobic/hydrophilic residues, the distribution of charges residues and the length of the fragment, or more detailed evaluation criteria could be specified based on machine learning models. Concerning the SCF calculation, to make the calculation more realistic, we may be able to further take into account factors such as disulphide bonds between amino acids and the secondary structure of the protein to some extent in the calculations.

Once we have theoretically identified a fragment with good emulsification properties, it would not necessarily be easy to obtain this in practice. The main difficulty is the isolation and purification of specific fractions from hydrolysed proteins to achieve the isolated relevant hydrolysates. For example, it is already complicated to isolate the β-conglycinin (what we require is β-conglycinin-α-subunit) involved in this study from the SPI in a sufficient amount for the preparation of emulsion experiments (Thanh et al., 1975; Thanh and Shibasaki, 1976; Iwabuchi and Yamauchi, 1987; Nagano et al., 1992; Liu et al., 2007). Therefore, our study turned to the emulsification properties of the hydrolysates obtained at the sub-optimal hydrolysis pH of the enzyme. As illustrated in Figure 6.1, hydrolysis of proteins at sub-optimal pH is more likely to result in peptide fragments of sufficient length to be used as emulsifiers, even though these might not involve one particular fragment but will be mixtures of several hydrolysates. In future work, we can combine theoretical calculations with trying to obtain hydrolysates with excellent emulsification properties for different proteins at their sub-optimal pHs using a more comprehensive range of enzymes.

In conclusion, as there are many different types of proteins and enzymes in nature, it is also possible to select the suitable emulsifier and stabiliser from a wide range of protein hydrolysates.

List of References

Ettelaie, R., Zengin, A. and Lee, H. 2014. Fragmented proteins as food emulsion stabilizers: A theoretical study. Biopolymers. 101(9), pp.945-958.

Iwabuchi, S. and Yamauchi, F. 1987. Determination of glycinin and. beta.-conglycinin in soybean proteins by immunological methods. Journal of Agricultural Food Chemistry. 35(2), pp.200-205.

Liu, C., Wang, H., Cui, Z., He, X., Wang, X., Zeng, X. and Ma, H. 2007. Optimization of extraction and isolation for 11S and 7S globulins of soybean seed storage protein. Food chemistry. 102(4), pp.1310-1316.

Murray, B. S., Ettelaie, R., Sarkar, A., Mackie, A. R. and Dickinson, E. 2021. The perfect hydrocolloid stabilizer: Imagination versus reality. Food Hydrocolloids. 117, p106696.

Nagano, T., Hirotsuka, M., Mori, H., Kohyama, K. and Nishinari, K. 1992. Dynamic viscoelastic study on the gelation of 7 S globulin from soybeans. Journal of Agricultural Food chemistry. 40(6), pp.941-944.

Thanh, V. H., Okubo, K. and Shibasaki, K. 1975. Isolation and characterization of the multiple 7S globulins of soybean proteins. Plant physiology. 56(1), pp.19-22.

Thanh, V. H. and Shibasaki, K. 1976. Major proteins of soybean seeds. A straightforward fractionation and their characterization. Journal of Agricultural Food Chemistry. 24(6), pp.1117-1121.

