Modelling Protein Interactions in Solution

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Furio Surfaro

aus Rom, Italien

Tübingen

2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	12.07.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Frank Schreiber
2. Berichterstatter:	Prof. Dr. Martin Oettel

Contents

Zusammenfassung auf Deutsch

Die Wechselwirkungen zwischen Makromolekülen in Lösung können aufgrund ihrer hohen Komplexität schwer zu verstehen und zu charakterisieren sein. Proteine besitzen insbesondere eine hierarchische Struktur, bei der mehrere Aminosäuren verbunden sind, um ein Molekül zu bilden, das eine spezifische Struktur aufweist. Daher hängen die effektiven Wechselwirkungen zwischen Proteinen stark von der Struktur und der relativen Ausrichtung der Teilchen ab, was wesentlich komplexer ist als bei einfachen isotropen Teilchen mit einer kurz- oder langreichweitigen Wechselwirkung. In dieser Arbeit versuchen wir, die Wechselwirkungen zwischen Proteinen mithilfe der klassischen Thermodynamik und der Theorie der assoziierenden Flüssigkeiten zu modellieren. Der erste Teil der Arbeit bezieht sich auf das grundlegende Verständnis von Proteinen und der Thermodynamik von Phasenübergängen, mit einem kurzen Überblick über die Theorien von Gasen und Lösungen. Diese Rahmenbedingungen dienen als Referenz für das Verständnis des komplizierten Phasenverhaltens in Proteinsalzlösungen. Der zweite Teil der Arbeit ist im Wesentlichen eine Zusammenfassung der Ergebnisse meiner Doktorarbeit. Um Protein-Salzlösungen zu modellieren, ist ein Schlüsselparameter der zweite Virialkoeffizient. Dieser Parameter ist direkt mit dem mittleren Wechselwirkungspotential zwischen Teilchen in Lösung verbunden und erhellt das thermodynamische Verhalten des Systems. Der zweite Virialkoeffizient $B2$ wurde unter Verwendung eines alternativen Ansatzes berechnet, der die Gleichgewichtsthermodynamik und die McMillan-Mayer-Theorie der Lösung nutzt. Die Ergebnisse zeigen, wie die Werte des zweiten Virialkoeffizienten von Humanem Serumalbumin (HSA) und Bovinem Serumalbumin (BSA) mit den Werten übereinstimmen, die unter Verwendung von Kleinwinkel-Röntgenstreuung (SAXS) bestimmt wurden. Im zweiten Teil des Ergebnisteils haben wir das Ionen aktivierten 'patchy' Partikelmodell erweitert, um die Abstoßung zwischen 'patchy' Teilchen einzubeziehen. Dies ermöglicht eine bessere Rationalisierung der experimentellen Beobachtungen

für Protein-Salzlösungen. Wir schließen den Ergebnisteil mit einem kinetischen Modell ab, das effektiv zur Modellierung des Nukleationsdichteplots verwendet wurde, der durch Zählen der Anzahl von Kristallen auf der Glasplatte erhalten wurde. Aus diesem Modell können charakteristische Ratenkonstanten und zusätzliche Parameter extrahiert werden, die verwendet werden können, um den Gesamtmechanismus hinter der Proteinkristallisation besser zu verstehen.

.

Part I

Introduction

Chapter 1

Understanding proteins

In the last six centuries, since the development of modern science, humans were able to give answers to many complicated questions and to reach a fundamental understanding about the reality and the physical laws that govern our universe. Despite the astonishing results obtained on the comprehension of the world several questions are still present, especially regarding complex systems such as biological organisms. What is life? How complex molecules can interact in complex systems? In this regards proteins play a crucial role in understanding this processes since they are ubiquitous in biological systems and in life. Therefore, understanding the behaviour of such intricate molecules might be the key to understand much more complicated behaviours in living organisms such as, cell compartmentalization, signal transduction, disease development and progression etc.. Proteins are constituted of amino acids and present different levels of organization. In literature, the protein structure is normally classified in a hierarchical way. Aminoacids in protein are connected together through a peptide bond that is formed between the amide extremity of one residue with the carboxylic part of the other by a condensation reaction that eliminate water, as describe in Fig 1.1. [33]. These process is repeated several time during the

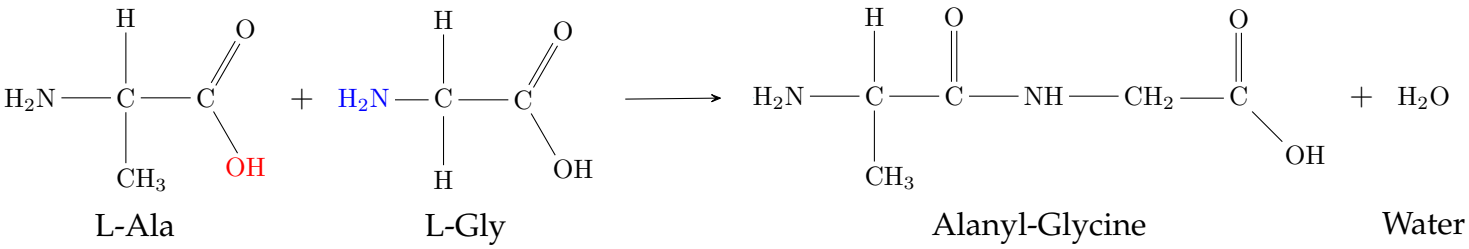

Figure 1.1: Condensation between aminoacids residues: the reaction proceeds by the addition of aminoacids and water elimination. after reaching a specific length the aminoacids in the sequence (primary structure) organize themselves in domains with specific secondary structures.

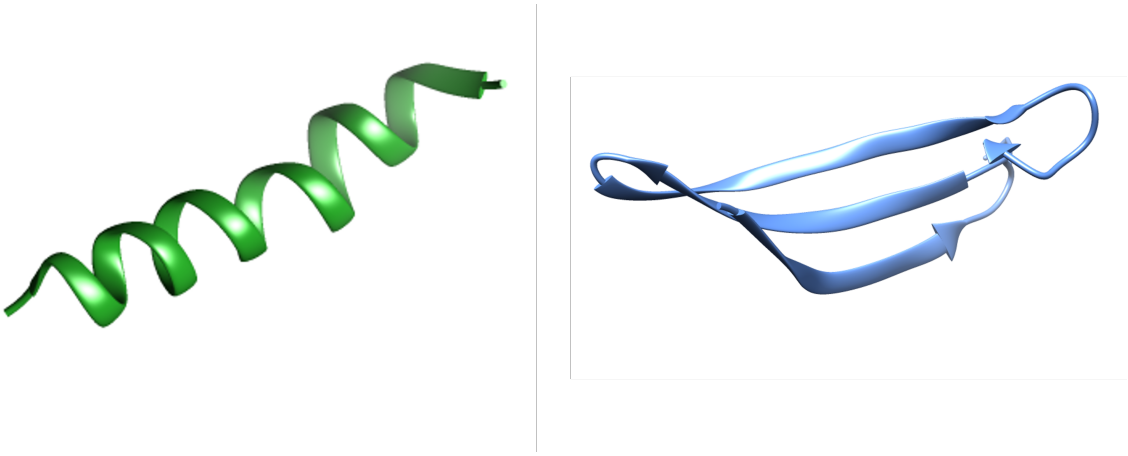

Figure 1.2: Cartooned representation of the two most common secondary structures in the proteins, α-elix (left) and β-sheet. The images are obtained from the software Chimera from reference PDB files[71, 116, 140]

.

translation process in cell's ribosomes. At the end of this process a polypeptide chain is formed, this first level of organization is called primary structure, and correspond to the sequence of aminoacids in the protein chain known as "backbone" [28]. A second level of organization arises when the aminoacids in the primary structure start to form intramolecular bonds. Depending on the type of residues involved in the intramolecular bonds and on the aminoacids sequence in the primary structure, different type of arrangements are possible, the most communes are called α-elix and β-Sheets that are represented as cartooned images in Fig 1.2. This further level of organization is called secondary structure [15, 28]. The secondary structure is stabilized by both intramolecular non covalent and covalent bonds such as hydrogen bonds and disulfur bridging, between distant residues on the protein chain. The type of secondary structure that can be generated strongly depends on the type and the organization of the aminoacids in the backbone, especially because the sterical repulsions between lateral groups on the side chains [99].

The primary and secondary structures have a specific organization in three dimensional space. This further level of organization is called tertiary structure. The tertiary structure possesses one aminoacidic chain with one ore more secondary structures and different protein domains that are structurally and intrinsically stables and functional independent regions. In Fig. 1.3 is shown a typical tertiary structure of two of the three proteins analyzed during the

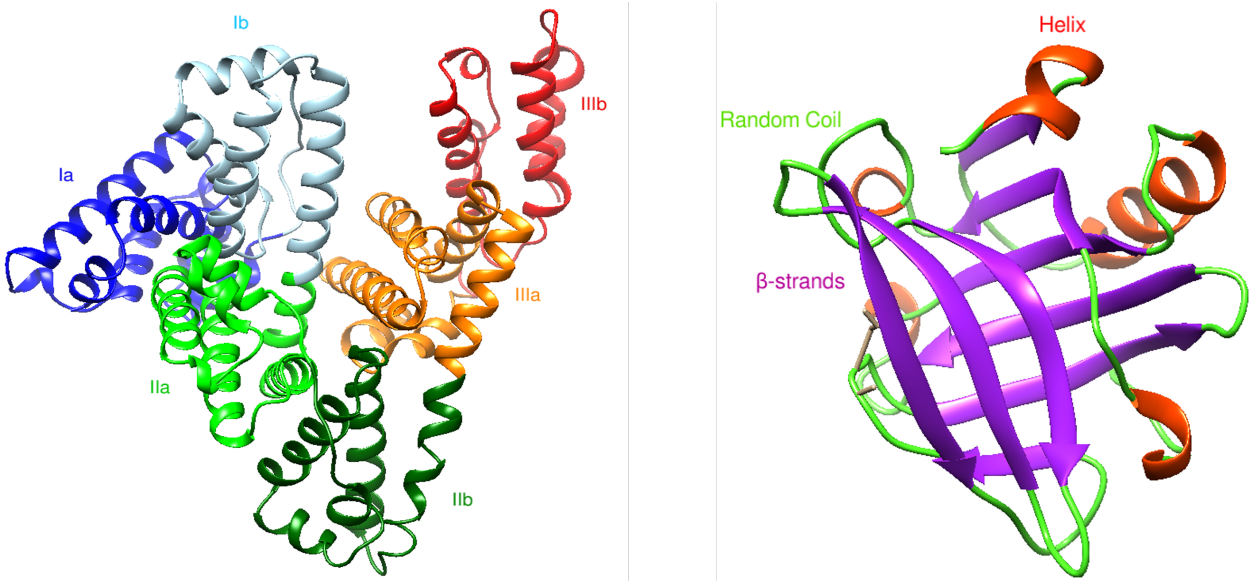

Figure 1.3: Cartooned representation of the tertiary structure of two of three globular proteins employed in the experiments covered in this thesis, Human Serum Albumin abbreviated as HSA on the (left) with the specific domains and β-lactoglobulin abbreviated as β-Lg on the right with the secondary structure motifs highlighted. The images are obtained from the software Chimera from reference PDB files [71, 116, 140]

.

experimental part of my doctoral studies [71, 116, 140].

In general proteins can be further classified in three different classes (globular, fibrous and membrane) considering their role and conformation inside the cells. For example Human Serum Albumin (HSA), Bovine Serum Albumin (BSA) and β-Lactoglobulin (β-Lg) are so called globular proteins for their specific globular conformation and solubility in water, they usually behave as carriers, messengers or enzyme inside the body, whereas other proteins such as collagen, keratin, elastin and fibirin are called structural proteins and are generally only slightly soluble in water and organize themselves as elongated helix with structural functions in mammals, where they are present in the connective tissue, bones, cartilage, hairs and skin [29, 105]. Another class of protein is so called membrane proteins, this class of protein are in abundance and they constitute approximately the 25% of all proteins. [2] Those proteins are usually working as receptors to connect the interior and exterior of the cells conducting the signal or to allow the exchange of ions between regions inside and outside the cell. While for globular and structural proteins the structural characterization is relatively simple, membrane proteins are particularly difficult to characterize and it is

still a wide open field of research. This is due to their particular constitution [42]. In fact, they are floating between the phospholipidc double layer exposing some of the side chains to the non polar chains of the phospholipids and other parts interacting with the polar head of them. Therefore, is really difficult to crystallize without destroying the rich network of interactions and their conformations and even in this case the high flexibility of some regions make this process extremely complicated [13, 42]. Most of the time this proteins are trapped in detergents micells in the attempt of crystallizing them, these micells are extremely sensitive to the environment and can be easily disrupted by small changes in pH and ionic strength [42, 79]. As aforementioned, the proteins structures are hierarchically classified. So far, the definition of the protein tertiary structure was used to introduce from the protein arrangement and chemistry the further classification based on the biological function (globular proteins, structural proteins, membrane proteins). However, proteins with their tertiary structure can arrange in larger supramolecular entities to perform biological functions. A protein that is composed by two or more subunits with a complete tertiary structures possesses an additional level of organization that is called quaternary structure [61]. An example of a protein that possesses a quaternary structure is Hemoglobin, where four different sub units (α_1, α_2 and β_1, β_2) are clusters together in order to perform the biological function [67]. In case of Hemoglobin, the sub-units contain a heme group that coordinate a Fe^{2+} ion, this groups are called prosthetic groups. A prosthetic group is a region of the protein with a specific biological function that is note made by a polypeptide chain. An example of prosthetic groups are organic molecules and metal ions. Those groups, help proteins to perform their biological functions. In this regards, the biological function of the proteins is intrinsically connected with their structures as well as to the presence of prosthetic groups [61, 67, 90]. In this chapter, we want to highlight the need to elucidate the importance of the protein structure as well as the experimental available to determine it and the complexity of the interaction network to better introduce protein-protein interaction mediated by trivalent ions, which is the core of this doctoral thesis.

Solution	Colloids	Suspension
10^{-12}-10^{-9} m	10^{-9}-10^{-6} m	$> 10^{-6}$ m
Homogeneous	Heterogeneous	Heterogeneous

Table 1.1: Classification of mixtures based on the relative size of the solute dissolved in the medium.

1.1 Proteins as colloidal system

In order to properly introduce protein interactions in solution it is necessary to step back to colloid chemistry. When a substance is dissolved in a medium such as water different mixtures are possible and these are classified based on the size of the particles inside the medium as explained in Tab. 1.1. In this regards, while in a solution the particles of solute are homogeneously dispersed in such a way that only one phase is present, in colloids we have an heterogeneous mixture in which a phase (solute) is dispersed in an other phase (medium) [51, 53, 106]. The properties of colloids are particularly intriguing for the ubiquitous presence in nature and for the possible application in different areas of research, such as medicinal chemistry, drug delivery and food chemistry [41, 91]. More fundamental question can be answered by studying colloidal systems, such as nucleation and liquid-liquid phase separation. Due to the size of the particles dispersed in the medium colloids interact with visible light, when UV light pass through a colloidal system it is scattered by the particle of solute, since they have comparable length, resulting in the so called "Tyndal effect" or "Tyndal scattering" in which the system appears to be colored or opalescent. For this reason, it is possible to use scattering experiments to get important information about the particles in solution, i.e. size of the molecules, shape, molecular weight, degree of polymerization and the type and magnitude of interactions [69, 135, 139]. When the particles in the colloidal dispersion strongly interact with each other, the colloidal dispersion become unstable, forming bigger clusters that can flocculate or coagulate in solution, creating a suspension. At this stage, these large clusters start to be affected by gravity and sediment at the bottom of the solution or they can be separated mechanically from the medium by centrifuging or filtration. An interesting theoretical framework that elucidate this behavior and more in general the one given by lyophobic colloids such as proteins is given by the DLVO theory and discussed in the next section.

1.2 General behaviour of lyophobic colloids: DLVO theory

The DLVO theory, named after the scientists Derjaguin, Landau, Verwey, and Overbeek, is a theoretical framework used to explain the stability and interaction forces between colloidal particles suspended in a liquid medium. It is regard as a milestone in colloidal science and one of the first theory adopted in field of colloids [23, 91]. In the DLVO theory the interaction between particles in solution are modelled considering them as charged spheres surrounded by an electric double layer. The particles interact via a potential that is given by two contributions [23].

$$V_{DLVO}(d) = V_{VdW}(d) + V_{SC}(d) \tag{1.1}$$

The first term is the Van der Waals potential, it is an attractive contribution that depends on the overall shape of the molecules and on the Keesom, Debye and London dispersion forces. A Van der Waals force is usually scaling as $-\lambda/r^6$ where λ is a constant that takes into account all the different types of Van der Waals interactions. The rationalization of the effective Van der Waals interactions between colloidal particles of different shapes and dimension is due to Hamaker. A potential between two spherical particles is given in his well known paper from which the following relations are adopted [45]. For two spheres interacting via Van der Waals potential the resulting expression is:

$$V_{VdW}(d) = -A\frac{1}{12}\left\{\frac{y}{x^2+xy+x} + \frac{y}{x^2+xy+x+y} + 2\ln\frac{x^2+xy+x}{x^2+xy+x+y}\right\} \tag{1.2}$$

Where $A = \pi^2 q^2 \lambda^2$ is the Hamaker constant, and q is the number of interacting molecules per unit of volume. The variables x and y are defined as:

$$x = \frac{d}{2R_1} = \frac{d}{\sigma_1} \quad ; \quad y = \frac{\sigma_2}{\sigma_1} = \frac{R_2}{R_1} \tag{1.3}$$

Where R_1 and R_2 are the radii of the two spheres and $\sigma = 2R$. In the limit of equally sized particles, Eq. 1.3 become

$$V_{VdW}(d) = -A\frac{1}{12}\left\{\frac{\sigma^2}{d(d+2\sigma)} + \frac{\sigma^2}{(d+\sigma)^2} + 2\ln\frac{d(d+2\sigma)}{(d+\sigma)^2}\right\} \tag{1.4}$$

This expression is particularly useful in the case of proteins-salt solutions especially in the case of monodisperse solutions in which the size of the interacting particles is in good approximation the same [104]. The second and last contribution on the DLVO potential is given by the Screening Coulomb potential. This potential is a repulsive contribution that arise from the screened charge around the spherical particles as the case of proteins in electrolytes solutions. In fact, when an 1:1 electrolyte, i.e. NaCl, KBr..., is dissolved in a polar medium such as water, an electrostatic screening of the charge occur by the medium that decrease the interaction between charge ions through the dielectric constant. When globular (spherical) proteins are in electrolyte solution, a double layer of electrolytes around the proteins surfaces that screens the direct charge-charge repulsion by an exponential factor is formed. This approach is based on the Poisson-Boltzmann model, which assumes a spherically symmetric distribution of charge around the particle. However, it is important to note that this assumption may not always hold true for proteins, as their charge distribution can be complex and non-spherical [68, 89]. The Screening Coulomb interaction potential is given, in the case of two spherical particles, by the following equation:

$$V_{SC}(d) = \frac{z^2 l_b}{\beta(1+k\sigma/2)^2} \quad \frac{exp(-k(d-\sigma))}{d} \tag{1.5}$$

where z is the charge of the microion, l_b is the Bjerrum lenght, β the inverse temperature, σ the particles diameter and d the distance between them. The parameter k is the inverse of the Debye lenght and it is defined as:

$$k^{-1} = D_l = \left[\frac{k_b T \epsilon}{N_A e^2 I}\right]^{1/2} \tag{1.6}$$

This part of the potential explicitly shows the dependence of the repulsion on the salt concentration, since the entity of the repulsion is going as the inverse square root of the ionic strength $I = \sum_i c_i z_i^2$. Therefore, by increasing the ionic strength the Debye length decreases and the attractions between particles increase, as depicted in Fig. 1.4. Therefore, it is possible to promote protein aggregation increasing the salt concentration.

If the salt concentration is high enough, the maximum of the potential vanish and the system is dominated by the Van der Waals attractions. The minimum concentration of salt necessary to promote the attractions is called critical coagulation concentration or CCC[81]. and corresponds to the point where the total

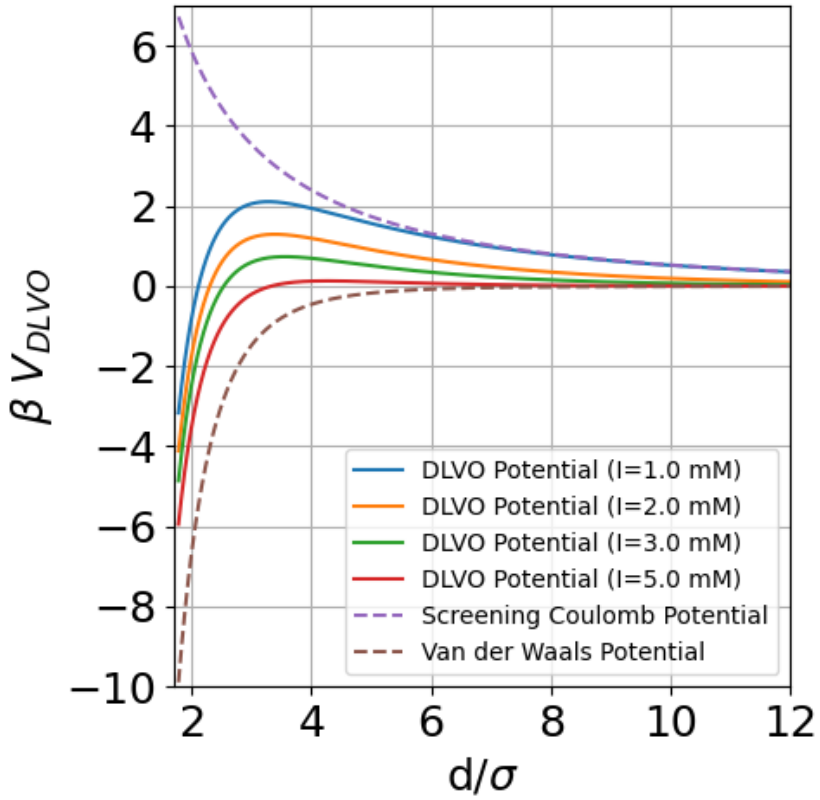

Figure 1.4: DLVO potential as a function of the normalized distance: in order to aggregate the particles need to overcome the maximum of the DLVO potential, this is possible at low separation distances and a relative high salt concentration. Increasing the ionic strength decrease the double layer repulsion and consequently promote aggregation.

potential and the first derivative with to respect the distance are both equal to zero:

$$\beta\, V_{DLVO}(d) = 0 \quad ; \quad \beta \frac{\partial V_{DLVO}(d)}{\partial d} = 0 \tag{1.7}$$

Considering a symmetric electrolyte the dependence of the CCC on the salt concentration is proportional to $1/z^6$ where z is the charge of the electrolyte. Therefore, for highly charged ions the amount of salt required to reach the CCC is lower. More complicated dependencies of the CCC on the charge of the electrolytes are used to describe in good approximation the behaviour of not symmetric electrolytes [62, 68, 121]. Despite the qualitative and partially quantitative description of colloidal systems and protein salt interactions, the DLVO theory fails to explain some trends observed in protein-salt systems and colloids. For example, the height of the energy repulsive barrier, monotonically decreases by increasing salt concentration. However, in many other systems,

such as protein solutions with multivalent salts, the trend of the interactions varying the salt concentration, shows a non monotonic behaviour, that lead to specific ions effects and to re-entrant condensation, in which the particles aggregates only at intermediate salt concentrations and repel each other as the ionic strength increases [12, 35, 36, 69, 71, 73, 118, 135, 138–140]. These effects are better discussed in the next sections and chapters. Since the description of specific ions effects is neglected treating the double layer as uniform charge distribution using the Poisson-Boltzmann model, other theories that are taking into account the specificity of the protein-surface and its interactions with the ions should be considered. Hydration forces, or highly repulsive forces at really short distances are usually neglected in DLVO as well as entropic forces such as depletion interactions [72, 117, 123].

1.2.1 Non-DLVO interactions: The Asakura–Oosawa model

As aforementioned, the DLVO theory is a milestone in the understanding of colloidal systems. However, different important interactions in the description of the colloidal solutions are neglected. An important non-DLVO interaction is described by the Asakura–Oosawa model [4, 10, 82, 113]. In this model, the interaction potential between two proteins arise from the excluded volume. To enhance realism while maintaining physical meaning, let's envisage a scenario involving, for instance, a protein within a bulk solution. Consider a situation wherein two "large" spherical particles are submerged in a medium containing significantly smaller spherical particles [83]. The big particles in solution interact via a simple hard sphere potential:

$$V_{HS}(d) = \begin{cases} \infty, & \text{if} \quad d < \sigma_b \\ 0, & \text{if} \quad d \geq \sigma_b \end{cases}. \tag{1.8}$$

When the bigger particles and the smaller particles are separated they exclude a part of the volume that is not accessible to the smaller spheres due to the hard sphere potential, the excluded volume contribution read as:

$$v_{excluded} = \pi \frac{(\sigma_b + \sigma_s)^3}{6}, \tag{1.9}$$

when the larger particles approach each other closely, they exclude from the previously excluded volume a region equivalent to the overlapping volume of the larger spheres. Consequently, the diminished contribution to the excluded volume is:

$$v_{excluded}^{red}(d) = \pi\frac{(\sigma_b + \sigma_s)^3}{6} + \frac{2\pi l^2}{3}\left[\frac{3(\sigma_b + \sigma_s)}{2} - l\right] \tag{1.10}$$

where σ_b and σ_s are the diameters of the big and small spheres respectively and l is the width of the overlapping lens formed by the spherical caps and defined as:

$$\frac{(\sigma_b + \sigma_s)}{2} - \frac{d}{2} \tag{1.11}$$

where d is the center to center distance. The variation of the Helmholtz free energy is directly connected with the force acting between two particles. Assuming that the system is diluted enough, It is feasible to treat our system by considering the interactions between small spheres, as well as the interactions between small and large spheres, as negligible. The total differential for the Helmholtz free energy is $dF = SdT - PdV$ and at constant temperature $dF = -PdV$. For a non interacting system with a hard core, it is possible to rewrite the expression as:

$$dF = -nk_bT\frac{dV}{V - nb}, \tag{1.12}$$

where V is the total volume of the system and b is the excluded volume. As aforementioned, we have two possible cases for the excluded volume one when the big and the small spheres are apart and another where the big spheres are overlapping so, depending on the relative distance between two particles, we have two possible different values for the parameter b:

$$b = \begin{cases} v_{excluded}, & \text{if} \quad d \geq \sigma_b + \sigma_s \\ v_{excluded}^{red}, & \text{if} \quad d < \sigma_b + \sigma_s \end{cases} \tag{1.13}$$

After integration it is possible to linearize the logarithm since $V >> b$:

$$\frac{F - F_{id}}{nk_bT} = \frac{\Delta F}{nk_bT} = ln\left(\frac{V - nb}{V}\right) \approx -\rho b, \tag{1.14}$$

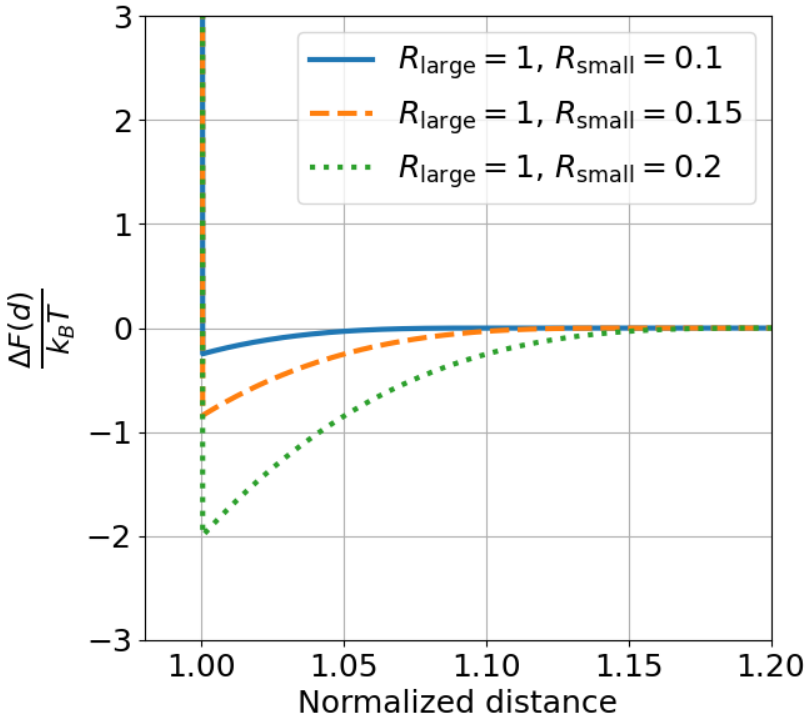

Figure 1.5: Sketch of the depletion interaction potential between large and small spheres: By increasing the radius of the small spheres the depth of the attractive region increase as well as the range of the interactions due to the increased excluded volume contribution.