Appendix A
Python Source Codes for the Programs

```
#The Python source code of classic SCF calculation


import numpy as np
from math import exp


np.set_printoptions(precision=6)


class SCF_v1:
    def __init__(self, sequence, surface_interaction, p_concentration, layers, w_type, initial1,
                 all_momomers_interaction, ion_concentration, initial2, mono_charges, ion_exist=1):
        self.sequence = sequence
        self.Ni = len(sequence)
        self.surface_interaction = surface_interaction
        self.p_concentration = p_concentration
        self.s_concentration = 1 - p_concentration - ion_concentration * 2
        self.layers = int(layers)
        self.monomer_types = len(surface_interaction)
        self.w_type = w_type
        self.convert_initial_guest = self.__convert_initial_guest(initial1)
        self.monomer_interaction = all_momomers_interaction
        self.ion_exist = ion_exist
        self.ion_concentration = ion_concentration
        self.initial_guesses_elec_field = initial2
        self.mono_charges = mono_charges
        self.a0 = 3 * 1.0e-10
```

```
        self.vacuum_water = 78.5
        self.vacuum_0 = 8.85 * 1.0e-12
        self.e = -1.6 * 1.0e-19
        self.kB = 1.38 * 1.0e-23
        self.Temp = 298

    def __convert_initial_guest(self, t):
        X = np.array(t).reshape([self.monomer_types, self.layers])
        return X

    def bulk_volume_fraction(self):
        monomers_bulk = np.empty(self.monomer_types)
        monomers_bulk[0] = self.s_concentration
        length = len(self.sequence)
        aa = None
        if self.ion_exist == 1:
            aa = -2
            monomers_bulk[-1] = self.ion_concentration
            monomers_bulk[-2] = self.ion_concentration
        else:
            pass
        for i in list(self.w_type.keys())[1:aa]:
            monomers_bulk[self.w_type[i]] = self.sequence.count(i) / length *
self.p_concentration
        return monomers_bulk

    def G1_values(self):
        G1r = np.empty([self.monomer_types, self.layers], dtype=np.float32)
        convert_guess = self.convert_initial_guest
        for w in range(self.monomer_types):
```

```
            for l in range(self.layers):
                G1r[w, l] = exp(-convert_guess[w, l])
        return G1r

    def Gn_values(self):
        G1_all = self.G1_values()

        def Gn_any(target_sequence):
            n = 1
            Gnr = np.empty([self.Ni, self.layers], dtype=np.float32)
            Gnr[0, :] = G1_all[self.w_type[target_sequence[0]], :]
            for m in target_sequence[1:]:
                for l in range(self.layers):
                    field_ti = G1_all[self.w_type[m], l]
                    if l - 1 == -1:
                        Gnr[n, l] = field_ti * (
                                (4 / 6 * Gnr[n - 1, l]) + (1 / 6 * Gnr[n - 1, l + 1]))
                    elif l + 1 == self.layers:
                        Gnr[n, l] = field_ti * (
                                (1 / 6 * Gnr[n - 1, l - 1]) + (4 / 6 * Gnr[n - 1, l]))
                    else:
                        Gnr[n, l] = field_ti * (
                                (1 / 6 * Gnr[n - 1, l - 1]) + (4 / 6 * Gnr[n - 1, l]) + (1 / 6 *
Gnr[n - 1, l + 1]))
                n = n + 1
            return Gnr

        return Gn_any(self.sequence), Gn_any(''.join(reversed(self.sequence)))

    def monomer_concentration(self):
```

```
G1_all = self.G1_values()
Gn_all = self.Gn_values()
aa = None
con_m = np.empty([self.monomer_types, self.layers], dtype=np.float32)
con_m[0, :] = G1_all[0, :] * self.s_concentration
if self.ion_exist == 1:
    con_m[-1, :] = G1_all[-1, :] * self.ion_concentration
    con_m[-2, :] = G1_all[-2, :] * self.ion_concentration
    aa = -2
else:
    pass

for m in list(self.w_type.keys())[1:aa]:
    for r in range(self.layers):
        sum_G = 0
        Count_Num = 0
        for n in range(self.Ni):
            Gf = Gn_all[0][n, r]
            Gb = Gn_all[1][self.Ni - n - 1, r]
            if self.sequence[Count_Num] == m:
                theta = 1
            else:
                theta = 0
            Gti = G1_all[self.w_type[m], r]
            sum_G = sum_G + (Gf * Gb * theta / Gti)
            Count_Num = Count_Num + 1
        con_m[self.w_type[m], r] = self.p_concentration / self.Ni * sum_G
con_p = np.empty(self.layers)
for l in range(self.layers):
```

```
            con_p[l] = sum(con_m[1:aa, l])
        return con_m, con_p


    def average_concentration(self):
        av_con_m = np.empty([self.monomer_types, self.layers],
dtype=np.float32)
        monomer_con = self.monomer_concentration()[0]
        for w in range(self.monomer_types):
            for l in range(self.layers):
                if l - 1 == -1:
                    av_con_m[w, l] = 4 / 6 * monomer_con[w, l] + 1 / 6 * \
                                     monomer_con[w, l + 1]
                elif l + 1 == self.layers:
                    av_con_m[w, l] = 1 / 6 * monomer_con[w, l - 1] + 4 / 6 * \
                                     monomer_con[w, l]
                else:
                    av_con_m[w, l] = 1 / 6 * monomer_con[w, l - 1] + 4 / 6 * \
                                     monomer_con[w, l] + 1 / 6 * \
                                     monomer_con[w, l + 1]
        return av_con_m


    def f_ar(self):
        f_ar_all = np.empty([self.monomer_types - 1, self.layers],
dtype=np.float32)
        hard_core = np.empty(self.layers)
        line_num = 0
        av_con_m = self.average_concentration()
        monomers_bulk = self.bulk_volume_fraction()
        for m in list(self.w_type.values())[1:]:
            tag1 = str(m)
```

```
            for r in range(self.layers):
                if r == 0 or r == self.layers - 1:
                    theta = 1
                else:
                    theta = 0

                qa = self.mono_charges[m]
                ele_field = self.initial_guesses_elec_field[r]
                ele_total_field = qa * ele_field

                sum_kai_f = 0
                sum_kai_b = 0
                tem_ls = list(self.w_type.values()).copy()
                tem_ls.remove(int(tag1))

                tem_water = list(self.w_type.values()).copy()[1:]
                for i, ii in zip(tem_ls, tem_water):
                    tag2 = str(i)
                    kai_f = self.monomer_interaction[tag1 + tag2]
                    av_con_f = av_con_m[i, r]
                    av_con_b = av_con_m[ii, r]
                    sum_kai_f = sum_kai_f + kai_f * (av_con_f - monomers_bulk[i])
                    tag3 = '0' + str(ii)
                    kai_b = self.monomer_interaction[tag3]

                    sum_kai_b = sum_kai_b + kai_b * (av_con_b -
monomers_bulk[ii])
                f_ar_all[line_num, r] = (self.convert_initial_guest[m, r] -
                                         theta * self.surface_interaction[m] - ele_total_field -
sum_kai_f) - \
```

```
                (self.convert_initial_guest[
                    0, r] - sum_kai_b)
            hard_core[r] = self.convert_initial_guest[0, r] - sum_kai_b - self.initial_guesses_elec_field[r]
            line_num = line_num + 1

        return f_ar_all, hard_core

    def g_r(self):
        sum_con = np.log(self.monomer_concentration()[0].sum(axis=0))
        return sum_con

    def h_r(self):
        monomer_con = self.monomer_concentration()[0]
        h_r_all = np.empty(self.layers)

        for r in range(self.layers):
            q_total = 0
            for i in range(self.monomer_types):
                q_total = q_total + self.mono_charges[i] * monomer_con[i, r] * self.e ** 2 / (
                        self.vacuum_water * self.a0 * self.kB * self.Temp * self.vacuum_0)
            if r == 0:
                h_r_all[r] = self.initial_guesses_elec_field[r] - self.initial_guesses_elec_field[r + 1] - q_total
            elif r == self.layers - 1:
                h_r_all[r] = self.initial_guesses_elec_field[r] - self.initial_guesses_elec_field[r - 1] - q_total
            else:
```

```
            h_r_all[r] = ((2 * self.initial_guesses_elec_field[r] -
self.initial_guesses_elec_field[r - 1]
                          - self.initial_guesses_elec_field[r + 1])) - q_total
        return h_r_all


    def total_errors(self):
        far = self.f_ar()[0]
        gr = self.g_r()
        hr = self.h_r()
        xxx = np.vstack([far, gr, hr]).flatten()
        return xxx


    def get_average_distance(self):
        A_total = []
        Gf, Gb = self.Gn_values()
        G1_all = self.G1_values()
        con_n = np.empty([self.Ni, self.layers], dtype=np.float32)
        count_monomer = 0
        lll = int(self.layers / 2)
        for n in self.sequence:
            for l in range(self.layers):
                gf = Gf[count_monomer, l]
                gb = Gb[self.Ni - count_monomer - 1, l]
                Gti = G1_all[self.w_type[n], l]
                con_n[count_monomer, l] = (gf * gb / Gti) * (self.p_concentration /
self.Ni)


            count_monomer = count_monomer + 1
        con_n_sum = con_n[:, 0:lll].sum(axis=1)[0]
        for n in range(self.Ni):
```

```
            distance = 0
            for l in range(lll):
                con_nr = con_n[n, l]
                r = l + 1
                distance = distance + con_nr * r / con_n_sum
            A_total.append(distance)
        return np.array(A_total)

    def get_average_distance_surface(self):
        A_total = []
        Gf, Gb = self.Gn_values()
        G1_all = self.G1_values()
        con_n = np.empty([self.Ni, self.layers], dtype=np.float32)
        count_monomer = 0
        lll = int(self.layers / 2)
        for n in self.sequence:
            for l in range(self.layers):
                gf = Gf[count_monomer, l]
                gb = Gb[self.Ni - count_monomer - 1, l]
                Gti = G1_all[self.w_type[n], l]
                con_n[count_monomer, l] = (gf * gb / Gti) * (