By replacing the density with the osmotic pressure we obtain the final expression of the potential:

$$\Delta F = -\Pi b \tag{1.15}$$

The potential is therefore:

$$\Delta F = \begin{cases} \infty, & \text{if} \quad d \leq \sigma_b \\ -\Pi v_{red}, & \text{if} \quad d \geq \sigma_b + \sigma_s \\ -\Pi v^{red}_{excluded}(d), & \text{if} \quad \sigma_b < d < \sigma_b + \sigma_s \end{cases} \tag{1.16}$$

By combining the two contributions at different distances we obtain the final expression:

$$\Delta F(d) + \Delta F_{ls} = -\Pi v^{tot}_{excluded}(d) - \Pi v_{red} = -\Pi (v^{tot}_{excluded}(d) - v_{red}) = -\Pi \Delta v(d) \tag{1.17}$$

where $\Delta F(d)$ is the value of the potential in the depletion interaction regions, and ΔF_{ls} is the value of the potential at larger separation distances. The first

term depends on the relative distance between particles while the second term is a constant and independent of the distance, it can be seen as the constant interaction energy due to the hard spheres excluded volume. An example of the depletion potential is given in Fig. 1.5 in which the dependence of the potential on the radius of the small sphere is also shown. These kind of interactions are regarded as entropic forces, since the system is constituted by particles that have only repulsive interactions. In this framework, the attractions between larger colloids is due to the gain in configuration entropy of the small spheres when the bigger particles get close. In this short subsection, I want to emphasize the importance of the depletion effects, these contributions have a relevant impact on the behaviour of many colloidal systems, therefore they should be carefully considered and taken into account in the description of the interactions between colloidal particles.

1.2.2 Non-DLVO interactions: salt effects and Hofmeister series

One way to trigger interactions between proteins is by using a "precipitation agent" that promote the protein-protein interactions. Salts, that are strong electrolytes, are normally employed for such a scope. However as discovered by Hofmeister different salts posses different power in precipitate proteins, those effects are normally not included in the DLVO forces since it treats the ions as uniform electronic cloud in the double layer around the protein. However, different salts interact differently with the solvent and the surface of the proteins. In this regards, Hofmeister developed a series based on the effectiveness of cations and anions in trigger protein aggregations. As orginal formulated by Hofmeister in his original manuscript dated 1888 entitled "Zur Lehre von der Wirkung der Salze", [49, 50, 65] the series reads as:

$$C_4H_4O_6^{2-} > SO_4^{2-} > HPO_4^{2-} > C_3H_5O(CO_2)_3^{3-} > CH_3COO^- > HCO_3^- > Cr_2O_7^{2-} >$$
$$Cl^- > NO_3^- > ClO_3^- > SCN^-$$

In this series anions are classified based on their capability to precipitate proteins. The anions are ordered in decreasing precipitation power. The effects seems to be stronger for anions than cations. Despite that, a similar series for the cations can be written [55, 58]:

$$NH_4^+ > Cs^+ > Rb^+ > K^+ > Na^+ > Ca^{2+} > Mg^{2+} > Li^+$$

In the traditional Hofmeister series ions are divided in two categories based on the stabilizing or destabilizing effect on water hydrogen bonds network. Ions that stabilize the hydrogen bonds network and the water structure are called kosmotropes while ions that disrupt it are called chaotropics. Kosmotropes are strongly hydrated in solution while for chaotropics is the opposite [92, 129]. In the previous series, anions on the left are considered to be kosmotropes, while for the series of cations the elements on the left are chaotropics. Clearly, the effect on the solvent is translated on an effect on the solute-solute interactions. In case of proteins, kosmotropes anions ordering the water structure are stabilizing the native conformation of the proteins and the hydrophobic interactions, promoting aggregation. While for cations, an inverse effect is observed in which they stabilize the protein tertiary structure, promoting aggregation [58, 88]. However, the Hofmeister series might be reverted depending on the pH. When the solution has a pH $>$ pI where pI is the isoletric point of the protein, the series is generally respected, while at pH $<$ pI the negatively charged proteins follow the reversed series [111]. This is an indication of more specific interactions between ions and substrates. Recently, several groups provide detailed studies on such effects and the general classification of ions in kosmotropes and chaotropics is considered partially outdated. For example, in some recent papers it was found that the bulk solvent interactions are not affected by the salt, while the specific hydration shell of the protein change [86, 88]. Regarding cations, many of the observed effects can be explained considering the interactions between ions and negatively charged residues such as carboxylate groups on the proteins surface. While for anions, the effects seem to be more complicated, some studies reported that chaotropes, e.g., SCN^- and I^- interact with the amide nitrogen and the adjacent α-carbon. It is interesting to notice that those anions promote unfolding and changes in the conformations of the protein native structure, increasing the overall solubility of the proteins [142]. However, a comprehensive explanation of all the effects is not yet reached and more insights are needed to really understand the origin of these effects.

1.2.3 Non-DLVO interactions: salt effects and acidity of the cations

While some of the effects of the Hofmeister series can be rationalized on the basis of hydrated ion paars in solution,[16, 17, 59] other effects, especially regarding multivalent cations are more difficult to be taken into account. In fact, multivalent cations are Lewis acids and might change the pH of the solution, inducing changes in proteins conformation [3, 85, 101, 136]. An example of change in protein conformation due to the pH is given in Fig. 1.6 . The effect on pH of multivalent cations follows an empirical correlation in which their acidity is proportional to the square of the valency divided the radius of the cation in pm [136]. For not or only slightly electronegative cations, the relation adopted is the following:

$$pK_a = 15.14 - 88.16\,pm\left(\frac{Z^2}{r}\right) \tag{1.18}$$

While for highly electronegative cations ($\chi > 1.5$) it is necessary a correction:

$$pK_a = 15.14 - 88.16\,pm\left(\frac{Z^2}{r} - 0.096(\chi - 1.5)\right) \tag{1.19}$$

therefore, the acidity (lower pK_a) is higher for small, highly charged and electronegative cations, where χ is the Pauli's electronegativity.[136] In fact, these ions form hydrated complex in solution that can react with water forming hydronium ions:

$$[Me(H_2O)_6]^{3+} + \mathrm{H_2}O \rightarrow [Me(H_2O)_5\mathrm{OH}]^{2+} + \mathrm{H_3}O^+$$

In this regards metals are an intriguing way to tune the properties of proteins in solution and their interactions. However, in order to properly understand protein interactions it is important to take into account all the variables that might effect their behaviour in solution.

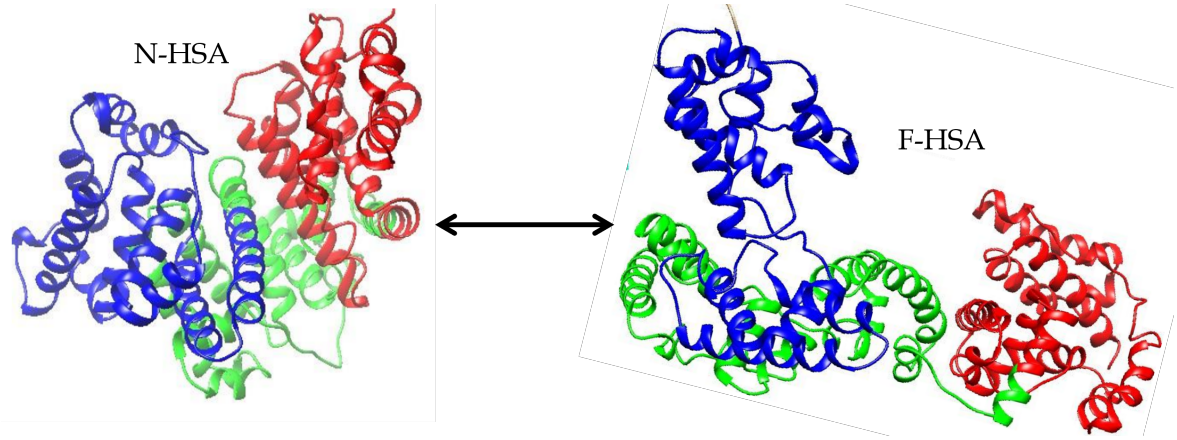

Figure 1.6: Change in conformation of HSA due to the change on pH: On the left side the HSA native conformation, that is favorable in physiological conditions. When the pH is in the range 5-3.5 the equilibrium between the two conformations is shifted towards the fast structure (F-HSA). [5]

1.3 Emerging theories in protein salt solutions: ion-activated attractive patches model

In recent years, new theories have been developed to gain a better understanding of protein salt solutions. One successful theory, that has been used to describe protein-protein interactions in presence of trivalent salts such as YCl_3 is the ion-activated attractive patchy particles model.[103] Protein in presence of trivalent salts usually shows an interesting and rich phase behaviour that include, liquid-liquid phase separation (LLPS), crystallization, and reentrant condensation [101, 135, 137, 141]. These effects, that are characteristic of solutions of proteins with trivalent ions are not captured in the frameworks of DLVO and other classical theories. In the patchy particles model the attractions between particles is given by:

$$V_{PP}(d_{12}, r_{12}^{ij}) = V_{HS}(d_{12}) + \sum_{i=1}^{m}\sum_{j=1}^{m} V_{pp}(r_{12}^{ij}) \tag{1.20}$$

where V_{HS} is the hard sphere potential and $V_{p-p}(r_{12}^{ij})$ is the potential between patchy particles and is given by :

$$V_{pp}(r_{12}^{ij}) = \begin{cases} \epsilon_{pp}, & \text{if} \quad r_{12}^{ij} < d_c \\ 0, & \text{otherwise} \end{cases} \tag{1.21}$$

ϵ_{pp} is an interaction parameter assuming a square well interaction potential between the patches. In our framework the particles can interact only if the patches i and j on the particles 1 and 2 have a distance r_{12}^{ij} smaller than the critical distance d_c, otherwise the interactions between them are considered to be zero. In this model, patchy particles are immersed in a reservoir of salt ions, that are approximate as an ideal gas of non interacting particles. The effect of the salt is to bind the empty patches on the surface of the particles. The probability to fill the empty spot on the surface of the particle is given by a Fermi-like distribution in the grand canonical ensemble:

$$\theta(c_s^r) = \frac{1}{1 + \exp\left(\beta\epsilon_b - \beta\mu_s^r(c_s^r)\right)} \tag{1.22}$$

where ϵ_b is the binding energy between the ions and the patch on the protein's surface, and $\mu_s^r(c_s^r)$ is the chemical potential of the salt in the reservoir, that is given by the chemical potential of an ideal gas and express as $\beta\mu_s^r(c_s) = ln(c_s^r)$. In this model, the overall probability to have m patches occupied with i ions is connected with the binding probability and follow a binomial distribution:

$$p(m, i) = \theta^i q^{m-i} \binom{m}{i} \tag{1.23}$$

where $q = (1 - \theta)$ is the no binding probability. The overall interaction energy between ions activated patchy particles is given by different contributions and is expressed by the following relation:

$$\beta\epsilon_{pp} = \beta\epsilon_{uu}(1-\theta)^2 + 2\beta\epsilon_{uo}\theta(1-\theta) + \beta\epsilon_{oo}\theta^2 \tag{1.24}$$

The first term accounts for interactions between unoccupied patches on the particles' surfaces, while the second term is negative and represents an attractive contribution due to interactions between filled and empty patches (ion bridging). The last term is a repulsive contribution resulting from interactions between patches filled with ions. In the previous model, the first two terms were set

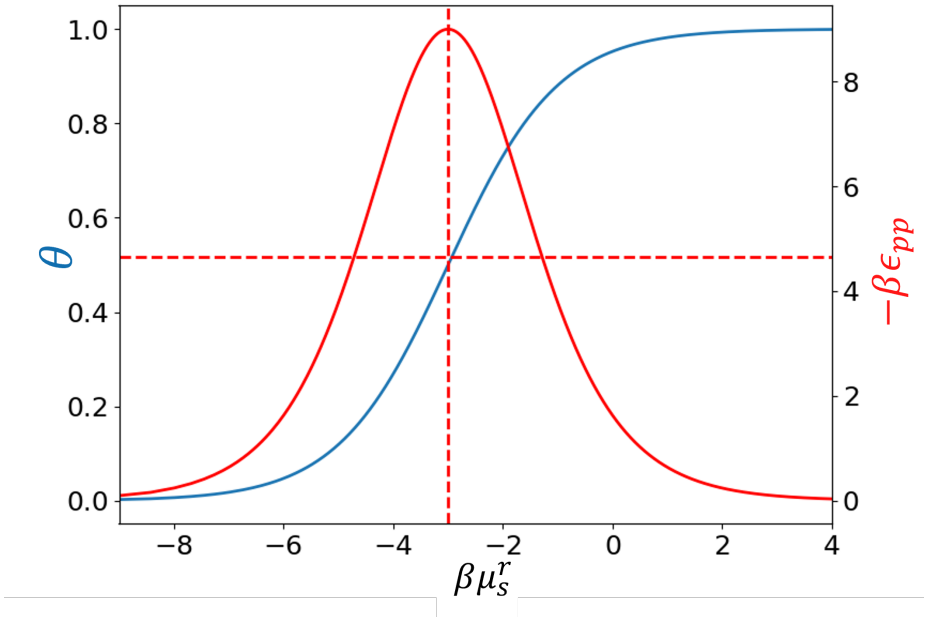

Figure 1.7: Chemical potential of the salt in the reservoir as a function of the binding probability and interaction energy. In this framework, the interactions energy ϵ_{uu} and ϵ_{oo} are considered to be zero and the maximum of the attractions are at $\theta = 0.5$. By increasing the salt concentration and therefore the number of occupied patches, slowly increase the interaction energy ϵ_{pp}.

to zero. In this work, we further extend the theory to include the effect of the repulsive terms. The theory makes use of the Wertheim theory of associating fluids for the calculation of the relative phase diagram of the protein-salt mixture, especially regarding the calculation of the liquid-vapor coexistence region. An important quantity that connect the theory with the experiment is the salt concentration c_s, which is the sum of the ions bounded on the surface of the protein with the free ions in solution:

$$c_s = m\theta\rho + c_s^r \left(1 - \eta \left(1 + \frac{R_s}{R_p}\right)^3\right) \tag{1.25}$$

where m is the number of patches, R_s is the radius of the salt and R_p the radius of the particle. The quantity ρ is the density of the particles and is connected with the packing fraction η by the relation $\eta = \frac{4}{3}\pi R_p^3 \rho$. In this thesis we explore the effect of different parameters on the liquid-vapor equilibrium. For the impatient reader the results section of the patchy particle model can be found in Chapter 5.

Part II

Gas theories and phase transitions

Chapter 2

Gas theories and phase transitions

This part of the thesis is focusing on the evolution of gas theories that are used as a reference for phase transition in more complicated system such as protein solutions with electrolytes. In fact, in some conditions the behaviour of protein solutions can be rationalized making use of theoretical frameworks derived from the behaviour of gaseous systems. In protein systems, we observe a wide range of different phenomena and phase transitions, such as crystallization and liquid-liquid phase separation. Understanding crystallization of protein is extremely challenging because the interactions between them depends on several parameters, such as pH, temperature and ionic strength. Liquid-liquid phase separation is an important process to comprehend different phenomena in living systems, such as compartmentalization, signal regulation and pathogenesis of diseases. Therefore, this chapter will focus on the evolution of the physical theories, from the ideal gas law to the more sophisticated theory of protein salt systems.

2.1 Thermodynamic ensembles

A key concept in statistical mechanics is the idea of statistical ensemble. An ensemble is a collection of different copies of a system of particles that have the same macroscopic conditions. The idea of a statistical ensemble is a key bridge between the microscopic and macroscopic descriptions of a physical system. While the behavior of individual particles in a system is governed by microscopic laws, statistical mechanics deals with the collective behavior of a large number of particles. The ensemble approach allows us to make statistical predictions about the macroscopic properties of a system without having to track the exact microscopic details of each particle [94].

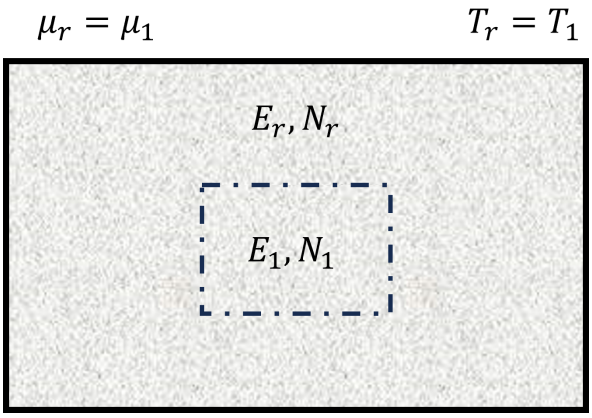

Figure 2.1: Schematic figure of the grand canonical ensemble in which the smaller box (the system) is coupled with the reservoir (larger box).

2.1.1 Grand canonical ensemble

In the grand canonical ensemble a system having a certain amount of energy E_1 and number of particles N_1 is in equilibrium with a particle reservoir having energy E_r and number of particles N_r. The particles are free to fluctuate from the reservoir to the system and viceversa. The system is in thermal and in chemical equilibrium with the reservoir where (T, V, μ) are fixed, however the particles N and the energy E are free to fluctuate between the system and the reservoir. For the system, the total energy and the total number of particles are conserved. Therefore, for any given system with energy E_1 and number of particles N_1 the total energy is $E_{tot} = E_r + E_1 = cosnt$ and the total number of particles is $N_{tot} = N_r + N_1 = const.$ In a thermodynamic ensemble, the properties of a system in thermodynamic equilibrium are derived from the partition function. In the grand canonical ensemble the partition function is defined as:

$$\mathcal{Z}_{gc}(T, V, \mu) = \sum_{N=0}^{\infty} \frac{1}{N!\,(2\pi\hbar)^{3N}} \int e^{-\beta(H(r,q)-\mu N)}\; d^{3N}r\, d^{3N}q \tag{2.1}$$

where H is the Hamiltonian function of the system, and depends on all the particle positions r and momentous p. From the partition function it is possible to calculate all the relevant thermodynamic variables of interest. For example, for the grand canonical ensemble the relevant thermodynamic potential is the grand potential also known as Landau potential and is given by the following

relation:

$$\beta\Omega(T,V,\mu) = -ln(\mathcal{Z}_{gc}(T,V,\mu)) \tag{2.2}$$

The partition function is therefore, directly connected with the grand potential that is defined from the classical thermodynamic theory as:

$$\Omega(T,V,\mu) = U - TS - \mu N = F - \mu N \tag{2.3}$$

For an homogeneous system at fixed external pressure and temperature, we have that $\mu N = G = U - TS + PV$, substituting in the first right insight in Eq. 2.3 we obtain:

$$\Omega(T,V,\mu) = -k_b T ln(\mathcal{Z}_{gc}(T,V,\mu)) = -PV \tag{2.4}$$

The grand potential is therefore an extensive quantity that scales as the volume of the system increases. From the grand potential is possible to access other important thermodynamic variables, such as;

$$S = -\left(\frac{\partial\Omega}{\partial T}\right)_{V,N} \tag{2.5}$$

$$P = -\left(\frac{\partial\Omega}{\partial V}\right)_{T,N} \tag{2.6}$$

$$N = -\left(\frac{\partial\Omega}{\partial \mu}\right)_{T,V}. \tag{2.7}$$

Fluctuations in the grand canonical ensemble

An important condition in the grand canonical ensemble is due to the number of particles that can fluctuate between the system and the reservoir. However, the entity of the fluctuations are someway dependent on the size of the system. The average number of particles in the system is defined as:

$$\langle N\rangle = \frac{\sum_N N e^{\beta(\mu N - E_N)}}{\sum_N e^{\beta(\mu N - E_N)}} = \frac{1}{\beta}\left(\frac{\partial \ln Z_{gc}}{\partial \mu}\right) = \frac{1}{\beta}\left[\frac{1}{Z_{\text{gc}}}\left(\frac{\partial Z_{\text{gc}}}{\partial \mu}\right)\right], \tag{2.8}$$

where $Z_{gc} = Z_{gc}(T, V, \mu)$.
Now deriving the Eq. 2.8 with to respect the chemical potential, we obtain:

$$\left(\frac{\partial\langle N\rangle}{\partial\mu}\right) = -\frac{1}{\beta}\left(\frac{\partial^2 \ln Z_{\text{gc}}}{\partial\mu^2}\right) \tag{2.9}$$

$$= -\frac{1}{\beta}\left[\frac{1}{Z_{\text{gc}}}\left(\frac{\partial^2 Z_{\text{gc}}}{\partial\mu^2}\right) - \frac{1}{Z_{\text{gc}}^2}\left(\frac{\partial Z_{\text{gc}}}{\partial\mu}\right)^2\right] \tag{2.10}$$

$$= \beta\left[\langle N^2\rangle - \langle N\rangle^2\right] \tag{2.11}$$

$$= \beta(\Delta N)^2 \tag{2.12}$$

The relative number of particles that fluctuate between the system and the reservoir can be expressed as:

$$\frac{\Delta N}{\langle N\rangle} = \frac{1}{\langle N\rangle}\sqrt{\frac{1}{\beta}\left(\frac{\partial\langle N\rangle}{\partial\mu}\right)} \propto \frac{1}{\sqrt{\langle N\rangle}} \tag{2.13}$$

hence, in thermodynamic limit when $\langle N\rangle \to \infty$ and V scale with the particles number, the fluctuations between particles in the systems can be neglected [120].

2.1.2 Canonical ensemble

The grand canonical ensemble is a generalization of the canonical ensemble in which the particles are free to fluctuate between the particle reservoir and the system. In the canonical ensemble, the system results to be in thermal equilibrium with the reservoir (heat reservoir) and the particles cannot freely be exchanged outside the system. However, the energy can fluctuate between the particles system and the reservoir. Instead to fix the chemical potential as previously seen in the grand canonical ensemble in the canonical ensemble (N, V, T) are kept constant. The partition function of the canonical ensemble takes a simpler form with to respect the grand canonical partition function, as it does not contain the term corresponding to the exchange of the particles, therefore:

$$\mathcal{Z}_c(T, V, N) = \frac{1}{N!\,(2\pi\hbar)^{3N}}\int e^{-\beta(H(r,q))}\; d^{3N}r\, d^{3N}q \tag{2.14}$$

In the canonical ensemble, the thermodynamic potential is the Helmholtz free energy F and is defined as:

$$\beta F = -ln(\mathcal{Z}_c(T, V, N)) \tag{2.15}$$

from which all the other quantities can be calculated:

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,N} \tag{2.16}$$

$$P = -\left(\frac{\partial F}{\partial V}\right)_{T,N} \tag{2.17}$$

$$\mu = -\left(\frac{\partial F}{\partial N}\right)_{T,V} \tag{2.18}$$

In the canonical ensemble, the relative fluctuation of energy due to the coupling with the heat bath are scaling as the inverse square root of the number of particles in the system in the same way of the relative fluctuation of the particles in the grand canonical ensemble Eq. 2.13 Again for large systems, in the thermodynamic limit, these effects can be neglected.

2.1.3 Microcanonical ensemble

The microcanonical ensemble, effectively describes an isolated system that possesses a fixed energy, number of particles and volume (N, V, E). Since the energy is fixed, only the states between E and $E + \Delta E$ with $E >> \Delta E$ are accessible. However, the central quantity of the microcanonical ensemble is not a proper partition function, but is called: "density of states". That is a direct consequence of the equal a priori probability postulate. Since the energy is fixed each configuration has the same probability. The density function between E and $E + \Delta E$ is defined as;

$$\rho(r, q) = \begin{cases} \dfrac{1}{\Gamma(N, V, E)}, & \text{if } E \leq H(r, q) \leq E + \Delta E \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

where $\Gamma(N, V, E)$ is the volume of the phase space occupied by the microcanonical ensemble and defined as:

$$\Gamma(N, V, E) = \int \int_{E \leq H(r,q) \leq E+\Delta E} d^{3N}r \quad d^{3N}q. \tag{2.20}$$

It is also convenient to define the volume of the phase space under the surface with energy E as $\Sigma(E)$ so that;

$$\Gamma(E) = \Sigma(E + \Delta E) - \Sigma(E) \tag{2.21}$$

from which it is possible to obtain the density of states by simply deriving $\Sigma(E)$ with respect to the energy:

$$g(E) = \left(\frac{\partial \Sigma(E)}{\partial E}\right) = \int d^{3N}q \int \delta(E - H(r, q)) \quad d^{3N}r \tag{2.22}$$

Since $\Delta E << E$, it is possible to expand $\Sigma(E+\Delta E)$ in the vicinity of E, obtaining $\Gamma(E) = g(E)\Delta E$. The key "potential" in the microcanonical ensemble is the entropy, that is directly connected with the density of states $g(E)$ and written as:

$$S(N, V, E) = k_b \ln(\Gamma(E)) \tag{2.23}$$

In this ensemble the equilibrium distribution is the one that maximizes the entropy. This principle provides a foundation for understanding the statistical behavior of isolated systems, where fluctuations in extensive parameters are not allowed. From the entropy it is possible to calculate all the related thermodynamic variables:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V} \tag{2.24}$$

$$\frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{E,N} \tag{2.25}$$

$$-\frac{\mu}{T} = \left(\frac{\partial S}{\partial N}\right)_{E,V} \tag{2.26}$$

2.2 From the ideal gas law to the phase transitions in multicomponent mixtures

2.2.1 The ideal gas law

With the statistical mechanical framework it is possible, for example, to get the ideal gas equation of state, for convenience we recover the result using the canonical ensemble, but it would be possible obtain the same results using other ensembles. As aforementioned, the key quantity in the canonical ensemble is the free energy, without indulging in further fundamental derivations, for an ideal gas it can be expressed as:

$$F = -Nk_bTln(\mathcal{Z}_c(T,V,N)) = -Nk_bT\left(\ln(V) + \frac{3}{2}\ln(T) + c\right) \tag{2.27}$$

by recalling Eq. (2.17) and deriving Eq. (2.27) with respect the volume at fixed (T, N) the final expression of the ideal gas is obtained:

$$-\left(\frac{\partial F}{\partial V}\right)_{T,N} = P = \frac{Nk_bT}{V} \tag{2.28}$$

The ideal gas law, assume that the gas can be represented by non interacting point-like particles. This simplified picture, is valid only when the gas is at high temperature or low pressure, or when the density of particles approach to zero.

In case of real gasses the dependence of pressure and temperature is not well described by the ideal gas law. Real gases are not point-like particles, in fact, they posses an hard core and the molecules can interact between each others. Therefore, they undergo to phase transition (condensation) at relatively low temperature. To better take into account the behaviour of real gases, more sophisticated equations of state are introduced [94, 120].

2.2.2 The Van der Waals equation of state

The Van der Waals equation of state, is a milestone in the understanding of the behaviour of real gases, in this equation b takes into account the volume of the particles and a takes into account the attractive interactions between particles. These terms are added to the volume and the pressure contributions respectively

[11, 125]. In this framework, the new equation of state (EOS) reads as:

$$\left(P + \frac{aN^2}{V^2}\right)(V - Nb) = Nk_bT \tag{2.29}$$

that can be rewritten in term of the the molar volume ν:

$$\left(P + \frac{a}{\nu^2}\right)(\nu - b) = RT. \tag{2.30}$$

When the interactions between particles are strong enough, the behaviour of the fluid strongly deviate from the ideal gas law, and a stationary inflection point, called critical point is reached. This situation is realized when the first and the second derivative of the pressure with respect to the molar volume are both zero.

$$\left(\frac{\partial P}{\partial \nu}\right)_T{}^c = -\frac{RT}{(\nu - b)^2} + \frac{2a}{\nu^3} = 0 \tag{2.31}$$

$$\left(\frac{\partial^2 P}{\partial \nu^2}\right)_T{}^c = \frac{2RT}{(\nu - b)^3} - \frac{6a}{\nu^4} = 0 \tag{2.32}$$

from which, with some algebra it is possible to analytically get all the critical variables. The critical molar volume, temperature and pressure for a gas that follows the Van der Waals EOS are:

$$\nu_c = 3b \tag{2.33}$$

$$T_c = \frac{8a}{27bR} \tag{2.34}$$

$$P_c = \frac{a}{27b^2}. \tag{2.35}$$

By further lowering the temperature below the critical point the system undergoes to phase separation, in which a high density phase (liquid) coexists with a low density phase (vapour). This points are represented on Fig. 2.2 where the Van der Waals isotherms are shown. The metastable regions, in which the two phases are coexisting, are connected by a tie line in which the area above a coexistence region equalize the area under the isotherm to the other side. This schematic representation is called Maxwell construction and ensures the mechanical and chemical equilibrium between the two phases. Other two important points are the one enclosed between the minimum and maximum of

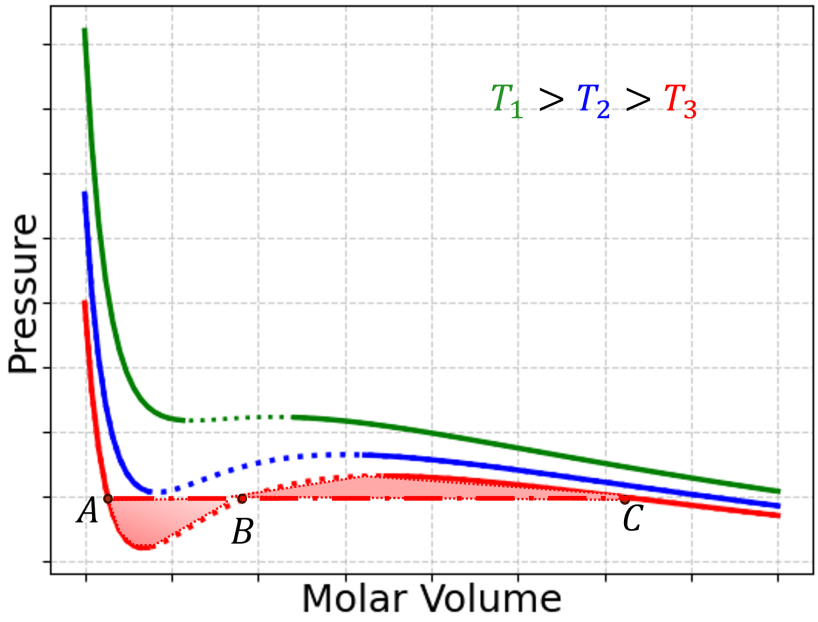

Figure 2.2: Van der Waals isotherms: the dotted regions are the regions of instability in which the derivative of the pressure with to respect the volume is greater than zero. Looking at the plot below the critical temperature (green line), the region of instability become larger (blue line), an high density phase (liquid) coexist with a low density phase (vapour). This coexistence regions are better represented in the last isotherm curve (red line), they correspond to the points along the tie line that follow the Maxwell construction. Between this points, the shaded area enclosed in segment AB equalize the area enclosed in segment BC.

the curve (dotted part of the line) in Fig. 2.2 in which the first derivative of the pressure with respect to the molar volume is bigger than zero. This region is a zone of instability in which phase separation without nucleation of a new phase occurs. The unstable zone of the curve is called spinodal region.

In this regards the Van der Waals equation is in qualitative agreement with the experimental behaviour of various real gasses. However, in several cases the quantitative predictive power of the Van der Waals EOS is not good enough, especially with strongly interacting gases, and more advanced equations of state are needed in order to better describe these fluids. Despite that, the Van der Waals EOS, established remarkably new insights on the behaviour of real gases. Another important observation coming from this EOS, is the so called "law of corresponding states". For example, the compressibility factor is defined as $Z = \frac{PV}{RT}$ it is unitary for an ideal gas, but for a Van der Waals gas at the critical

point a strong deviation from unity is observed:

$$Z_c = \frac{P_c \nu_c}{RT_c} = \frac{\nu_c}{\nu_c - b} - \frac{a}{\nu_c RT_c} = \frac{3}{8} \tag{2.36}$$

This value is independent of the type of gas involved. Despite that, many gasses further deviate from the Van der Waals predictions. It was also observed that different gasses deviate of the same degree from the ideal gas behaviour if compared using the reduced variables. The reduced variables are easily defined as:

$$P_r = \frac{P}{P_c} \tag{2.37}$$

$$V_r = \frac{V}{V_c} \tag{2.38}$$

$$T_r = \frac{T}{T_c} \tag{2.39}$$

from which a reduced form of the Eq. 2.30 equation can be obtained. Another important consequence of the Van der Waals equation is the possibility to rewrite the compressibility factor in a virial series of the molar volume. For larger molar volume or low density ($\nu >> b$) we can expand Z in a Taylor series:

$$Z = 1 + \left(b - \frac{a}{RT}\right)\frac{1}{\nu} + \frac{b^2}{\nu^2} + \frac{b^3}{\nu^3} + \dots \tag{2.40}$$

or as a function of density ρ:

$$Z = 1 + \left(b - \frac{a}{RT}\right)\rho + b^2\rho^2 + b^3\rho^3 + \dots \tag{2.41}$$

From which it is possible to define the second virial coefficent as:

$$B = \left(b - \frac{a}{RT}\right) \tag{2.42}$$

by replacing b^2, b^3 $\dots$, with the terms C, D $\dots$ we finally arrive to a more compact expression as a function of the density:

$$Z = 1 + B\rho + C\rho^2 + D\rho^3 + \dots \tag{2.43}$$

The second virial coefficent B is temperature dependent and its behaviour is well

represented by the Van der Waals EOS. The higher terms do not reproduce well the experimental observed trend. To summarize, the EOS established by Van der Waals introduces several concepts that are used in more advanced theory to describe, i.e. the phase behaviour of protein-salt solutions, the most prominent results obtained by Van der Waals can be summarized as follow:

- Explain the deviation from the ideal behaviour including the size of the particles as well as taking into account the interactions between them [11, 125].
- Establish the basic condition to explain the phenomenology of metastable, unstable first order phase transitions and coexistence between phases [125].
- Establish the law of corresponding states [97, 126].
- Establish an expression for the second (and high order) virial coefficient by expanding the compressibility factor in virial series [18].

In recent years some of the concepts introduced by Van der Waals and others eminent scientists in the past century, are successfully implemented not only in the case of pure substances as aforementioned, but are extended to describe complicated anisotropic systems, multi-component mixtures and even protein salt solutions [47, 124].

2.2.3 Phase separation in binary mixture: the regular solution model

Gasses are a perfect model to understand the basic physic behind phase separation and phase coexistence. However, if we work with mixtures we have to take into account not only the interactions between atoms or molecules of the same type, but also the interactions between different molecules in the mixture. Especially, most of the systems of interest, such as polymer-water systems, can have a limited range of composition, in which the components mix together. An early and still simple model to describe multicomponent systems, is the regular solution model. A regular solution, is a solution that possesses a behaviour that deviates only in small degree from an ideal solution. An ideal solution, does not possesses enthalpy of mixing ($\Delta H_{mix} = 0$), no change in volume ($\Delta V_{mix} = 0$) and

no change of internal energy ($\Delta U_{mix} = 0$) upon mixing. For an ideal solution, the only properties that change during a mixing process are:

$$\Delta G^{id}_{mix} = RT \sum_{k=1}^{c} x_k \ln(x_k) \tag{2.44}$$

$$\Delta S^{id}_{mix} = -R \sum_{k=1}^{c} x_k \ln(x_k) \tag{2.45}$$

where x_k is the molar fraction of the k component in the mixture. A regular solution, has the same entropy of mixing of an ideal solution, but the enthalpy of mixing is not zero and it is a function of the composition of the mixture [46, 47]. Since the excess entropy of mixing of a regular solution is zero, the mixing Gibbs free energy contains an excess term due to excess enthalpy contribution that takes into account the interactions between particles in the mixture. Therefore, the excess term reads as:

$$\Delta G_{mix} = \Delta G^{id}_{mix} + \Delta G^{ex}_{mix} \tag{2.46}$$

$$\Delta G^{ex}_{mix} = \Delta H^{ex}_{mix} + \cancel{T \Delta S^{ex}_{mix}} \tag{2.47}$$

A simple way to describe the heat of mixing as a function of the composition and an adjustable parameters (α_0) is to consider $\Delta H^{ex}_{mix} = \alpha_0 x_1 x_2$. Therefore the total Gibbs free energy of mixing of a binary mixture reads as:

$$\Delta G_{mix} = \alpha_0 x_1 x_2 + RT \left(x_1 \ln(x_1) + x_2 \ln(x_2) \right) \tag{2.48}$$

where α has the unit of energy per mole. The sign of the parameter α, that takes into account the interactions between particles, determines the type of departures from an ideal solution. For a positive deviation from the ideal solution, at low temperature, the free energy curves as a functions of composition develop additional maxima and minimum, resulting in a miscibility gap where two phases with different composition coexist. A schematic representation at fixed value of α is given in Fig. 2.3. In this simple framework α is independent of the temperature, however in more realistic scenarios the interaction parameter has a temperature dependence. For common tangent construction, the points that share the same minimum of free energy are the points where the mixture

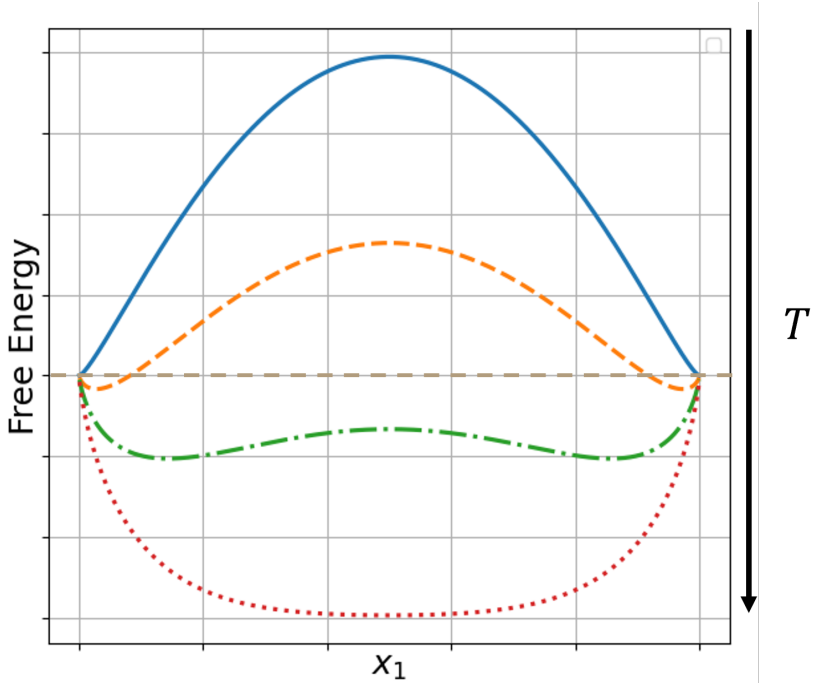

Figure 2.3: The graph depicts the free energy as a function of the composition of component one. The dotted horizontal line signifies the zero of the free energy of mixing. When the minimum in the free energy of mixing splits into two distinct minima, the unmixing process becomes energetically favorable. Consequently, the mixture phase separates into two coexisting phases with different compositions. Each curve on the graph represents a distinct isotherm. At high temperatures, the mixture is miscible across the entire composition interval.

phase separate in two coexistent phases with different composition. The phases are in thermodynamic equilibrium between them. From this simple picture, it is possible to extract the liquid-liquid phase separation coexistence boundaries as a function of the temperature as represented in Fig. 2.4, that in this simple framework has the critical point located at $x_1 = x_2 = 0.5$. The symmetry of the critical point with respect to the composition at $x_1 = x_2 = 0.5$ implies that the properties of the species 1 surrounded by a species 2 in a diluted solution are the same that the properties of a solution of species 1 that surround a species 2. This scenario is often not accurate and holds, as an useful approximation, only when the components are chemically similar. A consequence of this observation is that the excess term of one component can be rewritten as:

$$\Delta G_1^{ex} = \Delta H_1^{ex} = RT \ln(\gamma_1) \tag{2.49}$$

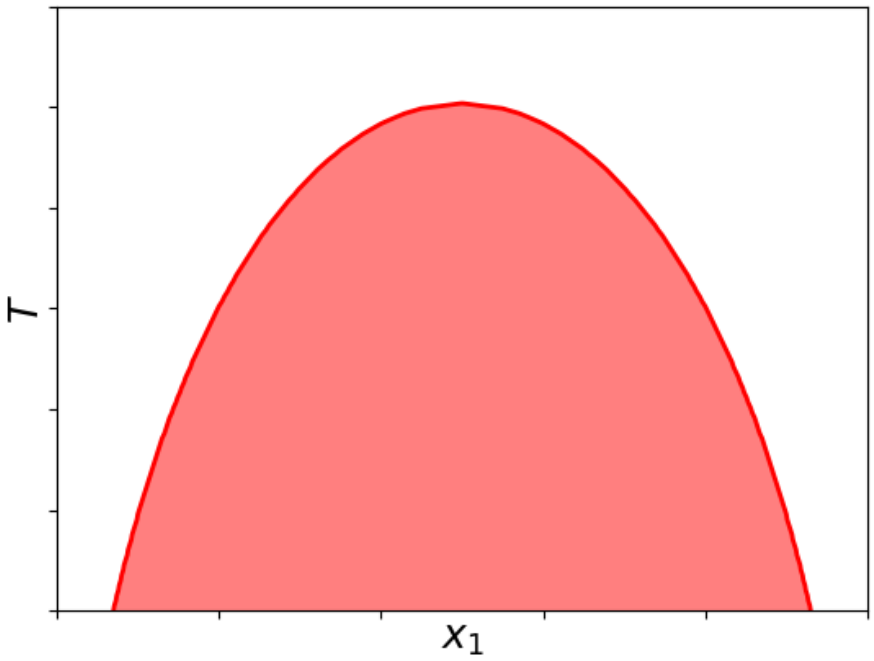

Figure 2.4: Temperature dependence of the miscibility gap for a regular solution with $\alpha = 5000 \quad Joule/mol$. The image is obtained by common tangent from a set of free energy curves at different temperatures, similar to the curves in Fig. 2.3. The image refers to the binodal coexistence points.

Since $\Delta H_1^{ex} = \alpha_0 x_2^2$ we obtain a computational accessible expression for the calculation of the activity coefficients.

$$\gamma_1 = e^{\alpha_0 x_2^2 / RT} \tag{2.50}$$

More realistic scenarios assume a two parameters model in which it is assumed, that the bond energies are functions of the composition. This model of solution is called suberegular and the enthalpy of mixing assumes a more sophisticated form described as follow:

$$\Delta H_{mix}^{ex} = \Delta G_{mix}^{ex} = x_1 x_2 (\alpha_0 x_1 + \alpha_1 x_2) \tag{2.51}$$

it can be further extended in the so called non regular solution model assuming that the parameters α_0 and α_1 can be temperature dependent and written as $\alpha_0(T)$, $\alpha_1(T)$. For components that have a great difference in volume and polarity

like polymer-water mixtures, a similar model, but specifically elaborated to describe water-polymer systems are available. The most popular and used is the Flory-Huggins model. However, to visualize liquid-liquid phase separation in mixtures, the regular solution model can be used as useful reference.

2.3 Theory of solutions used in the description of protein interactions and phase transitions

2.3.1 Crystallization: Classical Nucleation Theory

Proteins in solution display a wide range of phase transitions, including LLPS, fibrillation and aggregation. One relevant biological phase transition not completely understood is protein crystallization. This phase transition is particularly interesting, because from a crystallized protein, it is possible to determine the tridimensional structure using X-ray diffraction. One of the earliest theories, originated to describe the nucleation of a new phase from another and originally applied to describe condensation of a vapor into a liquid, is the Classical Nucleation Theory (CNT). The theory was later used to describe protein crystallization and liquid-solid phase transition [54, 56]. Despite the relative simplicity, the CNT is a useful tool to understand protein crystallization. In the CNT framework, the nucleation of a new phase is a thermodynamic driven process in which the system, has to overcome an energy barrier that correspond to a maximum in Gibbs free energy in order to generate stable nuclei. In the simplest scenario, the variation of the Gibbs free energy is described by two contributions:

$$\Delta G_{nucl} = \Delta G_{bulk} + \Delta G_{surf}. \tag{2.52}$$

Under the assumptions that particles are spherical and they can interact forming cluster that can become larger by the addition of single units, the bulk term can be expressed as:

$$\Delta G_{bulk} = -\frac{4\pi r^3}{3\nu} k_b T \ln\left(\frac{c}{c_{eq}}\right), \tag{2.53}$$

where r is the radius of the cluster, ν the molar volume, c the concentration at saturation and c_{eq} is the concentration at equilibrium when the crystal is in equilibrium with the supernate. The ratio between this two concentrations is

called supersaturation ratio and is a central quantity in the framework of the CNT. The bulk contribution on the free energy is negative and is a favorable term to the nucleation of a new phase. However, in order to form a new phase a certain amount of energy should be used to create e new interface between the crystal and the bulk solution. The energy cost to the formation of a new phase is given by the ΔG_{surf} term, a positive and unfavorable contribution to the nucleation. In the case of spherical particles the term is written as:

$$\Delta G_{surf} = 4\pi r^2 \omega, \tag{2.54}$$

where ω is the interfacial free energy. The overall contribution is therefore:

$$\Delta G_{nucl} = -\frac{4\pi r^3}{3\nu} k_b T \ln\left(\frac{c}{c_{eq}}\right) + 4\pi r^2 \omega, \tag{2.55}$$

from which, by differentiation of the free energy with to respect to the radius of the cluster it is possible to obtain the critical radius r_c^*:

$$\frac{\partial \Delta G_{nucl}}{\partial r} = 0 \quad \rightarrow \quad r_c^* = \frac{2\nu\omega}{k_b T \ln(\frac{c}{c_{eq}})}. \tag{2.56}$$

The value of the critical radius is therefore temperature and supersaturation dependent, it is reduced by increasing the temperature or the supersaturation as shown in Fig. 2.5. In the CNT the cluster is formed by addition of single units and the effects of the interfaces are neglected. It is possible to incorporate the effects of an interface multiplying the bulk nucleation free energy with a function of the contact angle between the surface and the droplet:

$$\Delta G_{nucl}^{het} = \Delta G_{nucl}\, f(\Phi), \tag{2.57}$$

where Φ is the contact angle having the form of:

$$f(\Phi) = \frac{(2 + cos\,\Phi)\,(1 - cos\,\Phi)^2}{2}. \tag{2.58}$$

The resulting values of the function Φ are: $0 \leq f(\Phi) \leq 1$. For high contact angle, i.e. 180°, $f(\Phi) = 1$ and the bulk condition is recovered, while for smaller contact angle a decreased free energy barrier is achieved [112, 127]. Therefore in CNT, an interaction between the molecule and the surface results in an effect similar to

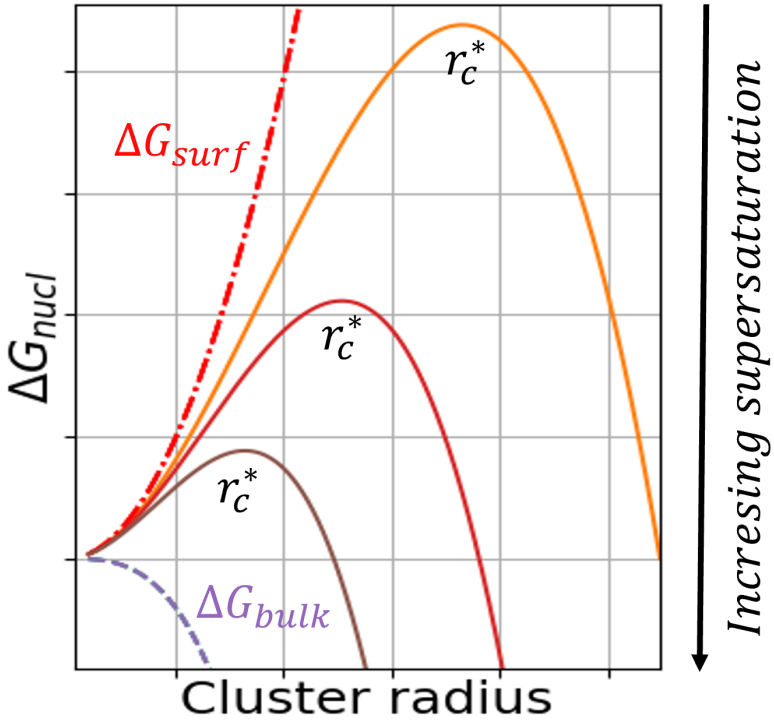

Figure 2.5: Gibbs free energy of nucleation: the red and purple lines correspond to the single contribution in free energy. The curves with the energy barriers correspond to the total contribution of free energy of nucleation. Increasing the supersaturation decrease the height of the energy barrier and reduce the critical cluster radius for the formation of a new phase.

the increased supersaturation in the bulk, with a faster kinetic of nucleation. In the CNT framework the rate of nucleation depends on the height of the energy barrier and is given by an Arrhenius like formula:

$$J = A \exp\left(\frac{-\Delta G^*_{nucl}(r = r_c^*)}{k_b T}\right). \tag{2.59}$$

The expression of the rate, takes into account the prefactor A, that contains different contributions, such as the number of collisions and residence time of the cluster at the top of the energy barrier. However, this simple model gives not reliable values for the nucleation rate as results.[112] More specific models result in a better estimation of the nucleation rate and other time dependencies of the nucleation process [7, 84]. A successful application of a more accurate

model, in case of protein nucleation in presence of trivalent salts is discussed in the experimental methods and in the results section where we used a sigmoidal model to predict the nucleation time and rate.

2.3.2 McMillan-Mayer theory of solutions

Writing about multicomponent systems, an important improvement in the description of mixtures was made by McMillan-Mayer [75, 76]. The theory provide useful analogies between the behaviour of the solutes in a solvent with the behaviour of real gasses. Within the framework of this theory, it is possible to isolate the contribution of the solutes from the solvent and obtain expression for the pair interaction between solute molecules. In addition, it expands the chemical potential at low density of solute in virial series of the osmotic pressure as a function of the concentration, in analogy with the virial series obtained previously in Eq. 2.43 for a Van der Waals gas. In the context of proteins-salt solution in water, this approach allows to treat the solvent-salt matrix as a mixed solvent in which the protein, being considered to be the solute is immersed. This approximation is valid, because usually the salt and water in solution are present in higher concentration with respect to the protein. The theory assumes that two solutions are in contact by a semipermeable membrane, have the same temperature and chemical potential of the solvent (water and salt), but different chemical potential and pressure. However, at equilibrium the pressure of diffusive species should be the same in both sides. The additional pressure necessary to stop the diffusion between the two sides, is due to the presence of the solute (protein) and is called osmotic pressure [48, 122]. As aforementioned, the osmotic pressure can be expanded in a virial series at low solute concentrations:

$$\Pi = c_p k_b T(1 + B_2(\mu_a, T)c_p + \dots, \tag{2.60}$$

where c_p is the protein concentration in the solvent. The second virial coefficient B_2, depends on the chemical potential of the solvent μ_a as well as on the temperature. In the Mcmillan-Mayer framework, the second virial coefficient is written as:

$$B_2(\mu_a, T) = -\frac{1}{2}\int [e^{-V(r)/kT} - 1]d\mathbf{r}, \tag{2.61}$$

where $V(r)$ is the potential of mean force in solution that depends on the solute-solute separation distance being a function of the solvent chemical potential. These results allow to compare the statistical mechanics of proteins in solutions with better known model applied on the description of gaseous systems, for example, in the calculation of the second virial coefficient, different pair potentials are applied, such as the sticky hard sphere potential. A brief introduction on the calculation of the second virial coefficents from $V(r)$ for the sticky hard sphere model is given in the next section. In addition, the Mcmillan-Mayer theory was successfully adopted in the result section to estimate the second virial coefficient from experimental data, the theoretical derivation of the result is done in Sec. 4.2.

2.3.3 The sticky hard sphere model

The sticky hard spheres model combines the concepts of hard sphere exclusion and attractive interactions. The model assumes that the spheres interact via a short-range attractive potential, often represented by a square-well potential [8]. The potential is zero for distances larger than the diameter of the spheres and becomes attractive within a certain range.

The key parameters of the model include the sphere diameter, the range and strength of the attractive potential. These parameters influence the phase behavior and thermodynamic properties of the system. The potential energy of interaction $V(r)$ between two sticky hard spheres at a distance r is often represented by a potential composed by an hard sphere term and an attractive term such as $V_{SHS}(r) = V_{HS}(r) + V_a(r)$ where $V_a(r)$ is the attractive contribution.:

$$\beta V(r) = \begin{cases} \infty & r \leq \sigma \\ -\beta V_o = \ln(\frac{12\tau\Delta}{\sigma+\Delta}) & \sigma < r < \sigma + \Delta \\ 0 & r \geq \sigma + \Delta \end{cases} \tag{2.62}$$

where τ is the stickiness parameter, β the inverse temperature and Δ the width of the square well.

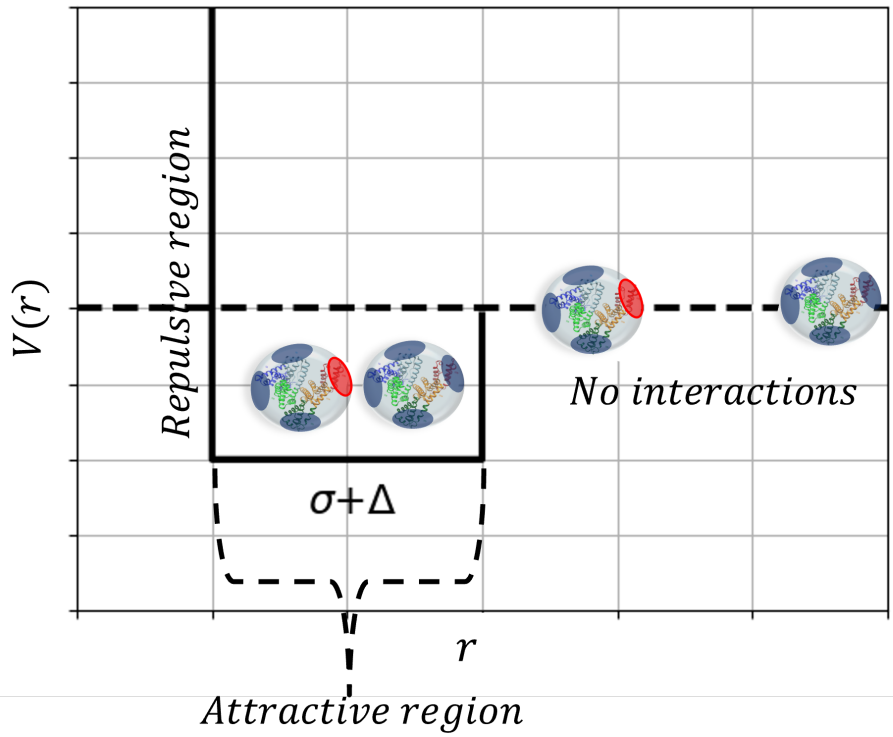

Figure 2.6: An example of the sticky hard sphere potential: At short distances the particles experience an hard core repulsion, the attractive range is inside the square well part at distances between r and $r + \Delta$, when the distance between particles is outside the interaction range the particles do not feel any potential.

The stickiness parameter can be written as:

$$\tau = \frac{1}{12\epsilon} \exp(V_0/kT) \tag{2.63}$$

where ϵ is the ratio between the width of the square well Δ and $\Delta + \sigma$, and therefore depends on the range of the interactions. The normalized second virial coefficient can be obtained in the limit of $\Delta \to 0$:

$$\lim_{\Delta \to 0} \frac{B_2}{B_{HS}} = 1 - \frac{1}{4\tau} \, . \tag{2.64}$$

The sticky hard spheres model provides a useful framework for studying the thermodynamic behavior of systems with short-range attractive interactions [8, 114]. Its simplicity allows for analytical and numerical investigations using, for example, Small-Angle-X-rays-Scattering (SAXS), making it a valuable tool

in modelling protein interactions under the assumption that the shape of the proteins is maintained spherical under the experimental conditions [135, 141]. Further research may explore extensions of the model, considering variations in the attractive potential or incorporating additional complexities to better capture the behavior of specific colloidal systems.

2.3.4 A brief introduction on associating fluids: The Wertheim theory

The Wertheim theory is a statistical mechanical theory used to describe the behaviour of associating fluids. An associating fluid is a species with strong directional bondings, such as hydrogen bonds in water or any kind of electrostatic directional interactions. A rigorous derivation based on the graph theory can be found in the following references [131–133]. The aim of this subsection is to briefly introduce the basis of the theory, that was successfully applied in the result section. In the Wertheim theory the particles posses an hard core (hard spheres) decorated with m binding sites or patches. In this framework the Helmholtz free energy is given by two contributions:

$$f = f_{HS} + f_{bond} \tag{2.65}$$

In order to make use of the theory different conditions shall be fulfilled:

- Each binding sites can form only one bond.
- Particles cannot associate in closed loops.
- For sterical reasons a particle can bind another particle only once.

With this conditions it is possible to write the binding free energy per particle as:

$$\beta f_{bond} = m\rho \left(\ln(1 - p_b) + \frac{1}{2} p_b \right), \tag{2.66}$$

where p_b is the probability of a patch having formed a bond. Note that p_b depends on the number density ρ and follows from the mass-action equation;

$$\frac{p_b}{(1 - p_b)^2} = m\rho\Delta, \tag{2.67}$$

Where the binding probability ($\rho\Delta_{ab} = \rho\Delta_{ba}$) between two independent sites on particle 1 and 2 is given by:

$$\Delta = 4\pi g_{HS}(\sigma) K F, \tag{2.68}$$

where $g_{HS}(\sigma)$ is the radial distribution function of the reference state, F is the Mayer function, and include the information about the potential between the patchy site a on particle 1 and the patchy site b on particle 2 and K is a contribution due to the geometry of the patchy interaction. The Mayer function is normally expressed as:

$$F_{ab}(12) = \exp\left[\beta V(r_{12}, \Omega_1, \Omega_2)\right] - 1. \tag{2.69}$$

The flexibility of the model allows to study different kind of patchy-patchy attractions by modify, for example the pair potential between sites. One simple case is to assume that particles interact by patches of the same kind. In this case the law of mass action is solved to obtained the expression in Eq. 2.66. The Wertheim theory was successfully extended to systems with different kind of patches by changing the corresponding law of mass action. A more detailed approach is given in the following references [52, 103]. The Wertheim theory allows to study a wide range of fluids and to simulate a huge class of systems, predicting experimental phase diagrams, including phase diagrams obtained from protein-salt solutions.