                        self.p_concentration / self.Ni) - self.p_concentration /
self.Ni

            count_monomer = count_monomer + 1
        con_n_sum = con_n[:, 0:lll].sum(axis=1)[0]
        for n in range(self.Ni):
            distance = 0
            for l in range(lll):
                con_nr = con_n[n, l]
```

```
            r = l + 1
            distance = distance + con_nr * r / con_n_sum
        A_total.append(distance)
    return np.array(A_total)

def get_total_absorption(self):
    poly_con = self.monomer_concentration()[1]
    AA = 0
    for i in poly_con:
        AA = AA + (i - self.p_concentration)
    AA = AA / self.Ni
    return AA

def get_free_energy(self):
    mono_con = self.monomer_concentration()[0]
    bulk_con = self.bulk_volume_fraction()
    p_con = self.monomer_concentration()[1]

    F_non_p = 0.0
    F_non_s = 0.0
    for l in range(self.layers):
        F_non_p = F_non_p + (p_con[l] - self.p_concentration)
    F_non_p = F_non_p / self.Ni

    for l in range(self.layers):
        F_non_s = F_non_s + (mono_con[0, l] - self.s_concentration)
    F_non = -F_non_p - F_non_s

    fields = self.convert_initial_guest
```

```
        E_non = 0
        for l in range(self.layers):
            E_non_tem = 0
            for w in range(self.monomer_types):
                E_non_tem = E_non_tem + fields[w, l] * mono_con[w, l]
            E_non = E_non + E_non_tem

        av_con = self.average_concentration()
        E_kai = 0
        for l in range(self.layers):
            for w in list(self.w_type.values()):
                for ww in list(self.w_type.values()):
                    kai = str(w) + str(ww)
                    if w != ww:
                        E_kai = E_kai + self.monomer_interaction[kai] * (mono_con[w,
l] - bulk_con[w]) * \
                                (av_con[ww, l] - bulk_con[ww]) / 2

        E_kai_bulk = 0
        for w in list(self.w_type.values()):
            for ww in list(self.w_type.values()):
                if w != ww:
                    kai = str(w) + str(ww)
                    E_kai_bulk = E_kai_bulk + self.monomer_interaction[kai] *
bulk_con[w] * (
                            1 / 6 * mono_con[ww, 0] - bulk_con[ww])
                    E_kai_bulk = E_kai_bulk + self.monomer_interaction[kai] *
bulk_con[w] * bulk_con[ww]
        E_kai = E_kai - E_kai_bulk
```

```
        E_elect = 0
        for l in range(self.layers):
            E_elect_tem = 0
            for w in range(self.monomer_types):
                E_elect_tem = E_elect_tem + self.mono_charges[w] * mono_con[w, l]
            E_elect_tem = self.initial_guesses_elec_field[l] * E_elect_tem
            E_elect = E_elect + E_elect_tem
        E_elect = E_elect / 2

        E_surf = 0
        for w in range(self.monomer_types):
            E_surf = E_surf + self.surface_interaction[w] * (mono_con[w, 0] + mono_con[w, -1])

        ###
        E_kk = 0
        for w in list(self.w_type.values()):
            tem_kk2 = list(self.w_type.values()).copy()
            tem_kk2.remove(w)
            for ww in tem_kk2:
                kai = str(w) + str(ww)
                E_kk = E_kk + self.monomer_interaction[kai] * bulk_con[w] * bulk_con[ww]

        Total_free = -F_non_p - F_non_s - E_non + E_kai + E_elect + E_surf
        return F_non_p, F_non_s, E_non, E_kai, E_elect, E_surf, Total_free

def from_field_to_con(p1, p2, p3, p4, p5, p6, p7, p8, p9, p10, p11):
    example1 = SCF_v1(p1, p2, p3, p4, p5, p6, p7, p8, p9, p10, p11)
```

```
print('G(1,r):')
print(example1.G1_values())

print('\nGf(n,r)：')
print(example1.Gn_values()[0])
print('\nGb(n,r)：')
print(example1.Gn_values()[1])

print('\neach monomer concentration：')
print(example1.monomer_concentration()[0])
print('\npolymer concentration：')
print(example1.monomer_concentration()[1])

print('\npolymer average concentration：')
print(example1.average_concentration())

print('\npolymer bulk concentration：')
print(example1.bulk_volume_fraction())

print('\nfar error：')
print(example1.f_ar()[0])

print('\ngr error：')
print(example1.g_r())

print('\nhr error：')
print(example1.h_r())
```

```
    print('\nall error：')
    print(example1.total_errors())
    print(len(example1.total_errors()))

    print('\naverage distance')
    print(example1.get_average_distance())
    print(example1.get_average_distance_surface())

    print('\nhard-core')
    print(example1.f_ar()[1])

    print('\nTotal absorption')
    print(example1.get_total_absorption())

    print('\nFree energy')
print(example1.get_free_energy())
```

```
# Calculation of pI
def cal_iso_point(CT,DT,ET): #for three types of chared aminos acid residues
    x = 1
    last_result = 99
    while True:
        result = 1/(1+10**(x-10))*CT+1/(1+10**(x-6.75))*DT-1/(1+10**(4.5-x))*ET
        if abs(result) < abs(last_result):
            last_result = result
            x = x + 0.01
        else:
            break
    return float(Decimal(x).quantize(Decimal('0.00')))
```