Part III

Experimental methods

Chapter 3

Experimental methods

3.1 UV-visible spectroscopy

In the UV-visible spectroscopy a light beam in the UV-visible region is used to promote electrons in a sample from an energy level to another. The entity of this transition can be measured as absorbed or transmitted light. It was shown, that an incident light wave that pass through a sample follow a decay on the transmitted light that is due to the absorbance of the sample [1, 40, 87]. The absorbance of the sample can be described by the Lambert-Beer law:

$$A = \epsilon \, l \, c \tag{3.1}$$

Where A is the absorbance of the sample, ϵ the characteristic extinction coefficients of the species, c the concentration and l the path length. For its easy applicability is now a standard tool to detect the concentration of molecules in solution. In particular, in the case of protein-salt solutions it can be used, for example, to calculate the supersaturation ratio by measuring the absorbance at the beginning and at the end of the phase separation or crystallization process, or can be used in real time during phase transitions obtaining information about the crystallization kinetics as shown in Fig 3.1.

The absorbance peak in case of proteins is usually present at 280 nm and is mostly due to the presence of aminoacids such as tryptophan and tyrosine with some contribution coming from phenylalanine [1, 35, 40]. During the nucleation of a crystal phase from a protein solution, a depletion of the protein bulk concentration and consequently a decrease in absorbance is recorded. The resulting curve, follow a logistic decay with an initial induction time, that follow

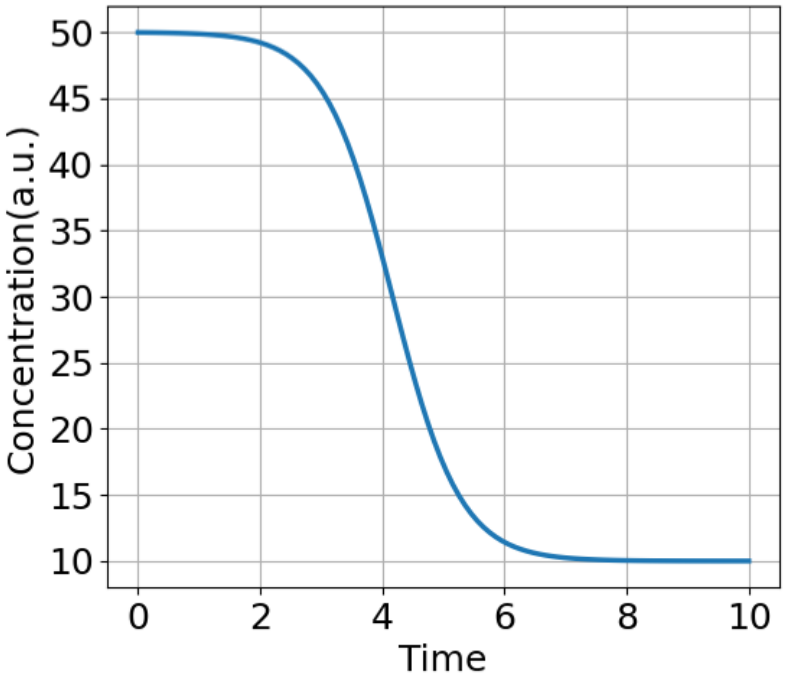

Figure 3.1: Example of application of UV-vis spectroscopy to reveal the change of concentration of a specie, for example the depletion of protein from the bulk solution, due to the crystallization or liquid-lquid phase separation.

an empirical law of the type:

$$C(t) = \frac{c_i}{1 + exp\left(k\left(t - t_c\right)\right)} + \frac{c_{eq}}{1 + exp\left(-k\left(t - t_c\right)\right)} \tag{3.2}$$

Where c_i is the initial concentration and c_{eq} is the equilibrium concentration of the crystals, k is the characteristic rate constant and t_c the induction time [71]. Another approach to quantify the extension and the rate of nucleation is by counting the number of crystals in a glass slide and fitting the resulting nucleation density plot with a similar class of logistic functions. An application of such method is discussed in Chapter 6.

3.2 Optical microscopy

An important tool to study phase transitions in protein solutions is the optical microscopy [12, 71, 109]. In the past years the application of optical microscopy for micro and even sub micrometers resolution allowed the detection and characterization of protein crystal growth, phase separation and fibrillation. In

transmission mode, a sample on a glass slide is positioned between an incident light source emitted from a light bulb, which passes through a series of lenses including the objective lens. The light transmitted through the sample is then focused by the objective lens and displayed as a magnified image to the operator through the eyepiece lens [20, 78]. The optical microscopy resolution is limited by diffraction and the theoretical limit or Abbe's limit is approximately given by:

$$R_{max} = \frac{\lambda_{min}}{2\,NA} \approx 200\,nm \tag{3.3}$$

where $\lambda_{min} \approx 400\,nm$ is the shortest UV-vis wavelength and NA is the numerical aperture and depends on the reflective index of the medium and on the cone of light entering into the objective lens, in air NA is less or equal than 1. An early application of the optical microscope to characterize the protein crystallization behaviour, was performed on lysozyme solutions with the aim of detect the nucleation time, as well as the growth rate. In this experiment a sample containing lysozyme and a crystallizing agent is placed in a spacer. The crystals are detected in an early stage and the growth is followed in real time, than a time series of pictures with a predefined time interval is taken. The size of the crystals is measured for each picture in time. The first time point, corresponds to the length of the crystals equal to zero and time equal to the preparation time of the sample. If the crystals are detected in the early stage of the growth, then the experimental data are fitted with a linear regression. The difference between the preparation time and the intercept with the x axis from the fitting is the induction time of nucleation, while the angular coefficient of the regression is the growth rate. This method provide a reliable fit only at the beginning of the crystal growth [27, 66]. In fact, the real curve of this process is better represented by a logistic function. As t increase the growth of the crystals slows down and a plateau is reached. At this point a linear regression will results in a nonphysical fitting, giving as an output a negative induction time for nucleation. Since the limitation of the linear regression analysis cannot be easily overcome, a different method to obtain the induction time of nucleation using an optical microscope can be employed. A recent method used to this purpose in this work, is to extract the induction time using a nucleation density plot. In this framework, a supersaturated solution is brought from the nucleation zone, where the nucleation and growth are simultaneously occurring, to a supersaturated metastable solution in which only the crystals growth occurs.

Typically, this is accomplished by raising the temperature from lower levels to higher ones. The resulting number of crystals is counted at different point in time, until no more crystals are detected. The resulting points are fitted using a sigmoidal model [84]. The details of the kinetic model are elaborated upon in the results section, where a modification of the Nenev model was employed on a solution containing HSA in the presence of $CeCl_3$.

3.3 Small Angle X-ray Scattering

Small Angle X-ray Scattering (SAXS) is a standard tool in the study of colloidal systems, including protein-salt mixtures. The advantage of using an X-ray source to investigate the sample lies in the small wavelength of the radiation, which is comparable in size to the atoms ($10^{-12} - 10^{-10}m$). Due to limitations inherent in optical instruments, scattering techniques are widely employed in the investigation of protein-salt solutions, both for characterizing the dynamic structure in solution and for studying the dynamics and kinetics of phase transitions [39, 69, 71, 114]. In the scattering process, a light source interacts with an object and is deflected from its original trajectory by a certain scattering angle ϕ. The scattered beam is then collected at small scattering angles. The information obtained is transformed from real space to reciprocal space. In the SAXS approach, the scattering intensity is plotted against the scattering vector q, defined as $Q = 4\pi \sin\theta/\lambda$. From the features of the scattering profile, it is possible to obtain information about the shape, interactions, and aggregation of the particles. An interesting approach in modeling SAXS curves to describe protein interactions involves using a specific potential, such as the sticky hard sphere potential described in Section 2.3.3. Essentially, the SAXS intensity can be described by the following relation:

$$I(Q) = \rho\,(\Delta\epsilon)^2\,V_p^2\,P(Q)\,S(Q), \tag{3.4}$$

where ρ is the number density of particles, $\Delta\epsilon$ is the scattering contrast, V_p is the particle volume, $P(Q)$ the form factor and $S(Q)$ is the structure factor. In modelling the SAXS curves it is important to use the right form factor $P(Q)$ that

for many proteins it is approximated as an ellipsoid of revolution:

$$P(Q) = \int_0^1 \left| \frac{3(\sin u - u\cos u)}{u^2} \right|^2 \tag{3.5}$$

$$\text{with} \quad u = Q\,b\left[\left(\frac{a}{b}\right)^2 x^2 + (1-x)^2\right]^{1/2} \tag{3.6}$$

Where a and b are the radii of the ellipse. The information related to the interactions are included in the structure factor $S(Q)$ being expressed as:

$$S(Q) = 1 + 4\pi\rho \int_0^\infty (g(\mathbf{r}) - 1)\frac{sin\,Q\,r}{Q\,r} r^2\,dr \tag{3.7}$$

In the pair correlation function $g(\mathbf{r})$ is encoded the thermodynamics and the information about the particles interactions [39, 119]. $S(Q)$, can be derived from the equilibrium distribution of particles as described by $g(\mathbf{r})$, which in turn is influenced by the interactions governed by the interparticle potential. In the case of a uniform isotropic fluid composed of spheres, the Ornstein-Zernicke (OZ) equation serves as a valuable tool for computing $S(Q)$. In the most general case the total correlation function is defined as:

$$h(\mathbf{r}) = g(\mathbf{r}) - 1 = c(\mathbf{r}) + \rho \cdot \int c(\mathbf{r} - \mathbf{x})h(\mathbf{x})d\mathbf{x} \tag{3.8}$$

The total correlation function $h(\mathbf{r})$ between two particles is the sum of the direct correlation $c(\mathbf{r})$ and a term that takes into account all the indirect correlations between particles, which indirectly affect the direct interaction. To solve the integral, several approximations can be used, with the most commonly used being the Percus-Yevick closure relation. Essentially, this relation provides a way to connect the pair correlation function $g(\mathbf{r})$ with the direct correlation $c(\mathbf{r})$. In this framework, the direct correlation is given by:

$$c(\mathbf{r}) = g(\mathbf{r})[1 - \exp\{\beta V(\mathbf{r})\}] \tag{3.9}$$

This relation can be adopted in the case of the aforementioned sticky hard sphere potential (SHS) to obtain a good approximation for $S(Q)$ and, therefore, gain information about the interaction between particles in the system. In Part V, we used a SAXS fitting procedure with the sticky hard sphere model for the pair interaction to estimate the second virial coefficient, comparing the results

obtained with the values derived from supersaturation measurements [118]. In a SHS system the reduced second virial coefficient, B_2/B_2^{HS} can be expressed as:

$$\lim_{\Delta \to 0} \frac{B_2}{B_2^{HS}} = 1 - \frac{1}{4\tau} \tag{3.10}$$

The critical value of B_2/B_2^{HS} at value of the stickiness parameter $\tau = \tau_c$ is approximately -1.56:

$$\frac{B_2(\tau = \tau_c)}{B_2^{HS}} = 1 - \frac{1}{4\tau_c} \approx -1.56 \tag{3.11}$$

Note that this value is not an hard limit for any kind of phase separation, but a threshold for phase separation in SHS fluids. In the context of the Percus-Yevick closure relation, the critical parameters can be obtained as follows:

$$\eta_c = \frac{\sqrt{2} - 4}{\sqrt{2}} \approx 0.1213, \quad \text{and} \quad \tau_c = \frac{2 - \sqrt{\frac{2}{6}}}{\sqrt{2 - \frac{4}{2}}} \approx 0.0976 \tag{3.12}$$

Where η_c is the critical packing fraction. In the result section, for the ion activated patchy particles model, the critical packing fraction is different and it varies as the number of binding sites on the surface of the particles (patches) change.

Part IV

Results

An Alternative Approach to the Osmotic Second Virial Coefficient of Protein Solutions and its Application to Liquid-Liquid Phase Separation: this part of the results section is mostly based on the homonyms paper published in the Journal of Chemical Physics, here are listed the contributions of the different authors to the realization of this work.

Furio Surfaro: wrote the original manuscript, developed the thermodynamics method, performed part of the experiments.

Dr. Ralph Maier: performed the Small-Angle-X-rays-Scattering experiments and contributed with some ideas to the manuscript.

Kai-Florian Pastryk: assisted the main author in performing the experiments and performed UV-vis experiments to determine the protein concentration before and after the phase separation process.

Prof. Dr. Roland Roth: contributed to the development of the manuscript, in the understanding of the thermodynamics model and supervised the theoretical part of the work providing supervision and funds.

PD. Dr. Fajun Zhang: contributed with the development of the manuscript and provided supervision and funds.

Prof. Dr. Dr. hc Frank Schreiber: contributed with the development of the manuscript and provided supervision, funds and facilities.

Chapter 4

An Alternative Approach to the Osmotic Second Virial Coefficient of Protein Solutions and its Application to Liquid-Liquid Phase Separation

4.1 Introduction

The osmotic second virial coefficient B_2 is an important parameter to describe the interactions and phase behaviour of protein solutions, including colloidal systems and macromolecular solutions. Another key parameter to describe the driving force of the nucleation of a new phase is the supersaturation, which is used in the Classical Nucleation Theory (CNT) framework and is connected with the favorable contribution in the Gibbs free energy in the bulk solution. Since the effective interaction potential between proteins is typically not known, a direct calculation of the second virial coefficient with Eq. (2.61) is impossible. Still, B_2 is accessible either by thermodynamic considerations, or experimentally. If B_2 is positive, the effective interaction is overall repulsive, while it is attractive for negative values of B_2. For LLPS a sufficiently strong attraction is required and it has been observed for several colloidal systems $B_2/B_{HS} \approx$ -1.5 [130] close to a critical point and more negative in the LLPS region. For protein solutions that display LLPS, even though the effective interaction is much more complicated than in the colloidal system, similar observations have been made close to the LLPS region [98, 135]. Furthermore, it seems that frequently B_2 falls into a narrow range when crystallization occurs, called crystallization slot [37]. In this work, we establish a thermodynamic relation in order to connect the

osmotic second virial coefficient B_2 with the supersaturation. The method here employed, is explained in the following subsections and can be applied if the system, such as a protein solution, shows a meta-stable LLPS, and relies solely on easily accessible macroscopic observables. We explain different experimental procedures that have successfully been used for the determination of B_2.

4.2 Theory

We start by considering a protein solution with concentration c. Its chemical potential μ, relative to a dilute reference system with concentration c_0 and chemical potential μ_0, is given by

$$\mu = \mu_0 + RT \ln \left(\frac{\gamma c}{c_0} \right), \tag{4.1}$$

where γ is the activity coefficient that takes the deviations of the solution from ideal behaviour into account. Here, we assume that all effects of the solvent, ions and crowding are included by γ. By definition, γ is equal to 1 for ideal solutions, i.e., at low protein concentrations. If γ deviates from unity, it indicates a concentration that is sufficiently high so that particles interact. A value of γ smaller than 1 corresponds to overall interactions between particles that are attractive, while a value larger than 1 hints towards overall interactions that are repulsive. We wish to describe an experimental situation, where a protein solution is prepared at an initial concentration c_i, that lies within the LLPS region, so that the solution phase separates into a low density phase with concentration c_l and a high density phase with concentration c_h. For protein solutions the LLPS is usually meta-stable, which makes the determination of c_l and c_h far from trivial. However, we know that along the tie line of the phase separation one can observe chemical and mechanical equilibrium, i.e., constant chemical potential and constant osmotic pressure. In the following, we will make use of the chemical equilibrium: $\mu(c_i) = \mu(c_l) = \mu(c_h)$. A schematic phase diagram describing this process can be found in reference [137]. After the LLPS occurs one can observe that the concentration of the low density phase reduces to its equilibrium value c_{eq}, as the solubility line is reached. During this step the high density phase is totally or partially consumed, if additional phases are present, i.e. precipitates or crystals. We assume that the resulting solution at

concentration c_{eq} is sufficiently dilute that we can treat it as an ideal solution with $\gamma_{eq} = 1$. The change in the chemical potential during this process is given by $\Delta\mu = \mu(c_{eq}) - \mu(c_l) = \mu(c_{eq}) - \mu(c_i)$, where we have employed the assumption that the chemical potential of the low density phase equals that of the initial solution (chemical equilibrium along the tie line of the LLPS). With the chemical potential given in Eq. (4.1) we can express change in the chemical potential as

$$\Delta\mu_a = RT \ln\left(\frac{c_{eq}}{\gamma_i c_i}\right), \tag{4.2}$$

where the index a in Eq. (4.2) indicates that we employ the *activity* formula of the chemical potential. Note that in this expression, c_i and c_{eq} are experimentally accessible, while γ_i is unknown.

Following McMillan and Mayer [76], we can also treat our system as an effective two-component system of a solvent, including ions, which we denote as component 1, and the proteins, which is component 2. The chemical potential of component 1, μ_1, can be expanded into a virial series in the concentration c of the proteins to obtain

$$\mu_1 = \mu_1^{\ominus} - RTV_1 c_2 [\frac{1}{M} + B_2 c_2 + B_3 c_2^2 + \dots], \tag{4.3}$$

where V_1 is the molar volume of the solvent, and M the molecular weight of the protein. B_2, B_3, ... are the second, third, and higher virial coefficients. If the protein solution is sufficiently dilute, as we have assumed in the low density phase, it is possible to truncate the virial expansion at the second term [43]. By differentiating μ_1 w.r.t. the protein concentration

$$\left(\frac{d\mu_1}{dc_2}\right)_{p,T} = -\frac{RTV_{m,1}}{M}[1 + 2B_2 M c_2], \tag{4.4}$$

and employing the Gibbs-Duhem relation [25], it is possible to connect the chemical potential of the solvent with the chemical potential of the proteins to obtain

$$\left(\frac{d\mu_2}{dc_2}\right)_{p,T} = -\frac{M}{c_2 V_{m,1}}\left(\frac{d\mu_1}{dc_2}\right)_{p,T}. \tag{4.5}$$

Combining Eqs.(4.4) and (4.5), we can calculate the chemical potential, relative

to a low density reference state with concentration c_0 as

$$\mu(c) = \mu_0 + RT \ln\left(\frac{c}{c_0}\right) + 2RTB_2M(c - c_0), \tag{4.6}$$

which is an alternative to Eq. (4.1), as long as the virial expansion can be truncated after the second term. We are now in the position to express the change in the chemical potential from the initial solution at concentration c_i to its final state with concentration c_{eq} by

$$\Delta\mu_{MM} = RT \ln\left(\frac{c_{eq}}{c_i}\right) + 2RTB_2M(c_{eq} - c_i), \tag{4.7}$$

which is an alternative expression to Eq. (4.2), where $\Delta\mu_{MM}$ indicate the chemical potential obtained from the McMillan-Mayer theory. By demanding that $\Delta\mu_a = \Delta\mu_{MM}$, we find that the second virial coefficient takes the form,

$$B_2 = -\frac{\ln(\gamma_i)}{2M(c_{eq} - c_i)}, \tag{4.8}$$

which still contains the unknown activity coefficient γ_i.

It is essential to note that the following considerations are primarily empirical, and the actual compositions, denoted by c_i, may not always align precisely with the tie line described by the Maxwell construction. Additionally, the equilibrium composition c_{eq} might fall outside the tie line, indicating deviations from the behavior that we want to present. In order to express the second virial coefficient solely by quantities that are accessible in the experiment we consider again the chemical equilibrium between the initial solution and the low density phase along the tie line, $\mu(c_i) = \mu(c_l)$, from which the following equation is obtained

$$\ln(\gamma_i) = \ln\left(\frac{\gamma_l c_l}{c_i}\right) \approx \ln\left(\frac{\gamma_{eq} c_{eq}}{c_i}\right) = \ln\left(\frac{c_{eq}}{c_i}\right), \tag{4.9}$$

where we have approximated the activity of the low density phase, which is not directly accessible, by that of the equilibrium solution, for which we have assumed that its activity coefficient $\gamma_{eq} = 1$. With this assumption we reach our theoretical main result, an estimate of the second virial coefficient based on

quantities that are easily accessible:

$$B_2 = -\frac{\ln\left(\frac{c_{eq}}{c_i}\right)}{2M(c_{eq} - c_i)}. \tag{4.10}$$

Note that assuming an interaction potential with a fixed interaction distance, like in the SAXS fitting procedure, or using a linear interpolation of the light scattering intensity like in the DLS framework, the second virial coefficient, should not depend on c_i, despite that, dependence on initial concentration was already found in systems of cyclodextrins, where the second virial coefficient is calculated from the structure factor and follow similar concentration dependencies as described by the Eq. (4.10) [63, 64].In order to remove the concentration dependence two assumptions can be adopted. The first one is to average the results over the concentration interval explored, keeping the salt-protein ratio constant, the second one is to normalize the values with respect to the first concentration point at any given initial concentration where LLPS is showed, as described in Sec.IV D. The second virial coefficient obtained is normalized by the second virial coefficient of hard spheres [110]. In order to demonstrate the validity of Eq. (4.10) we apply it to different protein solutions that display LLPS, induced by trivalent salts.

4.3 Materials and sample preparation

Proteins and salt were purchased from Merck, and used as received. The purities were 98% for BSA (product no. A7906), 97% for HSA (product no. A9511) and 99.99% for CeCl3 (product no. 429406). Stock solutions were prepared by dissolving the protein and salt in deionized (18.2 MΩ), degassed Millipore water. The resulting concentration of protein was determined with an ultraviolet visible (UV-vis) spectrophotometer (Cary 50 UV-vis spectrometer, Varian Technologies) using an extinction coefficient of 0.667 mL mg^{-1} cm^{-1} for BSA, and 0.531 mL mg^{-1} cm^{-1} for HSA at a wavelength of 278 nm [30]. All samples were prepared by mixing deionized, degassed Millipore water, protein stock solution and salt stock solution. All samples had a pH (between 6.3 and 7.0) above the respective pI of HSA and BSA, measured with a pH-Meter from Mettler Toledo (Germany). No buffer was added as neutral trivalent salts (i.e, $CeCl_3$) do not induce signifi-

cant pH variation. All samples were prepared, stored and investigated at 21 ± 1 °C. When the LLPS is approached in the phase diagram, Fig.1, the solution phase-separates into a dense, yellowish phase and in a low density, dilute phase. In addition, for HSA, crystallization was also observed [12], during this process the dense phase is almost totally consumed [70]. After 14 days, the concentration of the resulting dilute equilibrium phase was determined via UV-vis spectroscopy.

4.3.1 Salt versus protein concentration phase diagram

Samples of protein concentrations (c_p) at 35, 50, 65, 80, and 100 mg/ mL were prepared for BSA and HSA, varying salt concentrations (c_s). The mean value of c_s of the last clear and first turbid sample is referred to as $c^\star$ and the last turbid and first clear sample as $c^{\star\star}$. The macroscopic phase separation was ensured by visual inspection. No additional investigation with respect to the distribution of the salt in the two phases was made, and the concentration of the high density phase was not determined. Therefore, the points depicted in the phase diagram refer to the preparation conditions. For the complete knowledge of the LLPS loop additional analysis is needed, such as those performed in the references [73, 135, 138]. The phase diagram depicted in Fig. 5.2 shows that for BSA the amount of salt required to reach the $c^\star$ border is higher than for HSA. Such behavior could be due to the different surface charge of the proteins at such condition. BSA is slightly more negative with an overall charge of -11 compared to the value of HSA, which is around -9 [69]. Therefore, more salt has to be added to neutralize the excess negatively charged residues with counterions and trigger the aggregation. Also, the different locations of the $c^{\star\star}$ borders could be explained as a consequence of an increased amount of counterions in BSA compared to HSA to trigger the pair repulsion (reentrant condensation).

4.3.2 Small-Angle X-ray Scattering

SAXS measurements were performed at the Petra III beamline P12 (Hamburg, Germany). All the details related to the experimental procedures and the data analysis are available in reference [69]. For the determination of the reduced osmotic second virial coefficient $B_2' = B_2 / B_{HS}$ the data were fitted with a sticky hard sphere (SHS) as explained in Sec. 2.3.3, provided in the NIST macros, using

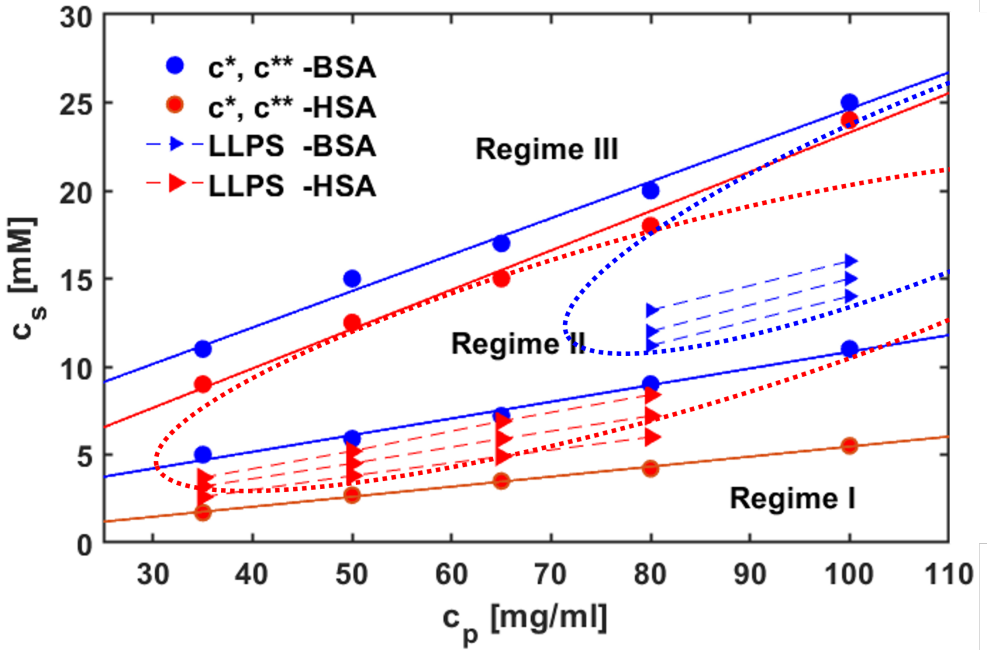

Figure 4.1: The experimental phase diagram displays triangular points that indicate the conditions where the samples inside the LLPS were prepared. Points that share the same c_s over c_p ratio are connected by dashed lines. The phase diagram boundaries are marked by full dots connected by solid lines. The c^* border is indicated by lower solid lines, while the upper solid lines represent the c^{**} border. Moreover, the phase diagram includes red and blue elliptical regions that represent a hypothetical complete LLPS loop for HSA and BSA [69].

IGOR Pro 6.37.

4.4 Discussion of salt versus protein concentration phase diagram

The determination of the phase diagram was carried out by visual inspection. For small amounts of salt, the overall interactions between proteins are repulsive, resulting in a clear solution without macroscopic aggregation (Regime I). Increasing the salt concentration the system reaches the boundary called $c^{\star}$ where the last clear sample and the first turbid was observed (Regime II). Here, the interactions between proteins are mostly attractive due to the ion bridging effect [103]. Further addition of trivalent salt ($CeCl_3$) moves the system towards the second boundary $c^{\star\star}$ where the resulting solution becomes clear again (Regime III or reentrant). A possible explanation of this behavior is the charge inversion of the proteins due to the complete saturation of the binding sites with trivalent ions, leading to strong repulsive interactions [103]. The experimental data points are fitted using a linear regression. It is interesting to note that the critical threshold value of protein concentration for macroscopic LLPS is shifted by about two-three times if we compare HSA and BSA (35-40 mg/ml and 72-80 mg/ml) for metastable LLPS [69, 71]. This behavior reflects the trend observed in the calculation of the osmotic second virial coefficient from supersaturation measurements explained in the next chapter, where it is observed that the values of the second virial coefficient for BSA is 2 to 3 times bigger than the values of the HSA system, in agreement with the magnitude of the attractive interactions found in this work. If the interactions are not sufficiently attractive at low concentration of protein, an increasing protein concentration greatly shortens the distance between proteins enhancing the likelihood of formation of a critical cluster large enough to allow the nucleation of a new phase. An additional threshold in concentration in Regime II is reached at high protein concentration, where the system does not undergo phase separation. This might be due to additional short range repulsion that can affect the overall attractive interactions in crowded environments as well as different entropic contribution [139]. Additional investigations are needed to understand protein interactions in Regime II at higher protein volume fraction, where additional phenomena are observed for several systems, such as kinetic aggregation and amorphous precipitation or gelation [22, 24].