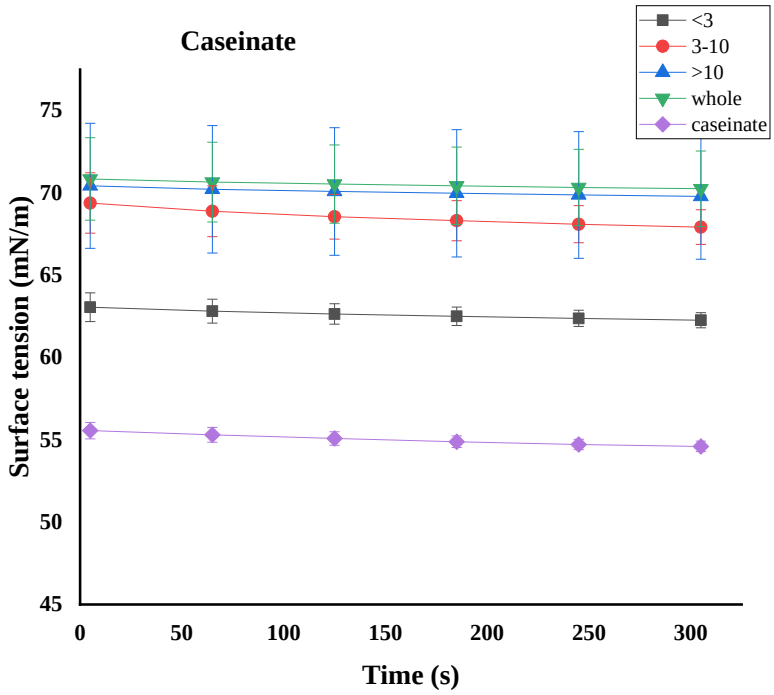
```
# Moving average method
from decimal import Decimal
import xlsxwriter
import re
import os

class moving_average_method():
    def __init__(self,Sequence,Period):
        self.Sequence = Sequence
        self.Period = Period
        self.AA_type                    =                    aaa={'P':'1',
'I':'1','G':'1','L':'1','V':'1','F':'1','A':'1','M':'1','W':'1','C':'1',
            'Q':'2','N':'2','S':'2','T':'2','Y':'2',
            'R':'2','K':'2',
            'H':'2',
            'E':'2','D':'2'}

    def split_by_period(self):
        output_total = []
        for i in self.Period:
            output = []
            start = 0
            end = int(i)
            total = len(self.Sequence)
            while end <= total:
                output.append(self.Sequence[start:end])
                start += 1
                end += 1
            output_total.append(output)
        return output_total
```

```
    def covert_Charcter_to_Num(self):
        total_output = []
        for i in self.split_by_period():
            output = []
            for ii in i:
                Num_sequence = ""
                for iii in ii:
                    Num_sequence = Num_sequence + self.AA_type[iii]
                output.append(Num_sequence)
            total_output.append(output)
        return total_output


    def get_average_points(self):
        total_output = []
        for i in self.covert_Charcter_to_Num():
            output = []
            for ii in i:
                hydrophobic_rate                                    =
float(Decimal(ii.count('1')/len(ii)).quantize(Decimal('0.000')))
                output.append(hydrophobic_rate)
            total_output.append(output)
        return total_output

    def insert_chart(self):
        all_data = self.get_average_points()
        max_values = []
        for i in all_data:
            max_values.append(len(i))
```

```
        max_value = max(max_values)
        distance = 1
        name_index = 0
        head_x = 0
        head_y = 0
        ###
        ds_x = 1
        ds_y = 0
        ys_x = 1
        ys_y = 1
        ###
        d_x = 1
        d_y = 0
        y_x = 1
        y_y = 1
        workbook = xlsxwriter.Workbook('Moving__Averagement_Results.xlsx')
        worksheet1 = workbook.add_worksheet(name="Moving averagement")
        chart_moving_averagement = workbook.add_chart({'type':
'scatter','subtype':'smooth'})
        for i in all_data:
            head_name = "Period = " + str(self.Period[name_index])
            worksheet1.write(head_x,head_y,head_name)
            name_index = name_index + 1
            head_y = head_y + 3
            for data in i:
                worksheet1.write(y_x,y_y,data)
                y_x = y_x + 1
                worksheet1.write(d_x,d_y,distance)
                d_x = d_x + 1
                distance = distance + max_value/len(i)
```

```
            chart_moving_averagement.add_series({
            'name':head_name,
            'categories':['Moving averagement',ds_x,ds_y,d_x,d_y],
            'values':['Moving averagement',ys_x,ys_y,y_x,y_y],
            })
            chart_moving_averagement.set_title({'name': 'Moving averagement'})
            chart_moving_averagement.set_x_axis({
            'major_gridlines':{'visible': False},"visible":False
            })
            chart_moving_averagement.set_y_axis({'name': 'Hdrophobicity',
            'major_gridlines':{'visible': False},
            'minor_gridlines':{'visible': False},'min':'0','max':'1'
            })
            chart_moving_averagement.set_legend({'position': 'bottom'})
            chart_moving_averagement.set_size({'width': 700, 'height': 400})
            y_y = y_y + 3
            y_x = 1
            d_y = d_y + 3
            d_x = 1
            ds_y = ds_y + 3
            ys_y =ys_y + 3
            distance = 1
        worksheet1.insert_chart('E2', chart_moving_averagement, {'x_offset': 0,
'y_offset': 0})
        workbook.close()
        cwd = os.getcwd()
        test = cwd + '\\Moving Averagement Results.jpg'
        print(f'The result of moving averagement has been saved to\n\n{test}')
```