4.5 Real time observation of phase separation

The phase separation was followed by visual inspection in glass vials and on glass slides under the microscope. The general behavior observed reveals that close to the binodal, the solutions form a stable high density phase, in the case of BSA, and a meta-stable high density phase, in the case of HSA. This is connected with a different crystallization behavior of the two proteins. In fact, as shown in our previous work a more stable crystalline phase exist for HSA at such condition while for BSA, crystallization does not take place. This is due to the different intermolecular interactions between this two proteins, since the BSA is more hydrophilic, additional hydrophobic interactions necessary for crystallization are suppressed. Therefore, for the HSA solution the LLPS is formed as kinetically driven process upon salt addition followed by the formation of the crystals from the low density phase that is the thermodynamic stable phase as here shown:

$$HSA(dense) \rightleftharpoons HSA(dilute) \rightarrow HSA(xtal).$$

The system first phase separates in a high density and low density phase until a dynamic equilibrium is reached between them. They remain stable until the first crystals appear in the low density phase. Therefore, the variation of the osmotic pressure between the two previously coexistent phases leads to the consumption of the dense liquid droplets to preserve the initial equilibrium. However, if the driving force of the crystallization process is sufficiently strong, an additional three phases equilibrium cannot be reached and the process continues until the complete high density phase is consumed and the equilibrium concentration c_{eq} is reached, which corresponds to the solubility limit [69]. The growth of the crystals generally is completed between 7 and 10 days at these conditions. In order to ensure that the phase conversion was completed, a period of 14 days from the sample preparation was chosen for our supersaturation measurements.

4.6 Calculation of the second virial coefficient from the supersaturation

Here, the experimental setup for the calculation of the second virial coefficient and the results are discussed. Upon the addition of $CeCl_3$ the samples in glass

vials, sealed with parafilm, were kept at temperature of 21 ± 1 °C for 14 days. This was done to ensure complete phase separation and crystallization. If HSA samples after preparation are centrifuged at 5000 rpm for 10 min and the high density phase is precipitated, the resulting concentration of the low density phase is always higher than the equilibrium concentration obtained 14 days after preparation. This behavior is essentially due to the crystallization that affects the final equilibrium concentration. For the BSA samples, which only show phase separation but no crystallization, the equilibrium concentration after 14 days is close to the values obtained after centrifuging. Therefore, to consider the total contribution (phase separation and crystallization) on the partitioning, only the values of the concentration obtained after 14 days are employed for the estimation of the osmotic second virial coefficient. Considering the normalized second virial coefficient obtained as a ratio of Eq. (4.10) and B_{HS} in ml·mol·g^{-2}, the only parameter not known experimentally is the hydrodynamic hard sphere radius. Hence, for the calculation of the hydrodynamic radius a previously resolved crystallographic structure in presence of YCl_3 was used as an input in the software HullRad from fluidic Analytics, which uses a convex hull model to estimate the hydrodynamic volume of a macromolecule [32]. Further details about this method are given in references [6, 32]. The resulting value is estimated to R_h = 3.55 nm and was used together with a molecular weight of 66.4 kDa for the calculation of the normalized second virial coefficient of both HSA and BSA from the supersaturation measurements. A similar value for R_h was found experimentally using SAXS on BSA diluite solutions in reference [102], where the scattering curves are well fitted by a prolate ellipsoid with an effective hydrodynamic radii of R_h = 3.62 nm. In the same reference a value of R_h = 3.66 nm obtained via DLS is also found. For the calculation of the second virial coefficient from the SAXS fitting procedure, an ellipsoid form factor was chosen. The ellipsoid form factor chosen was the same for both the proteins, with axes fixed to r_a = 1.8 nm and r_b = 6.1 nm. The results for the normalized osmotic second virial coefficient for HSA obtained from our approach and from SAXS measurements are plotted in Fig. 4.2. Both experiments were performed under similar conditions and the overall agreement, despite the fact that in the different experiments B_2 is extracted from different data, is good.

For BSA the experimental conditions explored with SAXS were far from the LLPS border and additional analysis shall be performed for direct comparison of

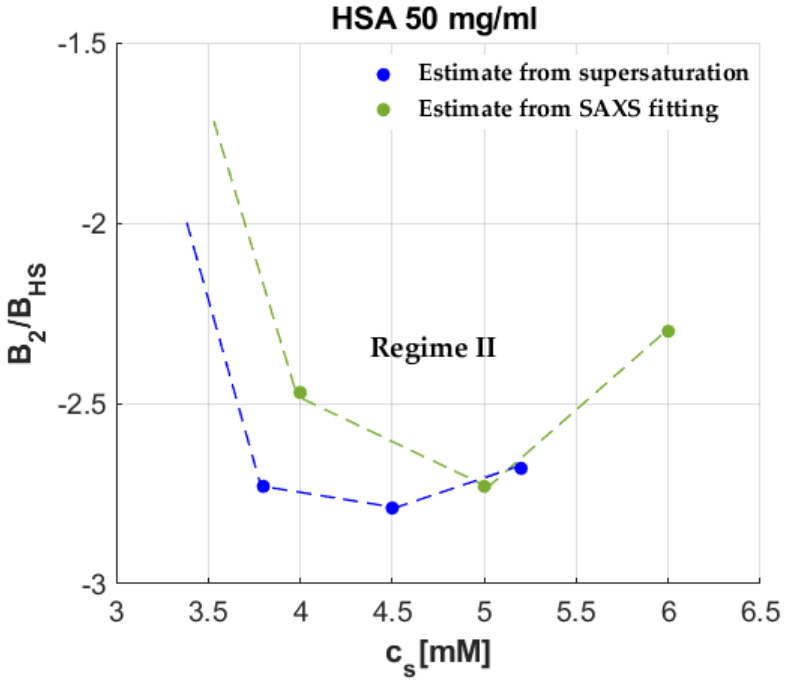

Figure 4.2: Normalized second virial coefficient estimate from supersaturation and SAXS fitting procedure, in the SAXS data points, the errors bars are smaller then the symbols shown here. For the points obtained by supersaturation measurements the error is estimate to be around 10% – largely due to the uncertainty of the concentration measurements, determined via UV-vis spectroscopy at equilibrium. The dotted lines act as a guide for the eyes. Although the different routes to B_2 make use of different data, the agreement between the estimates is good.

the B_2' calculated with these two methods for the BSA-$CeCl_3$ system. However, a good agreement was found comparing the results obtained at 80 mg/ml from supersaturation measurements of BSA with $CeCl_3$ with those obtained with the SAXS fitting procedure at 80 mg/ml BSA concentration with $LaCl_3$ from ref. [74] as shown in Tab. 7.1. Clearly, the effect of salt should produce differences in the magnitude of B_2', due to effects such as differences in binding sites affinity on the surface of the protein or ions polarizability that leads to different intermolecular interactions [19, 26]. We have chosen solutions with $CeCl_3$ for comparison, because these exhibit LLPS with to respect both HSA and BSA, while for $LaCl_3$ we obtain LLPS with HSA, but not for BSA. In BSA with $LaCl_3$ we observed transitions between Regime I (clear) – Regime II (turbidity without LLPS) and Regime II (turbidity without LLPS) - Regime III (reentrant). In principle we could induce partitioning in BSA with $LaCl_3$ by centrifuging

HSA				
c_p[mg/ml]	c_s[mM]	c_s/c_p	$B_2'(supersaturation)$	$B_2'(SAXS)$
35	2.6	5	-3.85	
35	3.2	6	-3.92	
35	3.7	7	-3.81	
50	3.8	5	-2.73	-2.47 (c_s/c_p = 5.3)
50	4.5	6	-2.79	-2.72 (c_s/c_p = 6.7)
50	5.2	7	-2.69	-2.31 (c_s/c_p = 8.0)
65	4.9	5	-2.24	
65	5.9	6	-2.31	
65	6.9	7	-2.28	
80	6	5	-1.86	
80	7.2	6	-1.94	
80	8.4	7	-1.88	

Table 4.1: Normalized second virial coefficient values for HSA obtained via supersaturation. The results at 50 mg/ml are compared with those obtained from SAXS fitting. The error of the second virial coefficient obtained from supersaturation is about 10%. B_2'. Values of c_s/c_p are approximated by an integer number.

the samples, while with $CeCl_3$ this is not necessary. Since our method relies only on thermodynamical assumptions, we apply the method exclusively on systems that show spontaneous driving force for LLPS. Independent of the type of protein used in this work, the general trend observed in both protein systems was an increasing of second virial coefficient as the protein concentration is increased, as shown in Tabs. 4.1 and 7.1. It is not clear if this behavior is directly derived from the mass transfer treatment discussed in Sec. 4.2 or if it is because there are deviations from ideality at higher protein concentration and the approximation γ_{eq} = 1 might not be valid even at the protein volume fraction of the low density equilibrium phase. However, this trend seems to describes the phase behavior of colloidal fluids near coexistence [77, 80, 128]. For HSA with $CeCl_3$ at this condition, the solubility line is located at ≈10 mg/ml and the LLPS border is around ≈ 30 mg/ml. In comparison, BSA solution is far more soluble than HSA and a smaller equilibrium concentration found was around 67 mg/ml. The values of the solubility line and LLPS borders depend of the initial protein concentration, salt concentration and their ratio. Using the data obtained in Tab. 4.1, as well as the initial concentration of HSA and the equilibrium one, it was possible to rebuild an experimental and qualitative phase diagram that resembles the theoretical diagram for colloidal fluids Fig. 4.3. For the complete phase diagram additional samples at higher protein concentration

could be investigated as this might provide additional information for the observed trend of the second virial coefficient. In fact, from the theory of colloidal fluids a turning point should be reached, corresponding to the critical point, and after that a decrease in B_2' should be observed at higher protein volume fractions. However, this could not be detectable with the previous assumptions for the limitation due to the approximations made. Furthermore, for a complete description of the phase diagram, the concentration in the high density phase should be determined so that the border at higher protein concentrations of the LLPS can also be calculated. In Fig. 4.3 a straight dotted line is used to describe the concentration of the low density phase after phase separation. This line could in principle be determined experimentally, if the sample is centrifuged at 5000 rpm for 10 minutes immediately after salt addition, to separate the contribution of LLPS (a kinetically driven process) to the contribution of the crystallization (a thermodynamically driven process).

BSA				
c_p[mg/ml]	c_s[mM]	c_s/c_p	$B_2'(supersaturation)$	$B_2'(SAXS)$
50	7	9	NO	-2.41
50	8	11	LLPS	-2.43
50	9	12		-2.43
80	11	9	-1.0	
80	12	10	-1.0	
80	13	11	-0.96	
80	8	7		-1.0 ($LaCl_3$)
80	12	10		-1.5 ($LaCl_3$)
80	17	14		-1.2 ($LaCl_3$)
100	14	9	-0.79	
100	15	10	-0.8	
100	16	11	-0.83	

Table 4.2: Normalized second virial coefficient values for BSA obtained via supersaturation. Note that no phase separation was observed at 50 mg/ml. The estimated error on the second virial coefficient is about 10%. For 80 mg/ml the results are similar to those obtained with $LaCl_3$ from ref. [74]. Values of c_s/c_p are approximated by an integer number.

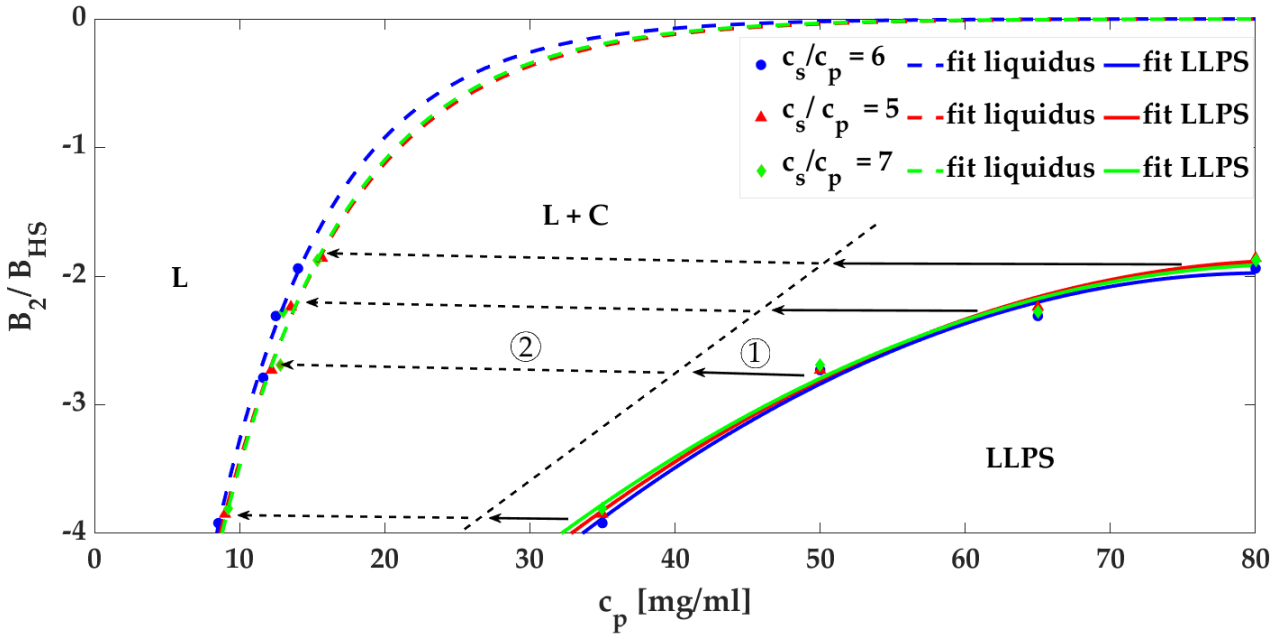

Figure 4.3: Experimental phase diagram of HSA with $CeCl_3$. The position of the solubility line (liquidus) follows the magnitude of the interactions and the different partitioning effect at different protein salt-ratios. For c_s/c_p = 6 the solubility of the protein diminish due to the stronger attractions. The overall mechanism of LLPS and crystallization of HSA in presence of $CeCl_3$ might be explained as follows: the system from the initial preparation conditions phase separate in a low and in a high density phase (1), from the low density phase the crystals start to nucleate and the consumption of material proceed until the solubility line is reached (2). The amount of high density and low density phase produced follow the arrows intensities.

4.7 Variation of the chemical potential near coexistence

In order to take into account variations of the chemical potential and to extend the range of application of the derived formula outside phase coexistence, it is possible to normalize the previous relation, Eq. (4.10) with respect to the minimum concentration where LLPS is shown $\tilde{c}$:

$$B_2 = -\frac{\ln\left(\frac{c_{eq}}{c_i}\right)}{2M(c_{eq}-c_i)}\frac{c_i}{\tilde{c}} \tag{4.11}$$

where $\tilde{c}$ is the lowest protein concentration for which LLPS is observed a given temperature and pressure. Taking into account this change in the chemical potential it is also possible to re-scale the second virial coefficient in such way

that at constant supersaturation correspond always the same value of the second virial coefficient. Using this method, the concentration dependence vanishes. However, this is an empirical correction and additional studies are needed to explore the validity of the assumptions done. The minimum concentration $\tilde{c}$ where LLPS is shown for HSA at salt-protein ratio of 6 is around 35-40 mg/ ml and for BSA, at salt-protein ratio of 10 the minimum concentration $\tilde{c}$ is around 72-80 mg/ml. By using those values, the normalized second virial coefficient becomes, -3.4 for HSA at initial concentration of 35 and 50 mg/ml. The HSA system shows also an increased supersaturation by increasing protein concentration, by taking into account this behavior, the resulting second virial coefficient normalized for the minimum concentration where LLPS is shown $\tilde{c}$ at initial concentration of 65 and 80 mg/ml is approximately -3.7. For BSA the initial-equilibrium concentration ratio is approximately the same at each condition explored, around 1.2. Normalizing such system with respect to the critical concentration of 72 mg/, we obtain a normalized second virial coefficient of about -1.12 close to the theoretical limit for phase separation of sticky hard sphere if specific energies routes are followed [135]. With this information, we speculate that for the BSA system we are moving close to the critical point in the phase diagram while for HSA we observed the optimal crystallization window below the critical point as predicted by G. A. Vliegenthart and H. N. W. Lekkerkerker [130].

4.8 Conclusions

In conclusion, we calculated the normalized second virial coefficient for a protein solution close to LLPS using a thermodynamic approach for a two component system under the assumption that salt addition only changes the surface charge density of the proteins. Furthermore, we linked the macroscopic behavior of protein partitioning with the magnitude of the second virial coefficient that is a free model calculation and experimentally accessible. The magnitude of the second virial coefficient follows the partitioning ratio between two phases as previously observed for a different system [74, 135]. The mass transfer consideration might be in principle applied also outside the LLPS region, when the system reaches the solubility line. However, this strongly depends of the protein phase behavior, the molecular weight and the precipitation agent employed.

Our expression was developed with the aim of being applicable to solutions containing two to three components, which exhibit both stable and metastable liquid-liquid phase separation (LLPS). This includes globular proteins with molecular weights ranging from 10 to 400 KDa, which covers, all the proteins investigated in reference [26]. We have compared our results with those obtained from a different method finding a good agreement. The main constraint of the thermodynamical relation adopted is its 2-component approach. In fact, we access the chemical potential of the protein using the chemical potential of the solvent and the Gibbs-Duhem equation. As many solutions contain multiple components, such as buffers, salts, proteins, PEG (polyethylene glycol), water, preservatives, impurities and surfaces of glass vials or containers, the LLPS process and the supersaturation measurements may be influenced by the presence of those compounds, decreasing the accuracy on the estimation of the second virial coefficient. Another constraint stems from the prerequisite understanding of the phase diagram and accurate identification of the binodal region. Additionally, smaller negatively charged proteins may experience the effects of bridging, resulting in significant kinetic amorphous aggregation or a greater driving force for crystallization than observed in the current system. When the interactions between proteins are too strong, the dynamic aspect of the interactions plays a crucial role on determine the final state of the system. At this condition, Monte Carlo simulations of particles with hard cores and isotropic, square-well interactions, using the fluctuation-dissipation theorem, show that if no ordered bounds are present, like amorphous aggregation, the system do not relax in the more favourable thermodynamical state. Consequently, if not ordered cluster are formed the bonding breaking energy is too high that the system is confined in an energetic trap [60] . In fact, the pathway for the formation of a new phase can be near equilibrium or far from it. The liquid-liquid phase separation process always competes with other forms of self-organization, including kinetic intermediates as previously discussed by Whitelam and Jack [134]. Under these conditions the treatment used could become inaccurate. However, by taking into account the distance from the LLPS border, with the normalization used in Eq. (4.11) it might be possible to estimate the second virial coefficient also outside the binodal region. The accuracy of the treatment strongly decreases if after phase separation the concentration of the protein in the low density phase is still high, probably due to protein-protein

interactions still present in this phase that should be also taken into account. In this work we have shown that under the conditions investigated the magnitude of the second virial coefficient calculated via supersaturation measurements is in agreement with the values obtained via SAXS fitting procedure. Working on the basic ideas presented here, might lead to a different approach to estimate the second virial coefficient that do not make use of expansive and advanced techniques. The method in fact could be employed routinely in any protein crystallization laboratories, without the need of training to perform complicated experiments. Furthermore, the method could be useful not only for proteins-salt solutions, but also for different colloidal systems that show crystallization or stable and metastable liquid-liquid phase separation.

Chapter 5

The Ion-Activated Attractive Patchy Particle Model and Its Application to the Liquid-Vapour Phase Transitions

5.1 Introduction

Patchy particles are a class of complex colloidal particles with anisotropic surface chemistry that enables them to selectively interact and bind with each other in a specific orientation. In recent years the interest in patchy particles model is increased for the interesting application in materials science, physics, chemistry, and biology due to their potential applications in self-assembly, drug delivery, catalysis, and many other fields [93, 107, 115]. The flexibility of patchy particle models allows for a wide range of possible interactions and structures to be studied [9, 21, 31]. These models can incorporate both directional and isotropic interactions between particles, as well as steric effects, to accurately capture the behaviour of complex systems [44, 95, 108]. We want to present an extension of our ion activated patchy particle model that we used before to capture the phase behaviour of protein-salt mixtures [35, 36, 103]. The aim of this work is to understand the effect of different parameters included in our model on the behaviour of the liquid-vapour phase transition, we will make use of the Wertheim theory, that provides a theoretical framework for predicting the thermodynamics properties of the patchy particle systems [131–133]. Additionally, We will show how with our approach, the anisotropic behaviour of the system arises from the different population of particles with occupied and unoccupied patches upon ion binding to the surface. In this regard our system can be seen as an effective multicomponent system, composed by particles with

different occupied and unoccupied binding site.

5.2 Theory

5.2.1 Ion Activated Patchy Particles Model

In this section we focus on the basic theory of the ion activated patchy particles model [103]. In our patchy particles system we assume that the probability to have an occupied patch, that is an ion bound to a binding site, as function of the salt or ions concentration $\Theta(c_s)$ is given by a Fermi-like distribution in the grand canonical ensemble (GCE):

$$\Theta = \frac{1}{1 + \exp\left(\beta(\epsilon_b - \mu_s(c_s^r)\right)}, \tag{5.1}$$

where $\beta = 1/(k_B T)$ is the inverse temperature, ϵ_b is the binding energy between a salt ion and a patch on the protein surface. We assume that the patches are independent and possess the same energy ϵ_b, which is kept constant and independent of the salt concentration. μ_s is the chemical potential of the salt in the reservoir, that can be approximate by the ideal gas expression $\mu_s^r(c_s^r) \approx k_B T \ln(c_s^r/\rho_0)$.

In our model, with the assumption of m independent patches per protein, the probability of finding m patches occupied by i ions is given by the binding probability Θ of a single patch via a binomial distribution

$$p(m, i) = \Theta^i (1 - \Theta)^{m-i} \binom{m}{i}, \tag{5.2}$$

where $q = (1 - \Theta)$ is the non binding probability. The overall patch-patch interaction energy between patches of different particles is given by

$$\beta\epsilon_{pp} = \beta\epsilon_{uu}(1 - \Theta)^2 + 2\beta\epsilon_{uo}\Theta(1 - \Theta) + \beta\epsilon_{oo}\Theta^2, \tag{5.3}$$

where ϵ_{uu} is the interaction energy between two unoccupied patches, ϵ_{oo} is the interaction energy between two occupied patches, and ϵ_{uo} is the contribution to the interactions between an occupied and an unoccupied patch.

Previously [103] we have simplified our model by assuming $\epsilon_{oo} = 0 = \epsilon_{uo}$. While we have found reasonable agreement between experiments of protein-

salt mixtures and predictions of our model, we want to explore the richness of the model by systematically studying the influence of the various parameters, including ϵ_{oo} and ϵ_{uo}, on the behaviour of the model. We expect that ϵ_{oo} and ϵ_{uo} account for repulsive contributions to the the overall interaction energy between patchy particles, as they account for interactions between patches with the same electrical charge.

5.2.2 Thermodynamics Model

The fundamental thermodynamics behaviour of our patchy particle model is based on the Wertheim theory. In this framework, the free energy density is given by the sum of the free energy density of the reference system and a perturbation contribution due to bonding between particles. In our work we use a hard-sphere fluid as reference system and employ the accurate thermodynamics based on the Carnahan-Starling [14] equation of state. Clearly, it would be possible to replace the hard-sphere fluid by a more general reference system. To this end one would have to replace the chemical potential and the pressure of the hard-sphere fluid by those of the reference system of choice.

The contribution due to bonding between patches of different particles is given by

$$\beta f_{bond} = m\rho \left(\ln(1 - p_b) + \frac{1}{2} p_b \right), \tag{5.4}$$

where $\rho = \frac{N}{V}$ is the bulk number density, m, as before, the number of patches per particle and p_b the probability of a patch having formed a bond. Note that p_b depends on the number density ρ and follows from the mass-action equation

$$\frac{p_b}{(1 - p_b)^2} = m\rho\Delta, \tag{5.5}$$

where Δ accounts for the spherical averaged interaction between bonds of patches of two particles. Here we follow Refs. [52, 103] and assume a hard-sphere reference system and a short ranged interaction between patches and obtain

$$\Delta = 4\pi g_{HS}(\sigma) K F, \tag{5.6}$$

where $g_{HS}(\sigma)$ is the contact value of the radial distribution function of the hard-sphere reference system, K account for the geometry of the patchy interaction

sites and F is the angular average of the Mayer-f function of the patch-patch interaction.

The total free energy density in our model is the sum of the ideal gas, the hard-sphere and the Wertheim contribution. The resulting chemical potential of the system consists therefore also of three terms, the ideal gas chemical potential, the excess chemical potential of the hard sphere reference system and the bonding term of the chemical potential

$$\beta\mu = \underbrace{\ln(\eta)}_{\beta\mu_{\text{id}}} + \underbrace{\frac{3\eta^3 - 9\eta^2 + 8\eta}{(1-\eta)^3}}_{\beta\mu_{\text{ex}}^{\text{CS}}} + \beta\underbrace{\frac{\partial f_{\text{bond}}}{\partial\rho}}_{\mu_{\text{bond}}}. \tag{5.7}$$

Here $\eta = 4\pi R^3\rho/3$ is the packing fraction of spheres with radius R. For convenience we have chosen without loss of generality the thermal wavelength λ in the ideal gas chemical potential so that $\lambda^3\rho = \eta$. The excess chemical potential of the reference system is here given by the expression of Carnahan-Starling [14].

The total pressure of the system is the sum of the Carnahan and Starling contribution βP_{cs} [14] plus the contribution given by the bonding term:

$$\beta P = \underbrace{\rho\frac{1+\eta+\eta^2-\eta^3}{(1-\eta)^3}}_{\beta P_{\text{cs}}} + \beta\left(\rho\mu_{\text{bond}} - f_{\text{bond}}\right). \tag{5.8}$$

While for the hard-sphere reference system the thermal energy $\beta = 1/(k_B T)$ is a trivial scaling factor, the bonding contributions are sensitive to it. For sufficiently low temperature the chemical potential and the pressure can develop a van-der-Waals loop that indicate the possibility for a liquid-liquid phase transition.

5.3 Results

In this section we study the effect of the different parameters on the liquid-liquid phase behaviour of the ion-activated patch model of proteins. Within the model we can adjust and change each parameter independently and thereby obtain an intuition of the influence of each parameter on the phase boundaries

5.3.1 Interaction energy curves

To this end we start with the interaction energy curves given by Eq. (5.3). In a previous study the parameters ϵ_{oo} and ϵ_{uu} were set to zero and all the focus was put onto the attraction between an unoccupied and an occupied patch via ϵ_{uo}. It was found [103] that in this case the minimum of the energy curve is at $\Theta = 1/2$ and that this minimum has to be sufficient deep (negative) for the system to phase separate.

In order to display the influence of the parameters ϵ_{oo} and ϵ_{uu} on the overall interaction energy between particles ϵ_{pp}, we compare it to the result of our previous model where those parameters were set to zero. In Fig. 5.1, we plot the reference curve where no repulsion is present, together with the curves obtained including the repulsive contributions. The value of the attractive interaction parameter $\beta\epsilon_{uo}$ is fixed so that it produces curves with the same minimum $\beta\epsilon_{min}$, for a given set of parameters the condition is given by

$$\beta\epsilon_{uo} = -\beta\epsilon_{min} - \sqrt{(\beta\epsilon_{uu} + \beta\epsilon_{min})(\beta\epsilon_{oo} + \beta\epsilon_{min})}. \tag{5.9}$$

The results in Fig. 5.1 show that including the repulsive parameters generate an asymmetry in the interaction energies curves. In Fact, we find that for the energy curve with no repulsions, the interaction energy curve is symmetric with the minimum located at $\Theta = 0.5$ (red curve), while for the other curves the location of the minimum is shifted to $\Theta \approx 0.44$ for $\beta\epsilon_{uu}$ = 2 , $\beta\epsilon_{oo}$= 8 and $\beta\epsilon_{uo}$ = -20.42 (blue curve) and $\Theta \approx 0.58$ for for $\beta\epsilon_{uu}$ = 12 , $\beta\epsilon_{oo}$= 3 and ϵ_{uo} = -22.60 (magenta curve). It is worth noticing that changing those parameters do not lead to a change in the probability distribution $p(m, i)$ at a given value of Θ, but the same histograms might refer to different energy values in the interaction energy curves Eq. (5.3). This is an intriguing behaviour of our model that we want to underline, since different experimental systems, for example proteins, are proved to be very sensitive on the type of salt used. In our model this effects can be described by the parameters $\beta\epsilon_{oo}$ and $\beta\epsilon_{uo}$. This in turn can change the resulting protein phase behaviour. In a sense our system, which is a mixture of proteins and salt ions, can be interpreted as a multi-component system of proteins with a different number of salt ions bound. As shown in Fig. 5.1 a given value of Θ, the binding probability of an ion to a patch together with the energy parameters determine the protein-protein interaction, Eq. (5.3), and the

distribution of proteins with a different number of ions bound to them, as shown in the histograms. The overall interaction energy is the key parameter to connect our ion activated patchy particle model with the framework of the Wertheim theory. In the next section we will describe the effect of $\beta\epsilon_{uu}$, $\beta\epsilon_{oo}$ and $\beta\epsilon_{uo}$ on the phase diagram of our model.