Appendix B
Surface tension values versus time



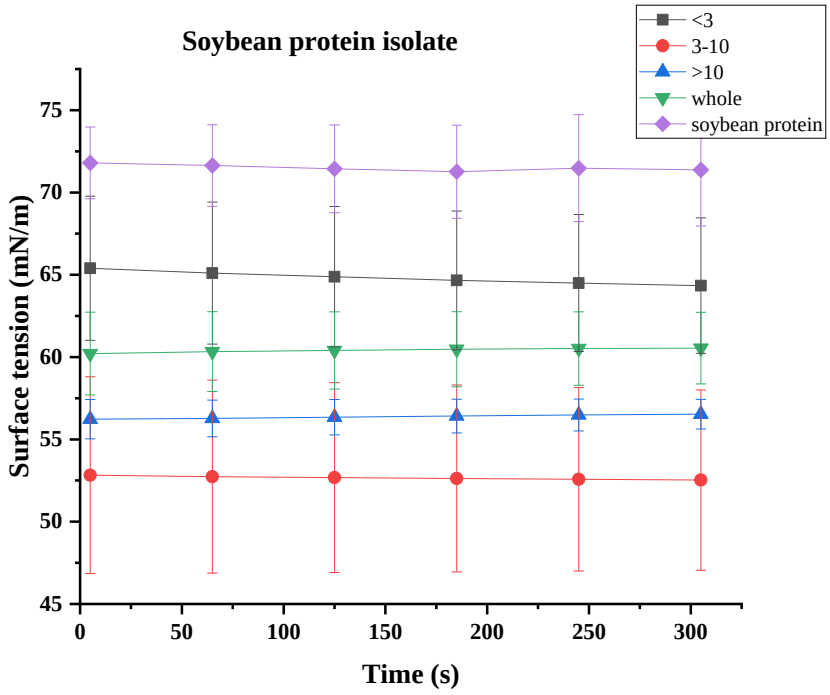


Figure B.1 Measurement of surface tension of caseinate and soybean protein isolate hydrolysed by alcalase at pH = 8 (10^{-3} wt.% protein solution in 0.05M phosphate buffer at pH 7)

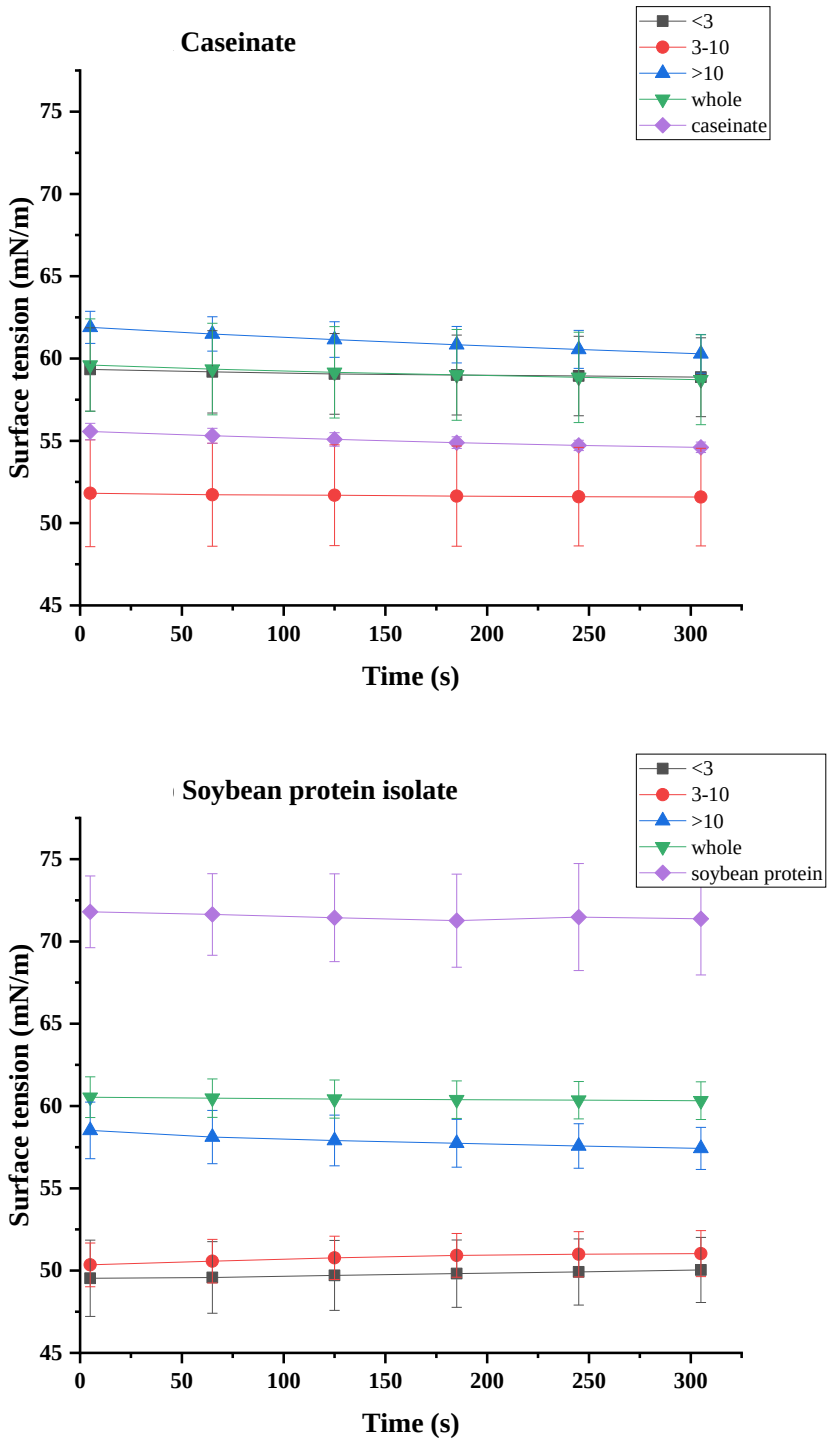


Figure B.2 Measurement of surface tension of caseinate and soybean protein isolate hydrolysed by pepsin at pH 1.3 (10^{-3} wt.% protein solution in 0.05M phosphate buffer at pH 7)