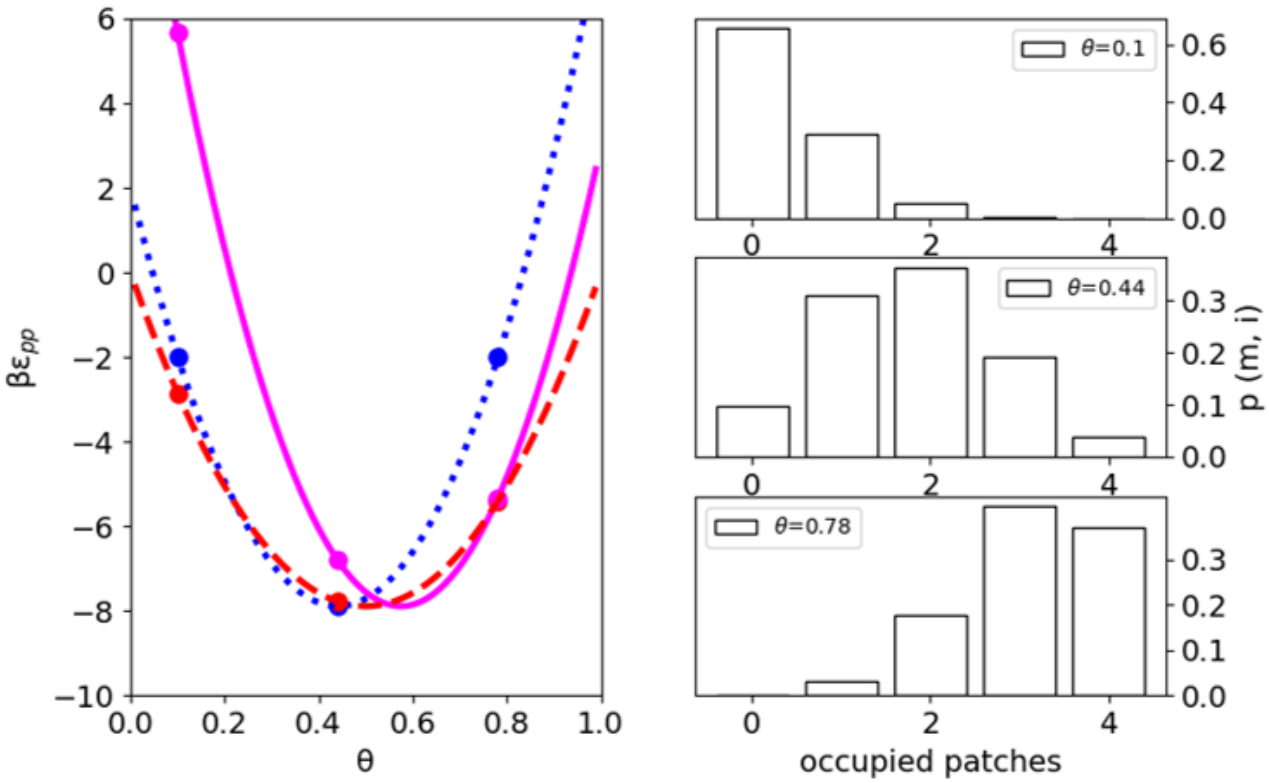

Figure 5.1: Interaction energies curves and corresponding histograms for three different choices of the interaction parameters. For the magenta curve the set of parameters used are $\beta\epsilon_{uu}$= 12, $\beta\epsilon_{oo}$= 3 and $\beta\epsilon_{uo}$= -22.60, for the blue curve $\beta\epsilon_{uu}$= 2, $\beta\epsilon_{oo}$= 8 and $\beta\epsilon_{uo}$= -20.42, and the red curve with $\beta\epsilon_{uu}$= 0 and $\beta\epsilon_{oo}$= 0 and $\beta\epsilon_{uo}$= -15.77. Note that for three different values of Θ, 0.1, 0.44, and 0.78, we show the corresponding histograms for the probabilities of finding proteins with a different number of salt ions bound to it. While the histograms are determined by Θ, the resulting protein-protein interaction energy depends on the energy parameters $\beta\epsilon_{uu}$, $\beta\epsilon_{oo}$ and $\beta\epsilon_{uo}$.

5.3.2 Effect of the interaction energy parameters on the liquid-vapour equilibrium

In order to test the effect of such parameters on the LLPS loops we want to recall the condition for phase equilibrium between a low density phase with

protein packing fraction η_1 and a high density phase at the same temperature with protein packing fraction, namely mechanical equilibrium η_2 $P(\eta_1) = P(\eta_2)$ and chemical equilibrium $\mu(\eta_1) = \mu(\eta_2)$. The Wertheim expressions for the pressure and the chemical potential do not admit analytical solution and therefore the liquid-vapour equilibrium is evaluated numerically. In Fig. 5.2 we show the effect of the repulsion on the shape of the LLPS loop.

As expected from the energy curves shown in Fig. 5.1, changes in the repulsion contributions to the energy curves shift the LLPS loop up or down. Both the pressure and the chemical potential are in fact, dependent on several parameters, including the number of patches m, the radius of the particles R, the packing fraction η, and the parameter F and K, introduced in Eq. (5.6). In our calculations we keep K and F fixed, unless mentioned otherwise. The quadratic form of the interaction energy, Eq. (5.3), implies that for a given interaction energy $\beta\epsilon_{pp}$ there are two different values of Θ or equivalently two different salt concentrations that give rise ro the same protein-protein interaction energy, but with different compositions of occupied and unoccupied patches. Therefore our model predicts a closed loop for LLPS with two critical points [103]. These points are highlighted by symbols in Fig. 5.2, both in the interaction energy curves and in the LLPS diagram. It is worth noting that the effect of repulsion not only shifts the interaction energy curves $\beta\epsilon_{pp}$ to the left or to the right, but also its width. As a result, the height of the LLPS loop is also reduced compared to LLPS loop with zero repulsion parameters.

The location of the critical point in the phase diagram is shifted along the probability or Θ-axes, but not in interaction energy $\beta\epsilon_{pp}$ or packing fraction η. For this set of parameters, the critical value of the protein-protein interaction is approximately at $\beta\epsilon_{pp} \approx -6.38$. However, the distribution of occupied and unoccupied patches at the critical point is significantly different among the different curves, as depicted in the histograms. These differences might impact not only the shape of the coexistence loop and the liquid-vapour equilibrium but also on the more complex features of the phase diagram of patchy particles, for example, in the case of the formation of crystals or amorphous solids, as observed in protein-salt mixtures. Therefore, setting the repulsion properly introduces an asymmetry in the interaction energy curves that makes the resulting behaviour more intriguing if compared with the more simplified model that does not take the repulsion parameters into account. The size of the

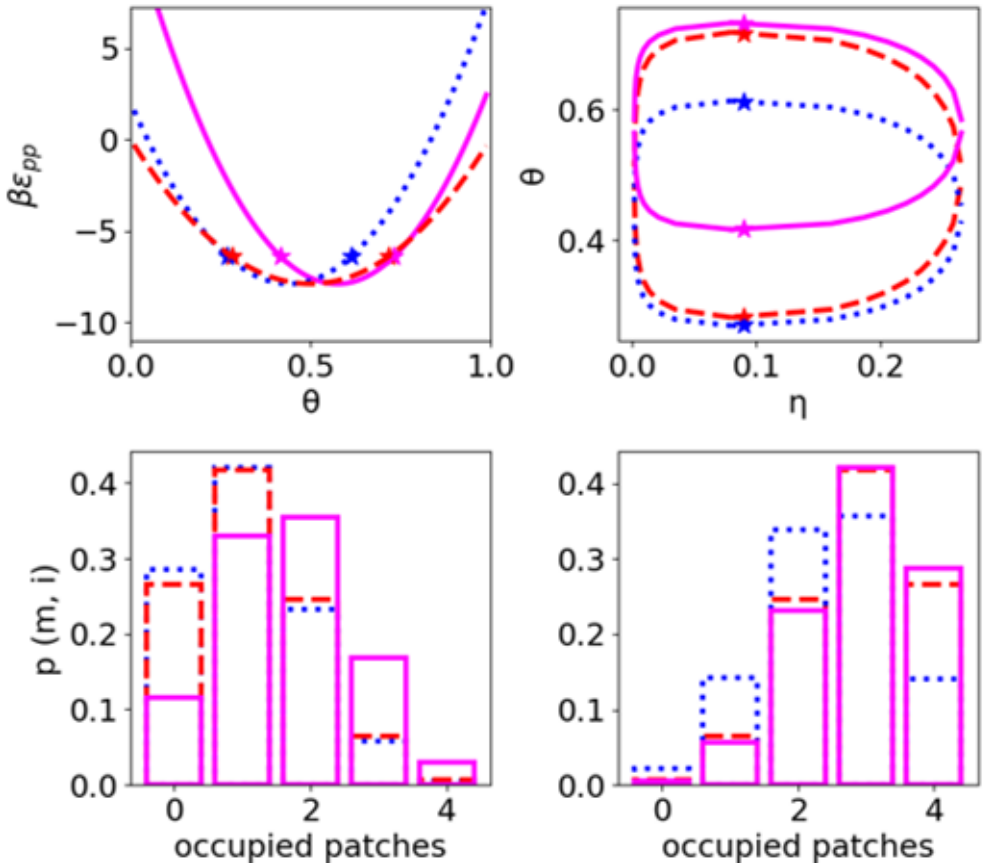

Figure 5.2: Energy curves and corresponding coexistence loops for three different choices of the interaction parameters $\beta\epsilon_{pp}$, as indicated in the label of Fig. 5.1 Note that the energy parameters are chosen so that the minima of the energy curves have the same depth, leading to the same width in the LLPS, while their locations vary from case to case, which causes an up or down shift and a variation in the height of the corresponding LLPS loops. The histograms show the occupancy-distributions for the two critical points for the three choices of energy parameters.

loop can be fine tuned by increasing or decreasing the depth of the interaction energy curve, while the position of the center along the y-axis can be tuned by changing the repulsion and shifting the interaction energy curve along the Θ axis. Those effects are included in Fig. 5.1 and Fig. 5.2. In Fig. 5.3, curves with the same values for the repulsion parameters $\beta\epsilon_{uu}$ and $\beta\epsilon_{oo}$ but with different values of attraction parameter $\beta\epsilon_{uo}$ are shown. Again, the critical points are always located at the same packing fraction value but shifted in probability Θ. The critical value for the packing fraction is equal to $\eta = 0.0898$ for the loops in both Fig. 5.2 and Fig. 5.3. However, the effect of increasing the attraction is not only changing the dimension of the loop but also, as shown in the previous

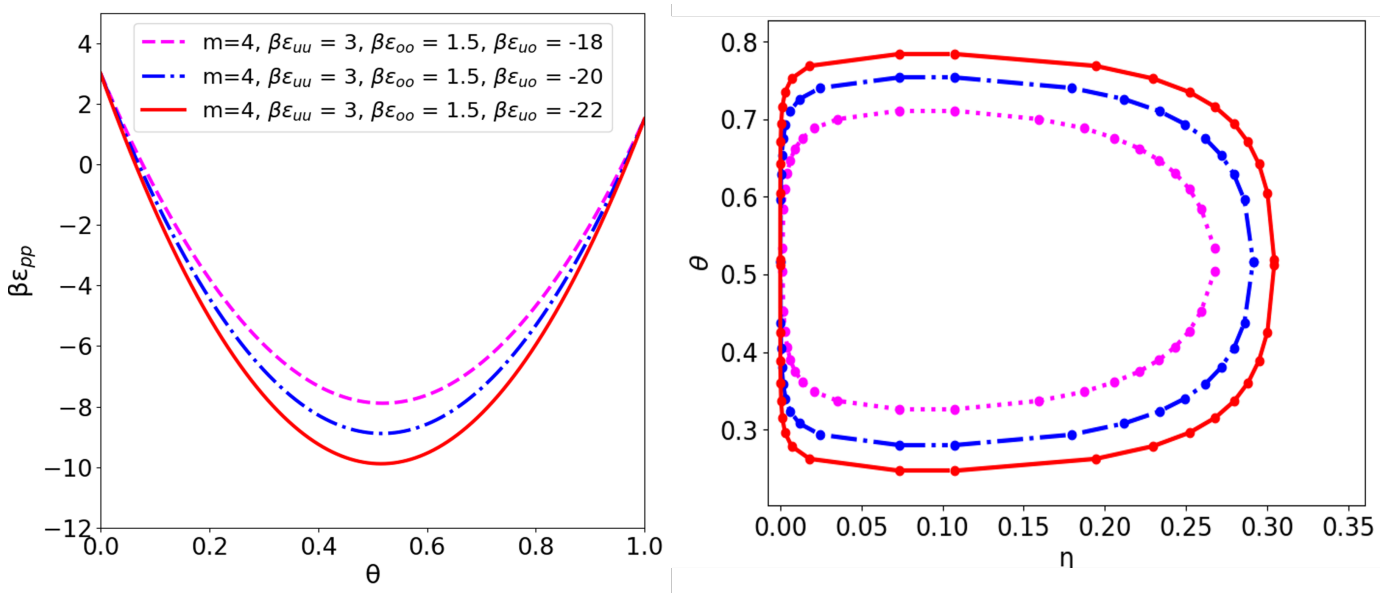

Figure 5.3: Interaction energies curves and LLPS loops for systems with the same repulsion parameters: Increasing the attractions increase the width of the interaction energy curves and therefore the size of the LLPS loop.

case, changing the distribution of the particles at the critical point and at any given probability Θ. In the next section we will discuss the effect of different thermodynamics parameters that are included in the Wertheim theory of the liquid-vapour equilibrium.

5.3.3 Effect of the thermodynamics parameters on the liquid-vapour equilibrium

Our model is based on the Wertheim theory of patchy particles that has a well defined set of parameters from which the thermodynamics quantities like the pressure and chemical potential follow. We want to explore the effect of changing these thermodynamics parameters employed within the Wertheim theory on the shape of the liquid-vapour equilibrium loop. The first parameter that we test is the number of patches m on each protein. It is important to notice that the number of patches does not affect the overall shape of the interaction energy equation [Eq. (5.3)]. The dependence of the number of patches can be seen by the probability $p(m,i)$ of the occupied patches Θ^i and of unoccupied patches $(1-\Theta)^{m-i}$, given in Eq. (5.2). In Fig. 5.4 we show how the effect of changing the number of patches m shifts the critical points for liquid-vapour

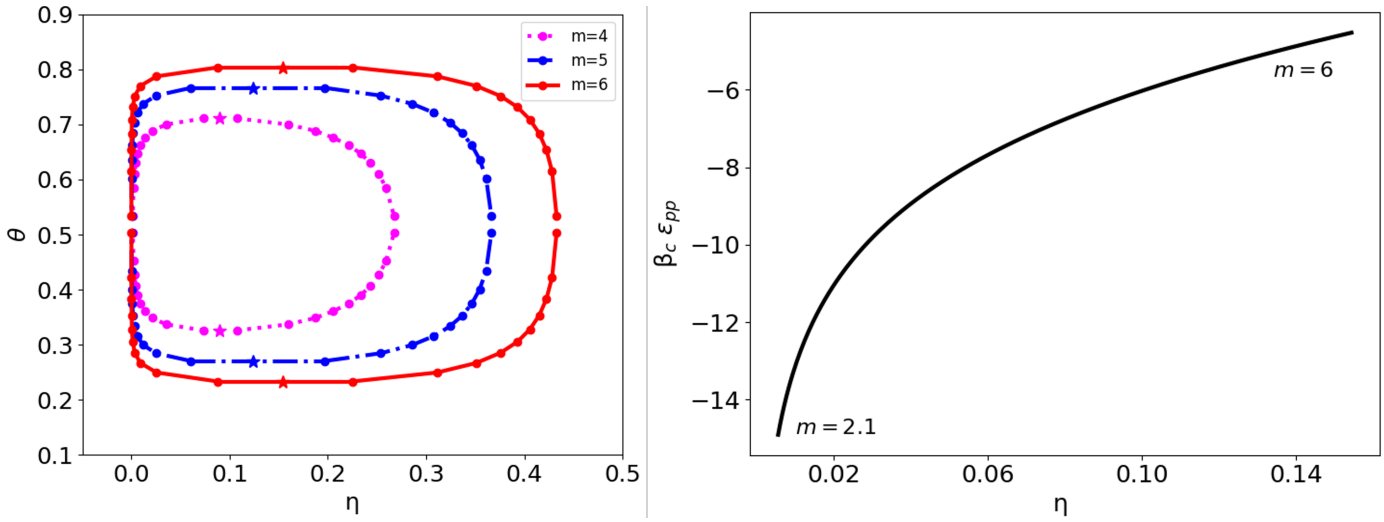

Figure 5.4: Effect of changing the number of patches on the patchy particles for interaction energy curves with $\beta\epsilon_{uu}$= 3, $\beta\epsilon_{oo}$= 1.5 and $\beta\epsilon_{uo}$= -18, increasing the number of patches increases the interaction energy necessary to reach liquid-liquid phase separation, it leads to an increased size of the coexistence loop and in a change of the critical value for the packing fraction.

phase separation. This is consistent with previous literature [9].

We want to emphasize that changing the number of patches does not only change the size of the coexistence loop but also the properties of the dense and diluted phase. In a system with only two patches, which does not phase separate [9], the possible geometry of clusters is limited to form linear chains that eventually might close into a ring. For a larger number of patches, different cluster morphology are possible, which can form in the dense phase as well as in the gel. The effect of changing the number of patches does not affect only the network of the dense phase but also the properties of the solid, the crystalline and the gel structures. A detailed treatment of such effects in patchy particle systems can be found in Refs. [100] and [38]. In particular, in Ref. [100], different type of crystalline phases are found for particles with valence of $m = 3$ and compared with particles with $m = 5$.

Another interesting effect is related to the other two parameters used in the Wertheim theory. The first one is the effect of the particle hard-sphere radius R and the second one is related to the volume of the interaction K. Changing the radius of the particle, the size of the loop changes as well as the value of the interaction energy at the critical point $\beta_c\epsilon_{pp}$, but not the critical packing fraction η_c, that remains constant. An example of those effects are included

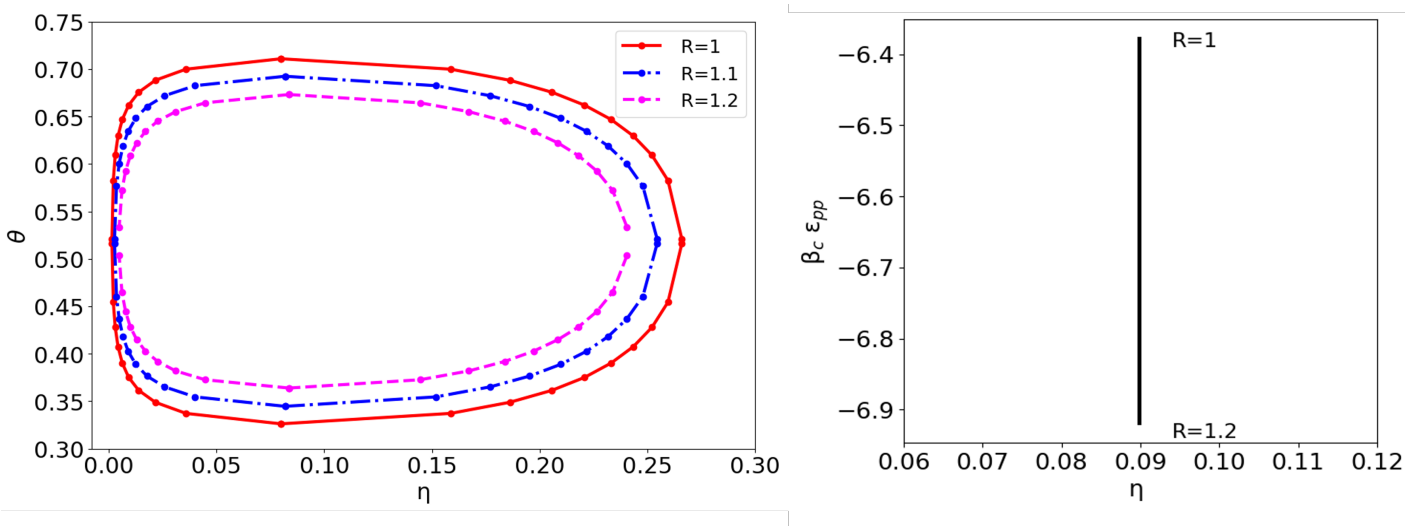

Figure 5.5: Effect of changing the radius of the particles for interaction energy curves with $\beta\epsilon_{uu}$= 3, $\beta\epsilon_{oo}$= 1.5 and $\beta\epsilon_{uo}$= -18, increasing the radius of the particles decrease the interaction energy necessary to reach liquid-liquid phase separation, it reduce the size of the LLPS loop.

in Fig. 5.5. This effect is due to the fact that the ratio between the interaction volume parameter K and the volume of the patchy particle increases as the radius decrease, resulting in stronger attractions. In other words the fraction of the surface that is covered by patches increases. The effect of the radius might explain several phenomena in experimental systems.

For example, in protein solution, the interactions between solvent and macromolecules can lead to a change of the hydrodynamic radius, depending on the protein-solvent interactions. We speculate, that within the framework of the Wertheim theory and ion activated attractive patches, it is possible to explain several effects arising from the protein-solvent interactions, without the need to explicitly treat the solvent, but indirectly infer the solvent contribution from the particle radius or from experimental systematic behaviour. In protein for example, the hydrodynamic radius is connected to protein-solvent interactions. A strongly polar solvent, for example, increases the hydrophobic effect and, depending on the amino-acid composition of the protein, might compact or relax the ternary structure. In Fig. 5.5 we show that small changes in the radius of the particle have a huge impact on the concentration of the liquid-vapour equilibrium. For example an increase of 0.2 units, increases the resulting packing fraction in the low density phase of about 3.3 times and decreases the packing fraction of the dense phase of about 12.5% at the extreme points of the coexistence

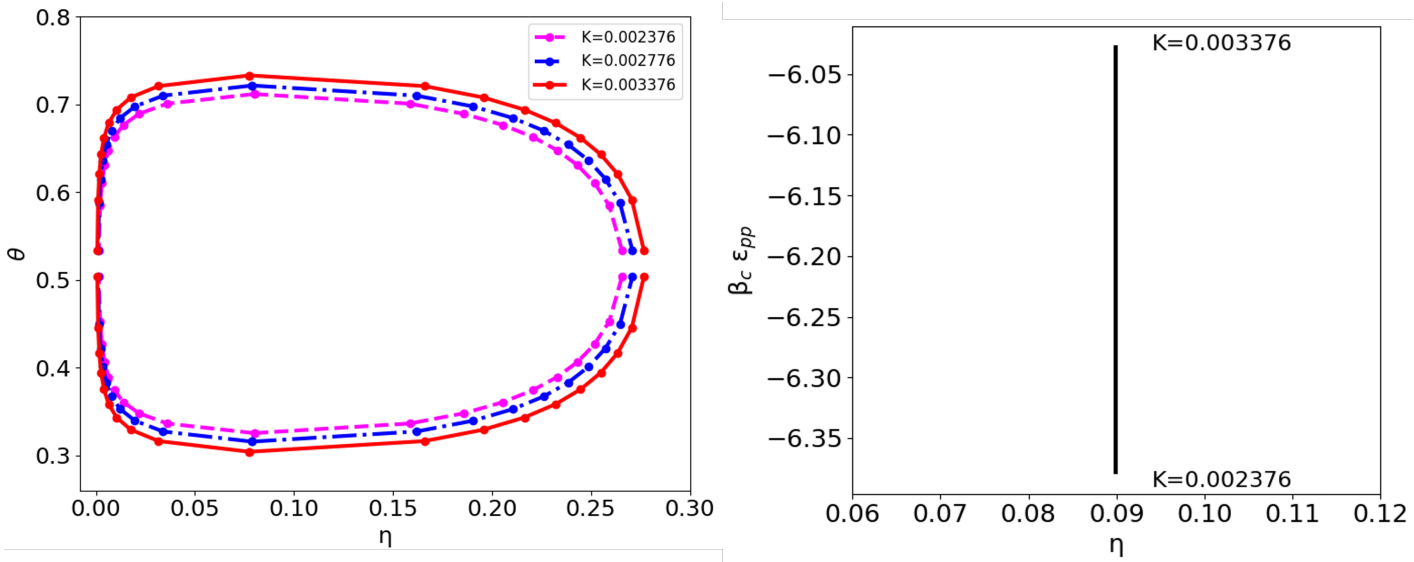

Figure 5.6: Effect of changing the Wertheim parameter K of the particles for interaction energy curves with $\beta\epsilon_{uu}$= 3, $\beta\epsilon_{oo}$= 1.5 and $\beta\epsilon_{uo}$= -18, increasing the Wertheim parameter of the particles increase the interaction energy necessary to reach liquid-liquid phase separation, it increase the size of the LLPS loop.

regions, while at central values the effects is less significant. We can also induce the same effect reported in Fig. 5.5 by fixing the radius of the protein and increasing the Wertheim parameter K. In this way we can tune the ratio between the interaction volume parameter and the volume of the patchy particle that is relevant to produce this effect.

5.4 Effect of the temperature

The temperature is a fundamental thermodynamic parameter which has been kept constant in the considerations so far. However, as we increase temperature in our system the attraction between proteins, induced by the formation of salt bridges between patches is weakened. If the temperature reaches a critical value T_c liquid-vapour phase separation vanishes and above T_c only the mixed fluid state is observed. Note that in this study we do not consider the solid phase. The full fluid phase diagram as function of temperature T, the protein packing fraction η and the binding probability Θ is shown in Fig. 5.7

However, the direct effect of temperature in experimental systems such as proteins can be far more complicated, involving not only a change of the interac-

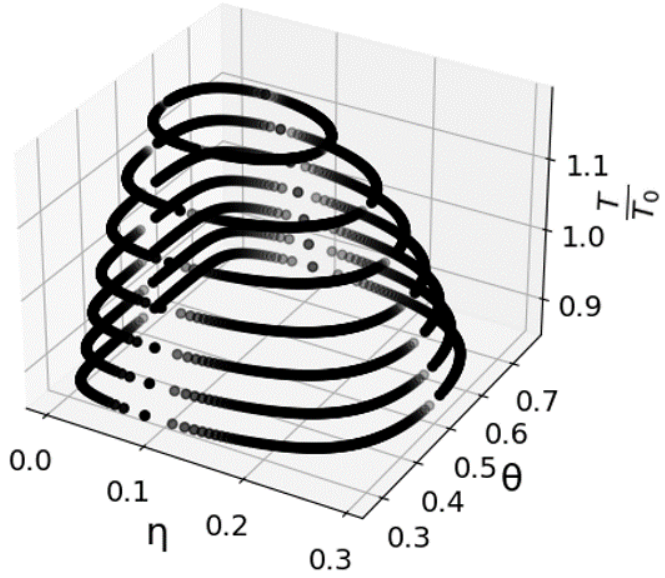

Figure 5.7: Effect of changing the temperature on the coexistence loop. The reference curve is at T = T_0 and refer to $\beta\epsilon_{uu}$= 3, $\beta\epsilon_{oo}$= 1.5 and $\beta\epsilon_{uo}$ = -18.

tions, but also strong conformations changes, that can make the spherical model here assumed not a good representation of the system. However, in the region of temperature where the globular structure of the protein is preserved, before denaturation, most of the effect due to the temperature are well reproduced within our model. It is worth noting that the temperature in the experimental system does not affect only the conformation of the proteins but also changes the ionic water equilibrium and the resulting hydration interactions between ions and water, protein and ions as well the specific pKa values of the aminoacidic residues [57]. Therefore, some intrinsic variables in the experiment are not investigated for the specific limitation due to the coarse grain approach.

5.5 Conclusion

In this work we have extended the ion activated patchy particles model for protein-salt mixtures, taking the interaction between binding sites of proteins in more details into account than before. While in previous studies the main focus was on the attraction between proteins induced by a salt bridge that forms between an occupied and an unoccupied binding site, here we have also taken the repulsion between two unoccupied and between two occupied

binding sites into account. The effect of this enriched interactions in our model produces a different and asymmetric ensemble of particles, consisting of proteins with no bound ion, proteins with one bound ions, etc. up to proteins with all binding sites filled. This rich multicomponent system, can be used to understand different phenomena, for example it can be use to rationalise different kind of phase transitions in colloidal system, such as liquid-liquid or liquid-vapour phase separation.

We have shown that although the model of ion activated patchy particle model based on the Wertheim theory is rather simple, the parameters have important effects on the liquid-vapour phase separation loop. Our main results are:

- The effect of the repulsion between two occupied and between two unoccupied binding sites on proteins influences the height of the liquid-vapour loop by shifting the interaction energy curves and resulting in different patchy particles distribution as described by the histograms.

- Increasing the attraction between an occupied and an unoccupied site increase the size of the coexistence loop as well as change the ensemble distribution at the critical point.