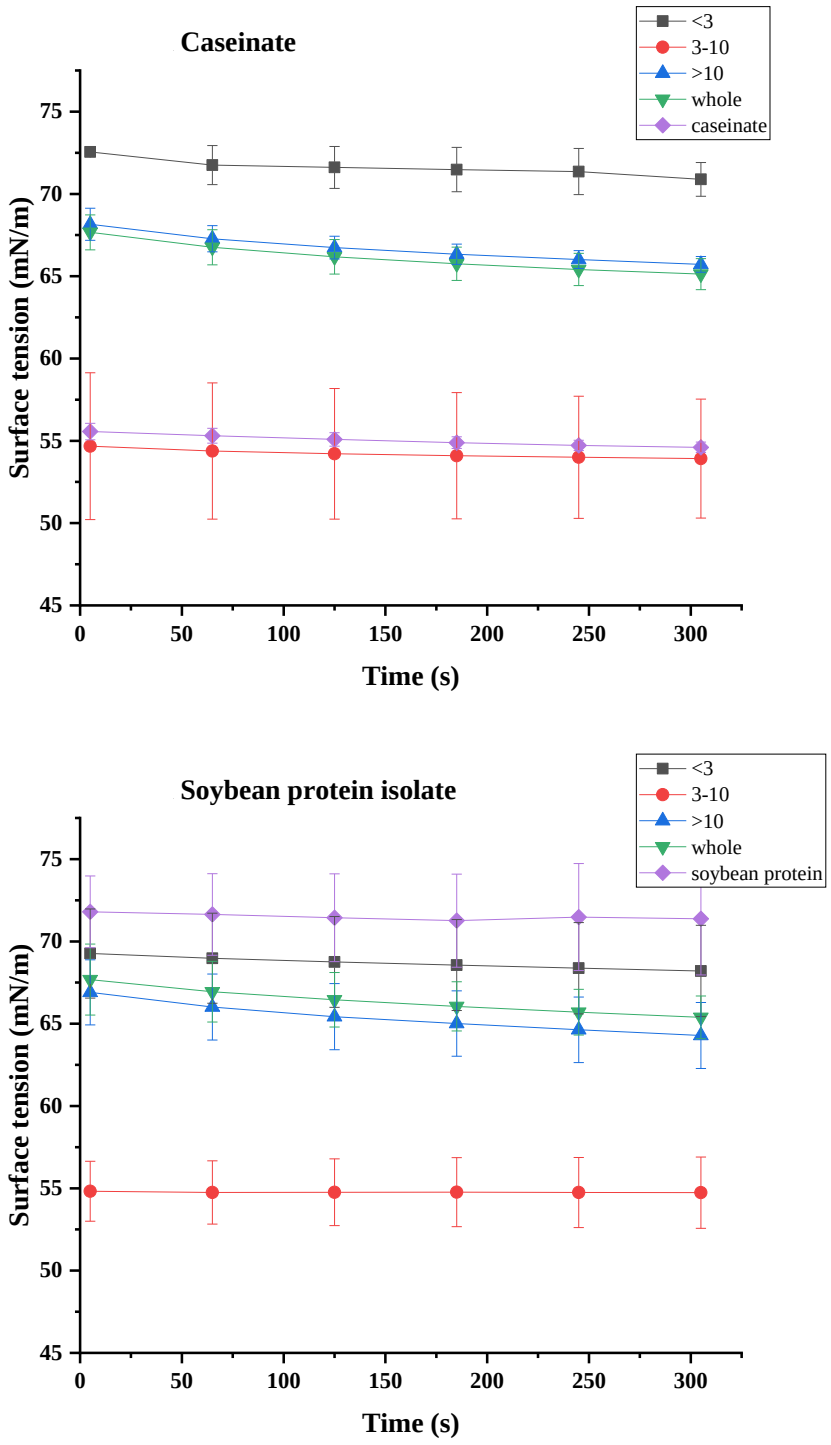


Figure B.3 Measurement of surface tension of caseinate and soybean protein isolate hydrolysed by pepsin at pH 2.1 (10^{-3} wt.% protein solution in 0.05M phosphate buffer at pH 7)