- The only sensitive parameter to change the critical packing fraction in the coexistence loop is the number m of patches on the surface of the particles. Increasing the number of patches lead to a bigger size of the loop and shift the critical interaction energy for phase separation to higher values. Increasing the number of patches also increase the number of possible components in the histograms.

- Increasing the radius or the patchy particle (protein) changes the ratio between the interaction volume parameter K and the volume of the particle, decreasing the interaction energy necessary to reach the critical point and reduce the size of the phase-separation loop.

- Changing the Wertheim parameter K produces the same effect of changing the radius of the particles, but in the opposite direction. An increasing of the Wertheim parameter leads to an increased size of the loop due to the increased ratio between the volume of the interaction and the volume of the particle, this effect is visible in Fig. 5.6.

The aim of our model is to describe the phase behaviour of proteins in salt solutions. In the ion activated patchy particle model the probability Θ to have an occupied patch on the surface of the protein is given by a Fermi-like distribution in the grand canonical ensemble, Eq. (5.1), that is a function of the salt concentration in the reservoir, or equivalently its chemical potential μ_s. The reservoir concentration of salt ions, c_s^r, is a quantity within the theoretical framework. However, the total salt concentration in the system, which is the quantity that can be controlled or measured in experiments, is directly connected with the salt concentration in the reservoir through [103]

$$c_s = m\Theta\rho + c_s^r(\mu_s)(1 - \eta(1 + R_s/R)^3). \tag{5.10}$$

The first term takes into account the ions bond on the surface of the patchy particle and the second term originates from the free ions in the solution, corrected for the volume excluded by the proteins. Using this relation, it is possible to access the concentration of ions on the surface of the patchy particles to obtain LLPS loop that can be compared to the experimental one, since they are function of the salt concentration that it is also our experimental variable.

It will be interesting to compare predictions of our enriched model to known experimental results. Furthermore our model can be further enriched by allowing for binding sites with different binding energies. This will increase the complexity of the model and should be an important step towards understanding complex systems such as proteins in salt solutions.

5.6 Outlook

In this article, we extended the patchy particles model including the effect of the repulsions. The first model from Felix Roosen-Runge et al.[103] assume an interaction energy between patchy colloids neglecting the repulsion between unoccupied patches as well as the repulsion between occupied patches on the surface of the particles. The general expression for the pair interaction of this kind is given by the following expression:

$$\beta\epsilon_{pp} = 2\beta\epsilon_{uo}\theta(1 - \theta) \tag{5.11}$$

In our more recent development, our total interaction energy contains two additional terms that takes into account the repulsion between unoccupied-unoccupied and occupied-occupied patches on the particle's surface. A more complete picture of the for the pair interaction is the following:

$$\beta\epsilon_{pp} = \beta\epsilon_{uu}(1-\theta)^2 + 2\beta\epsilon_{uo}\theta(1-\theta) + \beta\epsilon_{oo}\theta^2 \tag{5.12}$$

So far, even using this approach, we considered the binding probability of the patches on the particles the same. However, in real world scenario the binding energy of the patches can be different. We developed a new expression in this case. The resulting expression that takes into account two different kinds of patches is derived from the classical probability of distribute i ions on m patches. The resulting expression is the following:

$$\begin{aligned}
\epsilon_{pp} = & \frac{1}{N_A N_B}\epsilon_{uu}\left[m_\alpha(1-\theta_\alpha) + m_\beta(1-\theta_\beta)\right]^2 \\
& + \frac{1}{N_A N_B}\epsilon_{oo}(m_\alpha\theta_\alpha + m_\beta\theta_\beta)^2 \\
& + \frac{2m_\alpha^A m_\alpha^B}{N_A N_B}\epsilon_{\alpha\alpha}\theta_\alpha(1-\theta_\alpha) \\
& + \frac{2m_\beta^A m_\beta^B}{N_A N_B}\epsilon_{\beta\beta}\theta_\beta(1-\theta_\beta) \\
& + \frac{2m_\alpha^A m_\beta^B}{N_A N_B}\epsilon_{\alpha\beta}[(1-\theta_\alpha) + (1-\theta_\beta) - 2(1-\theta_\alpha)(1-\theta_\beta)]
\end{aligned} \tag{5.13}$$

Where N_A and N_B are the number of patches on the surface of particles A and B, θ_α is the probability to occupy a patch of kind α, θ_β is the probability to occupy a patch of kind β, m_α is the total number of patches of kind α on a particle, m_β is the total number of patches of kind β on a particle, m_α^A is the number of patches of kind α on the surface of particle A, m_β^A is the number of patches of kind β on the surface of particle A, m_α^B is the number of patches of kind α on the surface of particle B, m_β^B is the number of patches of kind β on the surface of particle B. The interaction term between an unoccupied and an occupied patches, now consists of three different terms, the interaction energy $\epsilon_{\alpha\alpha}$ between occupied and unoccupied patches of kind α, the interaction energy $\epsilon_{\beta\beta}$ between occupied and unoccupied patches of kind β and the cross term $\epsilon_{\alpha\beta}$ that takes into account the interactions between patches of different kind. The relation here derived,

allow us to explore different kinds of systems. For systems with three patches, for example, we can have different configurations that can drive the system to display a different phase diagram. For example, in a three patches system we can have 2 patches of type α and 1 of type β or the other way round. In this approach we consider a system of identical patchy particles, but in the future will be possible to extend this relation even for interactions between mixtures of patchy particles with different number of attractive sites or of different kind. For example in the case of a system of patchy particles with two binding sites (m=2) we could have scenarios where the particle A has 2 binding sites of the same type and particle B, two binding sites of different kind and or combination of them. Regarding our more recent development, Eq. 5.13, it is already possible to display some features of the model plotting the interaction energy curves as a function of the two different binding probabilities as shown in Fig. 6.1. With this model would be possible to investigate the phase diagram that will be now function of the two different binding probabilities. This enrichment of the model can be used to investigate and understand different experimental conditions and phase behaviours, such as differences in the crystal lattice and morphology as well as differences in the LLPS borders and coexistence regions. As a further improved would be interesting to connect the binding energy with the different type of ions that can bind the surface of the particles, for example, including specific quantum mechanics effects. In addition the theoretical framework can be used as a base to investigate the features of the model via Montecarlo and or molecular dynamics simulations.

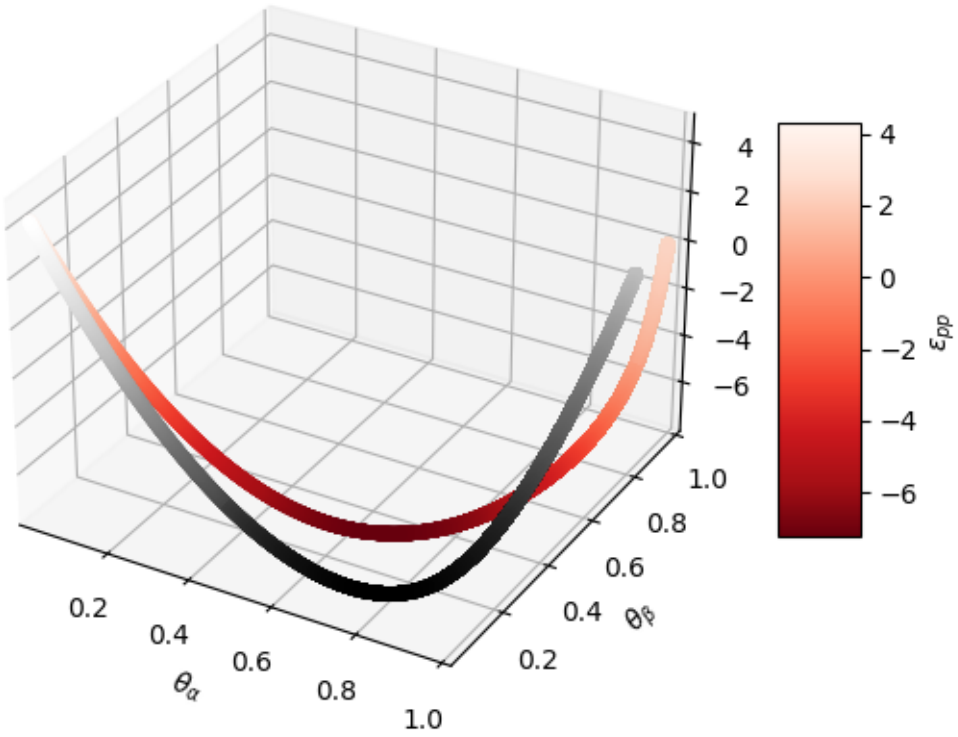

Figure 5.8: Two different energy curves for different set of attractive interaction parameters, it is possible now to solve the condition of equilibrium to get a 3-D phase diagram for the Liquid-Vapour coexistence

Chapter 6

Contribution on Kinetics of HSA crystallization and its relationship with the phase diagram

The investigation and rationalization of crystallization data can be challenging, in the paper of Buchholz et al. we employed a kinetic model based on the Nanev sigmoidal model of nucleation rate.[84] The model assumes that the variation of the number of nuclei (N) during time is given by two contributions, the first is a term that expresses the rate on which the nucleii are formed, the second contribution states for the rate of nuclei consumption. The overall expression can be written as follow:

$$\frac{dN}{dt} = k\,N - k\frac{N^2}{N_s} = \; k\,N(1 - \frac{N}{N_s}) \tag{6.1}$$

where N_s is the number of nuclei at saturation. The previous equation is a first order differential equation that can be analytically solved obtaining a valid expression to fit the experimental data from the nucleation density plot. The nucleation density plot, is obtained counting the number of crystals in the glass slides using an optical microscope at different times. The overall counting, in the case of HSA in presence of $CeCl_3$ can take up to 10 days, in which the number of visible crystals are counted at regular intervals of time. The procedure is repeated at least three times in order to have a reliable statistics. After solving the first order differential equation the final expression for the rate of nuclei formation is given by:

$$N(t) = \frac{N_s}{1 + exp(-k(t - t_c))} \tag{6.2}$$

From which values as the characteristic rate k and the critical time for nucleation t_c, when half of the nuclei are formed can be obtained. Some characteristic parameters of the model are listed below:

$$\text{Maximum nucleation rate} \qquad \frac{dN^2}{dt^2} = 0 \rightarrow \frac{N}{N_s} = 0.5 \rightarrow \frac{dN}{dt_{max}} = \frac{kN_s}{4} \tag{6.3}$$

$$\text{Tangent at the inflection point} \qquad N = \frac{N_s}{2} - \frac{k\,N_s\,t_c}{4} + \frac{k\,N_s\,t}{4} \tag{6.4}$$

The tangent at the inflection point can be used to calculate t_0 that is related to the induction time τ for nucleation. Therefore at $N = 0$, $t = t_0$ we obtain the following expression:

$$t_0 = t_c - \frac{2}{k} \qquad \rightarrow \qquad \tau \approx 0.6\,t_0 \tag{6.5}$$

The induction time τ can be connected with the supersaturation ratio, in fact, $\tau \propto \ln(c/c_{eq})^2$, from the CNT [84]. In Tab. 6.1 are shown the parameters obtained with the fitting procedure.

Protein $\left(\frac{mg}{ml}\right)$	Salt (mM)	$k(1/h)$	$k \times 10^{-5}(1/s)$
20	2	0.09583	2.6619
20	3	0.0715	1.9861
20	4	0.03608	1.0022
35	2.5	0.05721	1.5892
35	3	0.1187	3.2972
35	4	0.08513	2.3647
35	5	0.03069	0.85250
50	3.5	0.2563	7.1194
50	4	0.4282	**11.89**
50	5	0.2562	7.1167
50	6	0.1748	4.8556
80	6	0.1116	3.1000
80	8	0.1054	2.9278
100	7	0.1852	5.1444

Table 6.1: Values of the characteristic rate constants obtained through the fitting of the nucleation density plot for different protein concentrations.

Our approach, even if based on the Nanev model, slightly differ in the experimental procedure. In our experiments, we prepare the protein-salt solution at

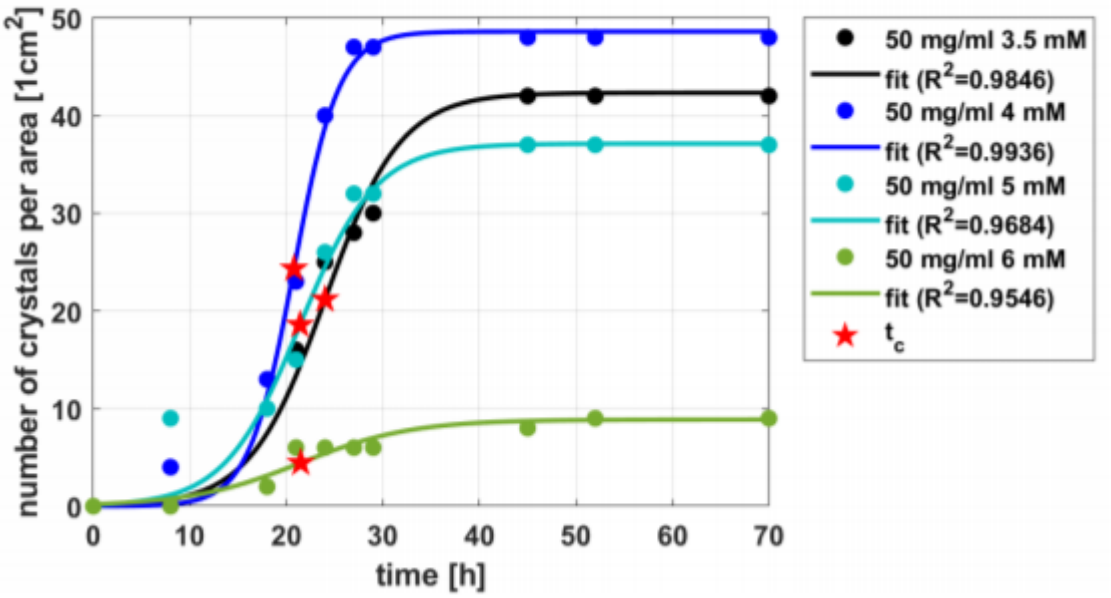

Figure 6.1: Fitting procedure of the nucleation density plot for different salt concentration at constant protein concentration from Ref. [12]. The resulting characteristic rates for this experimental conditions are given in Tab 6.1.

low saturation conditions, in the vicinity of c^*. At this conditions, the eventual drop in concentration due to the nuclei formation ensure that the system is investigated under conditions where only the already formed nuclei are free to growth in macroscopic crystals. In the Nanev approach, this is achieved changing rapidly the temperature from lower values to higher values with the purpose of limiting the growth of the first nucleii, obtaining a reliable statistic on the nucleation process. The values of the kinetics constants here obtained, are of the same order of magnitude of the rate constants present in literature and calculated with the same sigmoidal model, suggesting that our apporach can be used to model the kinetic of systems, containing negatively charged proteins in presence of trivalent salts.

6.1 Outlook

The present model, can be further extended to describe and discriminate between classical nucleation and non classical nucleation. In the non classical nucleation, the system pass through an intermediate step before nucleation. By

rewriting Eq. 6.2 it is possible to separate the overall equation in two different contributions:

$$\frac{N_s}{1 + a\ \exp(-kt)} = \frac{a\, N_s}{a + \exp(kt)} + \frac{N_s \exp(kt) - a\, N_s}{a + \exp(kt)} \tag{6.6}$$

The left hand side of Eq. 6.6 refers to the two step nucleation, while the right term contains the information about the homogeneous process. Where a is a constant that depends on the initial conditions of the system. This number can change depending on impurities and or crystallization conditions. In the Nanev approach a is chosen in such a way that the resulting equation can be written as Eq. 6.2 as a function of t_c, this equation contain implicitly also the two step contribution derived by Barlow et al. in Eq. 6.6 [7]. In order to test the validity of this two-step kinetics model, would be interesting in the future, to investigate different crystallization scenario to apply the model to non classical crystallization pathways. An intriguing opportunity to test this model, might involved crystallization condition close the LLPS border for solution of proteins with trivalent salts, since in this regions the formation of a dense phase can affect the overall crystallization pathway increasing the likelihood of having a non-classical crystallization process.

Part V

Conclusive remarks

Chapter 7

Overall Conclusion

7.1 Discussion and Outlook

Protein solutions in the presence of multivalent salts remain an intriguing field of research. In this work we combined an experimental approach on the investigation of complex system and a theoretical one, making use of several models to rationalize the kinetic and thermodynamics behaviour of these mixtures. Chronologically, the first part of the research was done in the laboratories, with focus on crystallization. During the investigation of the crystallization of HSA in presence of $CeCl_3$ we make use of the logistic model for the calculation of the nucleation time and the rate. The model is a valuable tool for the investigation of the classical and non classical crystallization, as described in this work, in the resulting publication and references. However, in order to fully make use of the model, would be interesting to extend the investigation close to the LLPS regions, from which is possible to separate the sigmoidal equation in the two contribution obtaining additional information on multi steps crystallization processes. Since the importance of the supersaturation ratio in the field of crystallization, we performed calculations of this quantities employing UV-vis measurements from which was possible to access the osmotic second virial coefficient, employing the Mcmillan-Mayer theory of solution and classical thermodynamics relations. The results obtained were successfully compared with the results obtained from the fitting of SAXS curves employing a sticky hard sphere potential. Regarding this results would be interesting to further extend the theory to different system in order to improve the estimation of this quantity and connect it with different kinds of pair potentials. The last part of the work, involved the extension of the ion activated patchy particle model including the repulsion between occupied binding sites and between unoccupied binding sites. These parameters have an

impact on the overall position of the coexistence loop between the liquid and vapor using the thermodynamics of the Wertheim perturbation theory. The ion activated patchy particle model was further extended deriving an expression for the pair interaction between patchy particles that posses two binding sites with different affinity for the ions. The next steps will be to fully characterize this system using both theory and simulation in order to connect the results with the experimental behaviour observed in protein salt solutions. Overall protein salt solutions were successfully investigated using different approaches, since the complexity of this systems does not allow to unequivocally characterize their behaviour using a single theoretical approach.

Appendix

7.2 Additional experiments and supersaturation measurements

The approach used in Chapter IV, can be further extended to different protein systems. Another system investigated during the publication of the manuscript involved the study of the partitioning between β-lactoglobulin ($\beta - lg$) in $LaCl_3$ (Lanthanum Chloride). From which the following table is derived.

βlg			
c_p[mg/ml]	c_s[mM]	c_s/c_p	$B_2'(supersaturation)$
5	0.35	2.5	-4.08
10	0.7	2.5	-4.23
20	1.4	2.5	-4.14
30	2.1	2.5	-3.96
40	2.8	2.5	-4.48
5	0.8	5.5	-17.01
10	1.5	5.5	-17.89
20	3.1	5.5	-17.58
30	4.6	5.5	-14.84
40	6.2	5.5	-14.82

Table 7.1: Normalized second virial coefficient values for βlg with $LaCl_3$ obtained via supersaturation obtained using Eq. 4.11. The estimated error on the second virial coefficient is about 10%. Values of c_s/c_p are approximated by an integer number. The conditions of c_s/c_p at ratio 2.5 is referred to the c^* border of the phase diagram of βlg with $LaCl_3$, while at ratio 5.5 the system is in the middle of Regime II. The value of the supersaturation ratio changes from about 1.25-1.5 near c*, to 40-130 in the middle of the Regime II.

The value of the second virial coefficient obtained by supersaturation near the c^* border is as order of magnitude, not different from the values obtained with HSA in presence of $CeCl_3$, and the supersaturation ratio remain quite small over

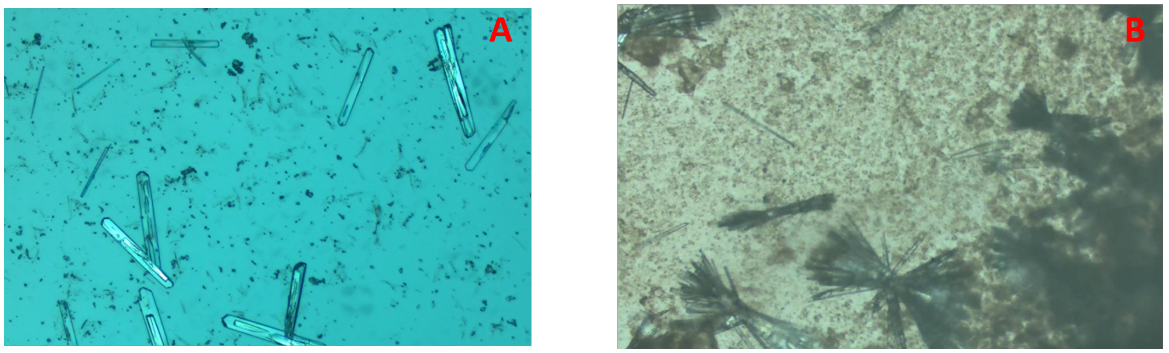

Figure 7.1: Different equilibrium shape of the crystals a t fixed protein concentration but a different position in the phase diagram: A) Condition close c* at mild supersaturation and B) condition at high supersaturation in middle of the Regime II.

all the interval of protein concentration explored. At this condition the crystals nucleate from a clear solution without the presence of visible aggregates. At this conditions would be possible to compare the results here obtained with the results from scattering experiments using static or dynamic light scattering (S/DLS) and SAXS, since the sample is still partially homogeneous. The first point where LLPS usually appear in the glass slides, is around $\tilde{c} \approx 20 mg/ml$ that is the normalization factor used in the Eq. 4.11 to compute the virial coefficient.

At low supersaturation, the equilibrium shape of the crystal appear different and more regular with to respect the crystals appearing from the solution prepared in the middle of the Regime II as shown in Fig. 7.1. This differences can be rationalize on the basis of the supersaturation ratio or on the second virial coefficient; when the driving force of nucleation is too high the kinetic aggregation compete with the formation of the crystals resulting in a not optimal binding between proteins. Alternatively, it is possible to connect the resulting observation, with the George and Wilson's crystallization slot. At high negative second virial coefficient the system is too far away from the ideal crystallization conditions as the system is too far from the crystallization slot [37]. It is worth noticing, that in the middle of the regime two, it is impossible to calculate the second virial coefficient with scattering methods, especially with SAXS fitting procedures, because the solution contain too many macroscopic aggregates and the concentration in the low density phase (supernate) is too low to detect even the protein form factor. Therefore, the estimation used in this work that rely only on solubility measurements allow us to investigate conditions that are not accessible to the standard scattering techniques. In addition, this difference in

the second virial coefficent between proteins can be rationalize on the basis of different degrees of patchiness, since for patchy particles systems, depending of different parameters, such as the number of patches and the extension of them on surface of the particles, the resulting values of the critical second virial coefficient, can be much more negative that the one obtainable from isotropic square well fluids as modelled from the SAXS curves. [34] In addition, the values of the second virial coefficents here obtained are in agreement with the one obtained by R. Piazza et al, in which they obtained the B_2 values using SLS and DLS for $\beta lg - A$ in presence of monovalent salt and in absence of macroscopic aggregates [96]. In order to explore the capability of this thermodynamic calculation would be interesting to analyze different protein-salt systems and further improve the calculation and the general understanding of such results.

7.3 Additional data on the ion activated patchy particles model

Additional information on the patchy particles model can be obtained investigating the effect of different parameters on the relative orientation of the binodal-spinodal loops. It is possible to speculate that changes in the repulsion can lead to different forms of the binoidal and spinodal points. The relative points can be obtained by solving the well know condition:

$$\left(\frac{\partial P}{\partial \rho}\right)_T = 0 \tag{7.1}$$

By Eq.5.10 it is possible to represent the curves with to respect the salt concen-

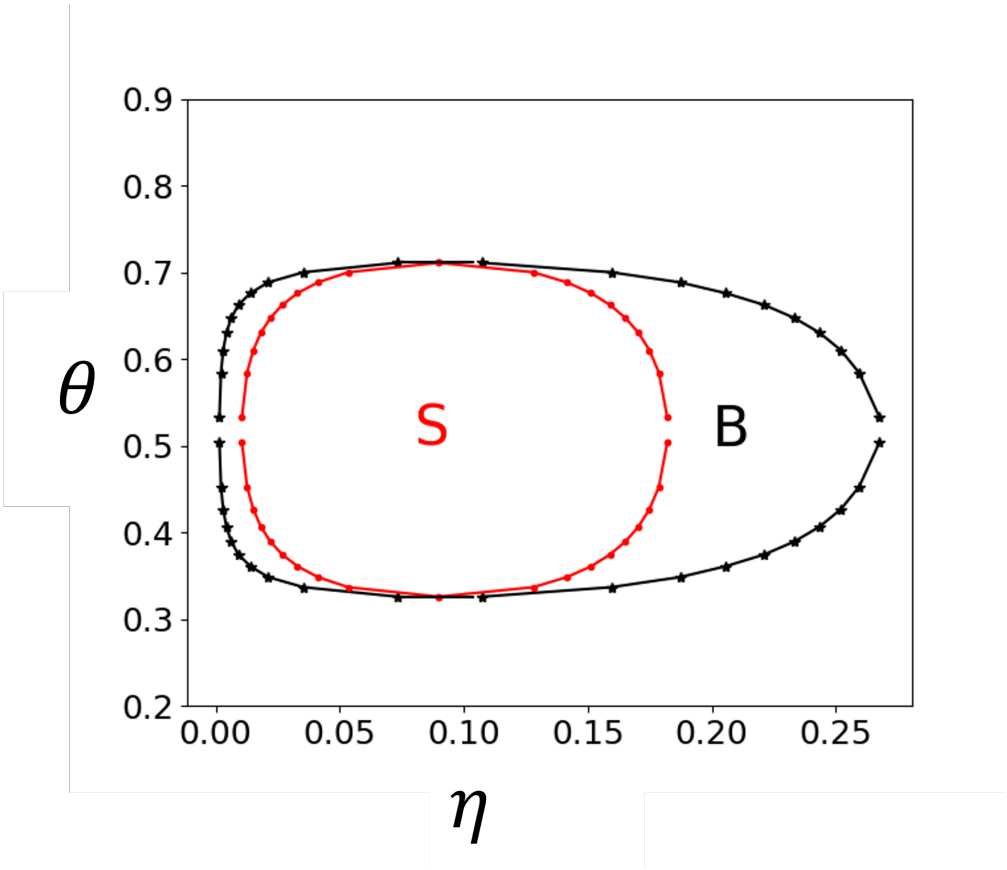

Figure 7.2: Binoidal and Spinoidal curves for value of parameters $\beta\epsilon_{uu}$= 3, $\beta\epsilon_{oo}$= 1.5 and $\beta\epsilon_{uo}$= -18.

tration of the system that is the variable used in the experiment that allowed a direct comparison of the theory with the experiments. In Fig.7.3 is displayed

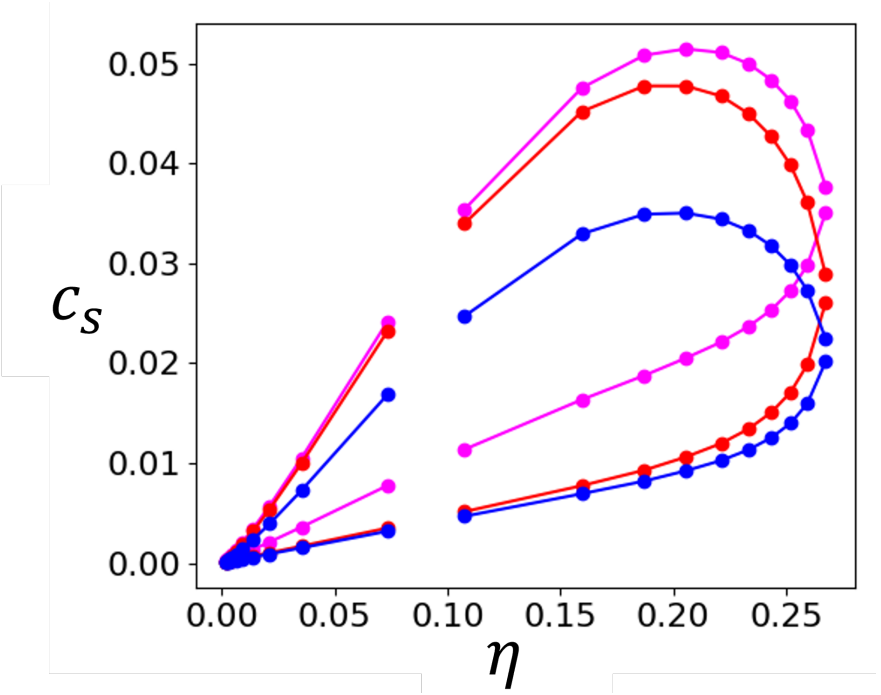

Figure 7.3: Liquid-vapour coexisting curve on the c_s-η plane generated from systems with different interaction energies, calculated from the curves in Fig.5.1.

the effect of the repulsions on the overall shape of the LLPS loop. The presence of the repulsions decrease the area of the loops and modulate the slope and the height of the LLPS coexistence regions. From the pressure plotted against the packing fraction is possible to better visualize the effect of the number of patches on the critical energy necessary to bring the system to the LLPS regions as shown in Fig. 7.4, 7.5 and 7.6.

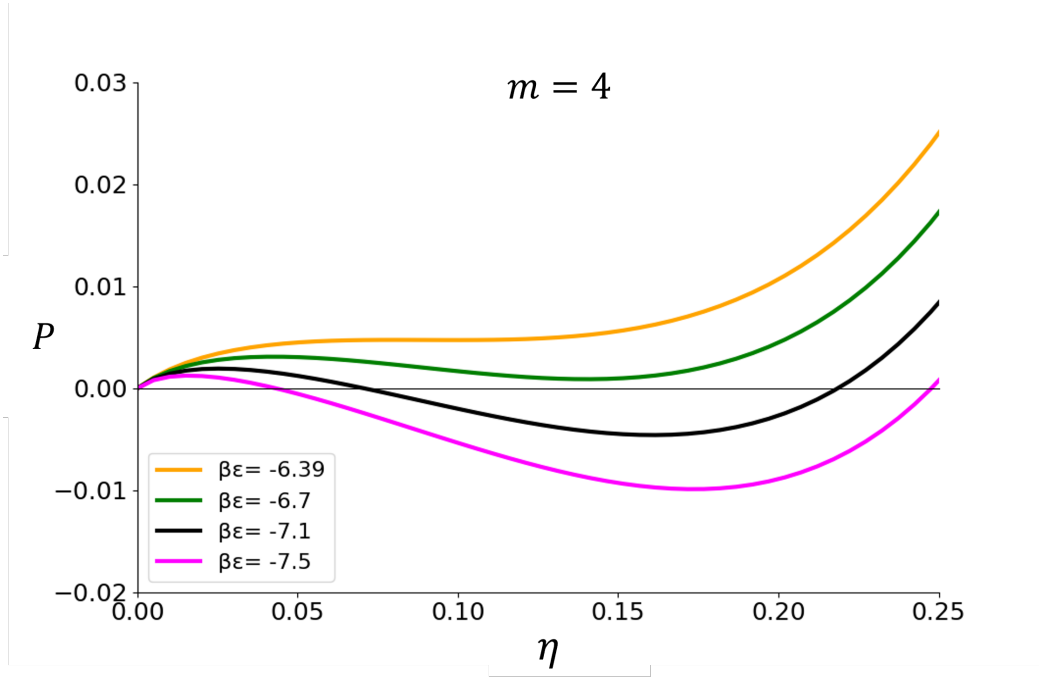

Figure 7.4: Pressure from the Wertheim equation of state, value on the vicinity of the critical point for a fixed number of patches (m=4) $\beta\epsilon$ in the plot refers to $\beta\epsilon_{pp}$

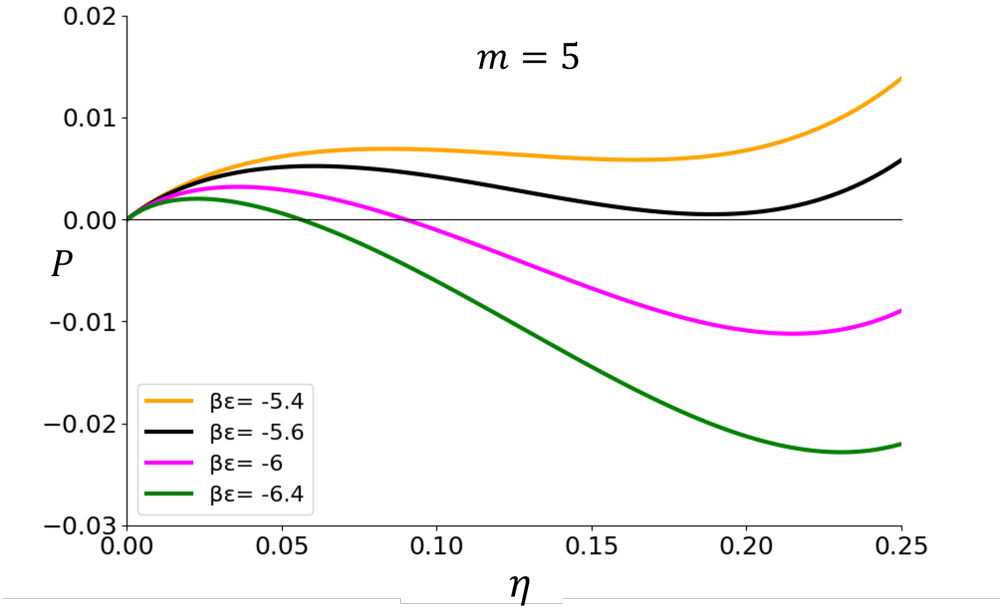

Figure 7.5: Pressure from the Wertheim equation of state, value on the vicinity of the critical point for a fixed number of patches (m=5) $\beta\epsilon$ in the plot refers to $\beta\epsilon_{pp}$.

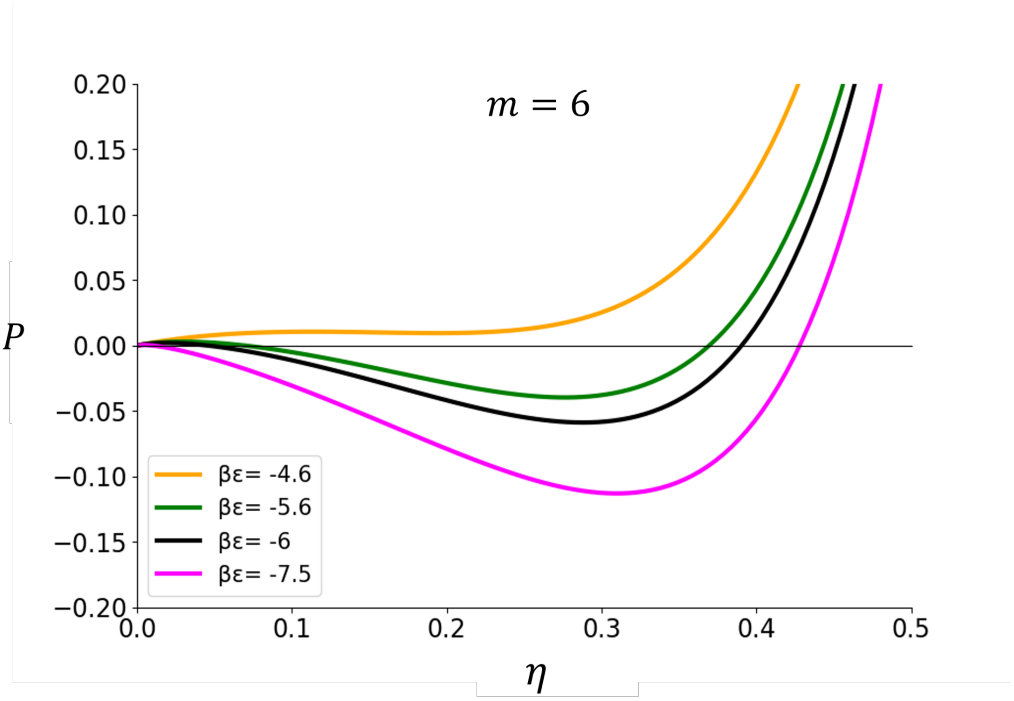

Figure 7.6: Pressure from the Wertheim equation of state, value on the vicinity of the critical point for a fixed number of patches (m=6) $\beta\epsilon$ in the plot refers to $\beta\epsilon_{pp}$.

List of publications

Kinetics of HSA crystallization and its relationship with the phase diagram

Cara Buchholz, Lara F Reichart, **Furio Surfaro**, Ralph Maier, Fajun Zhang, Alexander Gerlach, and Frank Schreiber, Journal of Crystal Growth 603 (2023), 126959.

An alternative approach to the osmotic second virial coefficient of protein solutions and its application to liquid–liquid phase separation

Furio Surfaro, Ralph Maier, Kai-Florian Pastryk, Fajun Zhang, Frank Schreiber, and Roland Roth, The Journal of Chemical Physics 158 (2023), no. 16

From Adsorption to Crystallization of Proteins: Evidence for Interface-Assisted Nucleation

Hadra Banks, **Furio Surfaro**, Kai-Florian Pastryk, Cara Buchholz, Ivan A Zaluzhnyy, Alexander Gerlach and Frank Schreiber, Colloids and Surfaces B: Biointerfaces (**under review**)

The Ion Activated Attractive Patchy Particle Model and Its Application to the Liquid-Vapour Phase Transitions

Furio Surfaro, Fajun Zhang, Frank Schreiber and Roland Roth
(**in preparation**)

Acknowledgments

I am deeply grateful to Prof. Dr. Dr. h.c. Frank Schreiber for providing me with the opportunity to conduct my PhD research within his group. His support and encouragement have been invaluable throughout this journey.

I extend my heartfelt appreciation to Prof. Dr. Roland Roth for his exceptional mentorship during this time. His dedication to my scientific and personal development has been instrumental in shaping my research path.

I would like to express my sincere gratitude to PD. Dr. Fajun Zhang for his insightful suggestions, scientific advice, and the meaningful time we have shared together collaborating on various projects.

I am indebted to my colleagues for their camaraderie, support, and the enriching experiences we have shared over the years.

To my family, with a special mention to my parents, I offer my deepest thanks for their unwavering support and understanding throughout this endeavor. Their love and encouragement have been my source of strength.

Special appreciation goes to Hadra Banks for her unwavering belief in my dreams and for always standing by my side. I am profoundly grateful for her love and companionship.

In loving memory of my grandmother Gabriella Moretti, whose encouragement and wisdom continue to inspire me.

To all the kind members of my family and friends that I loved and are not longer with us.

Thank you all for being part of this incredible journey.

Lo maggior corno de la fiamma antica
cominciò a crollarsi mormorando,
pur come quella cui vento affatica:
indi la cima qua e là menando,
come fosse la lingua che parlasse,
gittò voce di fuori e disse:
"Quando mi diparti' da Circe, che sottrasse
me più d'un anno là presso a Gaeta,
prima che sì Enëa la nomasse,
né dolcezza di figlio, né la pieta
del vecchio padre, né 'l debito amore
lo qual dovea Penelopè far lieta,
vincer potero dentro a me l'ardore
ch'i' ebbi a divenir del mondo esperto
e de li vizi umani e del valore;
ma misi me per l'alto mare aperto
sol con un legno e con quella compagna
picciola da la qual non fui diserto.
L'un lito e l'altro vidi infin la Spagna,
fin nel Morrocco, e l'isola d'i Sardi,
e l'altre che quel mare intorno bagna.
Io e' compagni eravam vecchi e tardi
quando venimmo a quella foce stretta
dov'Ercule segnò li suoi riguardi
acciò che l'uom più oltre non si metta:
da la man destra mi lasciai Sibilia,
da l'altra già m'avea lasciata Setta.

"O frati," dissi, "che per cento milia
perigli siete giunti a l'occidente,
a questa tanto picciola vigilia
d'i nostri sensi ch'è del rimanente
non vogliate negar l'esperïenza,
di retro al sol, del mondo sanza gente.
Considerate la vostra semenza:

fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza."

Dante Alighieri,

Inferno, Canto XXVI, vv. 85-120

Bibliography

[1] Alastair Aitken and Michèle Learmonth, *Protein determination by uv absorption*, The Protein Protocols Handbook (1996), 3.

[2] Markus Sällman Almén, Karl JV Nordström, Robert Fredriksson, and Helgi B Schiöth, *Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin*, BMC biology **7** (2009), 1–14.

[3] Andrea Antosova, Miroslav Gancar, Zuzana Bednarikova, Jozef Marek, Eva Bystrenova, and Zuzana Gazova, *The influence of cations on α-lactalbumin amyloid aggregation*, JBIC Journal of Biological Inorganic Chemistry **27** (2022), no. 7, 679.

[4] Sho Asakura and Fumio Oosawa, *On interaction between two bodies immersed in a solution of macromolecules*, The Journal of Chemical Physics **22** (1954), no. 7, 1255.

[5] K Baler, Osvaldo Antonio Martín, Marcelo A Carignano, GA Ameer, Jorge Alberto Vila, and Igal Szleifer, *Electrostatic unfolding and interactions of albumin driven by ph changes: a molecular dynamics study*, The journal of Physical Chemistry B **118** (2014), no. 4, 921.

[6] C Bradford Barber, David P Dobkin, and Hannu Huhdanpaa, *The quickhull algorithm for convex hulls*, ACM Transactions on Mathematical Software **22** (1996), no. 4, 469.

[7] DA Barlow and Jan Gregus, *The kinetics of homogeneous and two-step nucleation during protein crystal growth from solution*, International Journal of Chemical Kinetics **51** (2019), no. 11, 840.

[8] RJ Baxter, *Percus–yevick equation for hard spheres with surface adhesion*, Journal of Chemical Physics **49** (1968), no. 6, 2770.

[9] Emanuela Bianchi, Julio Largo, Piero Tartaglia, Emanuela Zaccarelli, and Francesco Sciortino, *Phase diagram of patchy colloids: Towards empty liquids*, Physical Review Letters **97** (2006), no. 16, 168301.

[10] Sophie Bourriot, Catherine Garnier, and Jean-Louis Doublier, *Phase separation, rheology and structure of micellar casein-galactomannan mixtures*, International Dairy Journal **9** (1999), no. 3-6, 353.

[11] William Pingry Boynton, *Gibbs thermodynamical model for a substance following van der waals' equation*, Physical Review (Series I) **11** (1900), no. 5, 291.

[12] Cara Buchholz, Lara F Reichart, Furio Surfaro, Ralph Maier, Fajun Zhang, Alexander Gerlach, and Frank Schreiber, *Kinetics of hsa crystallization and its relationship with the phase diagram*, Journal of Crystal Growth **603** (2023), 126959.

[13] Martin Caffrey, *Membrane protein crystallization*, Journal of Structural Biology **142** (2003), no. 1, 108.

[14] Norman F. Carnahan and Kenneth E. Starling, *Equation of state for nonattracting rigid spheres*, The Journal of Chemical Physics **51** (1969), no. 2, 635.

[15] Cyrus Chothia, *Principles that determine the structure of proteins*, Annual Review of Biochemistry **53** (1984), no. 1, 537.

[16] Kim D. Collins, *Charge density-dependent strength of hydration and biological structure*, Biophysical Journal **72** (1997), no. 1, 65.

[17] Kim D. Collins, *The behavior of ions in water is controlled by their water affinity*, Quarterly Reviews of Biophysics **52** (2019), e11.

[18] Angus F Core, *Xxvi. the second virial coefficient of gases*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **46** (1923), no. 272, 256.

[19] RA Curtis, J Ulrich, A Montaser, JM Prausnitz, and HW Blanch, *Protein–protein interactions in concentrated electrolyte solutions*, Biotechnology and Bioengineering **79** (2002), no. 4, 367.

[20] Michael W Davidson and Mortimer Abramowitz, *Optical microscopy*, Encyclopedia of Imaging Science and Technology **2** (2002), no. 1106-1141, 120.

[21] Cristiano De Michele, Simone Gabrielli, Piero Tartaglia, and Francesco Sciortino, *Dynamics in the presence of attractive patchy interactions*, The Journal of Physical Chemistry B **110** (2006), no. 15, 8064.

[22] Linda R De Young, Anthony L Fink, and Ken A Dill, *Aggregation of globular proteins*, Accounts of Chemical Research **26** (1993), no. 12, 614.

[23] Boris V Derjaguin, *Theory of the stability of strongly charged lyophobic sol and of the adhesion of strongly charged particles in solutions of electrolytes*, Acta Physicochimica U.R.S.S **14** (1941), 633.

[24] Moshe A Dessau and Yorgo Modis, *Protein crystallization for x-ray crystallography*, Journal of Visualized Experiments (2011), no. 47, e2285.

[25] Pierre Duhem, *On the general problem of chemical statics*, Journal of Physical Chemistry **2** (1898), no. 1, 1.

[26] André C Dumetz, Ann M Snellinger-O'Brien, Eric W Kaler, and Abraham M Lenhoff, *Patterns of protein–protein interactions in salt solutions and implications for protein crystallization*, Protein Science **16** (2007), no. 9, 1867.

[27] SD Durbin and G Feher, *Crystal growth studies of lysozyme as a model for protein crystallization*, Journal of Crystal Growth **76** (1986), no. 3, 583.

[28] AV Efimov, *Standard structures in proteins*, Progress in Biophysics and Molecular Biology **60** (1993), no. 3, 201.

[29] David R Eyre, Mercedes A Paz, and Paul M Gallop, *Cross-linking in collagen and elastin*, Annual Review of Biochemistry **53** (1984), no. 1, 717.

[30] Gerald D Fasman, *Crc handbook of biochemistry and molecular biology*, CRC press, 2018.

[31] Lang Feng, Rémi Dreyfus, Ruojie Sha, Nadrian C Seeman, and Paul M Chaikin, *Dna patchy particles*, Advanced Materials **25** (2013), no. 20, 2779.

[32] Patrick J Fleming and Karen G Fleming, *Hullrad: Fast calculations of folded and disordered protein and nucleic acid hydrodynamic properties*, Biophysical Journal **114** (2018), no. 4, 856.

[33] Paul J Flory, *Fundamental principles of condensation polymerization.*, Chemical Reviews **39** (1946), no. 1, 137.

[34] Giuseppe Foffi and Francesco Sciortino, *On the possibility of extending the noro- frenkel generalized law of correspondent states to nonisotropic patchy interactions*, The Journal of Physical Chemistry B **111** (2007), no. 33, 9702.

[35] Madeleine R Fries, Daniel Stopper, Michal K Braun, Alexander Hinderhofer, Fajun Zhang, Robert MJ Jacobs, Maximilian WA Skoda, Hendrik Hansen-Goos, Roland Roth, and Frank Schreiber, *Multivalent-ion-activated protein adsorption reflecting bulk reentrant behavior*, Physical Review Letters **119** (2017), no. 22, 228001.

[36] Madeleine R Fries, Daniel Stopper, Maximilian WA Skoda, Matthias Blum, Christoph Kertzscher, Alexander Hinderhofer, Fajun Zhang, Robert MJ Jacobs, Roland Roth, and Frank Schreiber, *Enhanced protein adsorption upon bulk phase separation*, Scientific Reports **10** (2020), no. 1, 1.

[37] Abraham George and W William Wilson, *Predicting protein crystallization from a dilute solution property*, Acta Crystallographica Section D: Biological Crystallography **50** (1994), no. 4, 361.

[38] Achille Giacometti, Fred Lado, Julio Largo, Giorgio Pastore, and Francesco Sciortino, *Effects of patch size and number within a simple model of patchy colloids*, The Journal of Chemical Physics **132** (2010), no. 17, 174110.

[39] Otto Glatter, *Scattering methods and their application in colloid and interface science*, Elsevier, 2018.

[40] Gerald R Grimsley and C Nick Pace, *Spectrophotometric determination of protein concentration*, Current Protocols in Protein Science **33** (2003), no. 1, 3.

[41] Lutz Grossmann and David Julian McClements, *Current insights into protein solubility: A review of its importance for alternative proteins*, Food Hydrocolloids **137** (2023), 108416.

[42] Guido Guidotti, *Membrane proteins*, Annual Review of Biochemistry **41** (1972), no. 1, 731.

[43] B Guo, S Kao, H McDonald, A Asanov, Leon L Combs, and W William Wilson, *Correlation of second virial coefficients and solubilities useful in protein crystal growth*, Journal of Crystal Growth **196** (1999), no. 2-4, 424.

[44] Amin Haghmoradi, Bennett D Marshall, and Walter G Chapman, *Beyond wertheim's multi-density theory: Steric hindrance and associated rings in a two-density formalism for binary mixtures of molecules with two associating sites*, Journal of Chemical & Engineering Data **65** (2020), no. 12, 5743.

[45] Hugo C Hamaker, *The london—van der waals attraction between spherical particles*, Physica **4** (1937), no. 10, 1058.

[46] JH Hildebrand and SE Wood, *The derivation of equations for regular solutions*, The Journal of Chemical Physics **1** (1933), no. 12, 817.

[47] Joel H Hildebrand, *Solubility. xii. regular solutions1*, Journal of the American Chemical Society **51** (1929), no. 1, 66.

[48] Terrell L Hill, *Theory of solutions. i1*, Journal of the American Chemical Society **79** (1957), no. 18, 4885.

[49] Franz Hofmeister, *Zur lehre von der wirkung der salze: Dritte mittheilung*, Archiv für experimentelle Pathologie und Pharmakologie **25** (1888), 1.

[50] Franz Hofmeister, *Zur lehre von der wirkung der salze: zweite mittheilung*, Archiv für experimentelle Pathologie und Pharmakologie **24** (1888), 247.

[51] Roy Hughes and T Cosgrove, *An introduction to colloids*, Colloid Science: Principles, Methods and Applications (2010), 1.

[52] George Jackson, Walter G. Chapman, and Keith E. Gubbins, *Phase equilibria of associating fluids spherical molecules with multiple bonding sites*, Molecular Physics **65** (1988), no. 1, 1–31.

[53] Bruno Jirgensons and Martins Eduards Straumanis, *A short textbook of colloid chemistry*, Elsevier, 2013.

[54] Vitaly I Kalikmanov, *Classical nucleation theory*, Nucleation Theory, Springer, 2012, p. 17.

[55] Beibei Kang, Huicheng Tang, Zengdian Zhao, and Shasha Song, *Hofmeister series: Insights of ion specificity from amphiphilic assembly and interface property*, ACS Omega **5** (2020), no. 12, 6229.

[56] Dimo Kashchiev, *Classical nucleation theory approach to two-step nucleation of crystals*, Journal of Crystal Growth **530** (2020), 125300.

[57] Nikolay N Khechinashvili, Joël Janin, and Francis Rodier, *Thermodynamics of the temperature-induced unfolding of globular proteins*, Protein Science **4** (1995), no. 7, 1315.

[58] Jaibir Kherb, Sarah C Flores, and Paul S Cremer, *Role of carboxylate side chains in the cation hofmeister series*, The Journal of Physical Chemistry B **116** (2012), no. 25, 7389.

[59] Michael Y Kiriukhin and Kim D Collins, *Dynamic hydration numbers for biologically important ions*, Biophysical Chemistry **99** (2002), no. 2, 155.

[60] Daphne Klotsa and Robert L Jack, *Predicting the self-assembly of a model colloidal crystal*, Soft Matter **7** (2011), no. 13, 6294.

[61] Irving M Klotz, NR Langebman, and DW Dahnall, *Quaternary structure of proteins*, Annual Review of Biochemistry **39** (1970), no. 1, 25.

[62] Mika M Kohonen, Marilyn E Karaman, and Richard M Pashley, *Debye length in multivalent electrolyte solutions*, Langmuir **16** (2000), no. 13, 5749.

[63] André Kusmin, *Hydration of cyclodextrins studied by neutron and x-ray scattering*, Ph.D. thesis, Freie Universität Berlin, Department of Biology, Chemistry and Pharmacy, 2007.

[64] André Kusmin, Ruep E Lechner, Martin Kammel, and Wolfram Saenger, *Native and methylated cyclodextrins with positive and negative solubility coefficients in water studied by saxs and sans*, The Journal of Physical Chemistry B **112** (2008), no. 41, 12888.

[65] S. Lewith, *Zur lehre von der wirkung der salze: Erste mittheilung*, Archiv für experimentelle Pathologie und Pharmakologie **24** (1887), 1.

[66] Yingxin Liu, Xiujuan Wang, and Chi Bun Ching, *Toward further understanding of lysozyme crystallization: phase diagram, protein- protein interaction, nucleation kinetics, and growth kinetics*, Crystal Growth & Design **10** (2010), no. 2, 548.

[67] Jonathan A Lukin, Georg Kontaxis, Virgil Simplaceanu, Yue Yuan, Ad Bax, and Chien Ho, *Quaternary structure of hemoglobin in solution*, Proceedings of the National Academy of Sciences **100** (2003), no. 2, 517.

[68] Thu Zar Lwin, Ruhong Zhou, and Ray Luo, *Is poisson-boltzmann theory insufficient for protein folding simulations?*, The Journal of Chemical Physics **124** (2006), no. 3.

[69] Ralph Maier, Madeleine R Fries, Cara Buchholz, Fajun Zhang, and Frank Schreiber, *Human versus bovine serum albumin: A subtle difference in hydrophobicity leads to large differences in bulk and interface behavior*, Crystal Growth & Design **21** (2021), no. 9, 5451.

[70] Ralph Maier, Benedikt Sohmen, Stefano Da Vela, Olga Matsarskaia, Christian Beck, Ralf Schweins, Tilo Seydel, Fajun Zhang, and Frank Schreiber, *Protein crystallization from a preordered metastable intermediate phase followed by real-time small-angle neutron scattering*, Crystal Growth & Design **21** (2021), no. 12, 6971.

[71] Ralph Maier, Georg Zocher, Andrea Sauter, Stefano Da Vela, Olga Matsarskaia, Ralf Schweins, Michael Sztucki, Fajun Zhang, Thilo Stehle, and Frank Schreiber, *Protein crystallization in the presence of a metastable liquid–liquid phase separation*, Crystal Growth & Design **20** (2020), no. 12, 7951.

[72] Marian Manciu and Eli Ruckenstein, *Role of the hydration force in the stability of colloids at high ionic strengths*, Langmuir **17** (2001), no. 22, 7061.

[73] Olga Matsarskaia, Michal K Braun, Felix Roosen-Runge, Marcell Wolf, Fajun Zhang, Roland Roth, and Frank Schreiber, *Cation-induced hydration effects cause lower critical solution temperature behavior in protein solutions*, The Journal of Physical Chemistry B **120** (2016), no. 31, 7731.

[74] Olga Matsarskaia, Felix Roosen-Runge, Gudrun Lotze, Johannes Möller, Alessandro Mariani, Fajun Zhang, and Frank Schreiber, *Tuning phase tran-*

sitions of aqueous protein solutions by multivalent cations, Physical Chemistry Chemical Physics **20** (2018), no. 42, 27214.

[75] Joseph E Mayer, *The theory of ionic solutions*, The Journal of Chemical Physics **18** (1950), no. 11, 1426.

[76] William G McMillan Jr and Joseph E Mayer, *The statistical thermodynamics of multicomponent systems*, Journal of Chemical Physics **13** (1945), no. 7, 276.

[77] SVG Menon, C Manohar, and K Srinivasa Rao, *A new interpretation of the sticky hard sphere model*, Journal of Chemical Physics **95** (1991), no. 12, 9186.

[78] Jerome Mertz, *Introduction to optical microscopy*, Cambridge University Press, 2019.

[79] Hartmut Michel, *Crystallization of membrane proteins*, Trends in Biochemical Sciences **8** (1983), no. 2, 56.

[80] Mark A Miller and Daan Frenkel, *Phase diagram of the adhesive hard sphere fluid*, Journal of Chemical Physics **121** (2004), no. 1, 535.

[81] Tiziana Missana and Andrés Adell, *On the applicability of dlvo theory to the prediction of clay colloids stability*, Journal of Colloid and Interface Science **230** (2000), no. 1, 150.

[82] K Miyazaki, KS Schweizer, D Thirumalai, R Tuinier, and E Zaccarelli, *The asakura–oosawa theory: Entropic forces in physics, biology, and soft matter*, The Journal of Chemical Physics **156** (2022), no. 8.

[83] Marko Mravlak, *Depletion force*, University of Ljubljana (2008), 3.

[84] Christo N Nanev and Vesselin D Tonchev, *Sigmoid kinetics of protein crystal nucleation*, Journal of Crystal Growth **427** (2015), 48.

[85] Hans Neurath, Jesse P Greenstein, Frank W Putnam, and John A Erickson, *The chemistry of protein denaturation.*, Chemical Reviews **34** (1944), no. 2, 157.

[86] Satoshi Nihonyanagi, Shoichi Yamaguchi, and Tahei Tahara, *Counterion effect on interfacial water at charged interfaces and its relevance to the hofmeister series*, Journal of the American Chemical Society **136** (2014), no. 17, 6155.

[87] James E Noble and Marc JA Bailey, *Quantitation of protein*, Methods in Enzymology **463** (2009), 73.

[88] Halil I Okur, Jana Hladílková, Kelvin B Rembert, Younhee Cho, Jan Heyda, Joachim Dzubiella, Paul S Cremer, and Pavel Jungwirth, *Beyond the hofmeister series: Ion-specific effects on proteins and their biological functions*, The Journal of Physical Chemistry B **121** (2017), no. 9, 1997.

[89] Lars Onsager, *Theories of concentrated electrolytes.*, Chemical Reviews **13** (1933), no. 1, 73.

[90] Paul R Ortiz de Montellano, *Control of the catalytic activity of prosthetic heme by the structure of hemoproteins*, Accounts of Chemical Research **20** (1987), no. 8, 289.

[91] J Th G Overbeek, *Colloids, a fascinating subject: Introductory lecture*, Colloidal Dispersions. Royal Society of Chemistry, London, UK (1982).

[92] Avanish S Parmar and Martin Muschol, *Hydration and hydrodynamic interactions of lysozyme: effects of chaotropic versus kosmotropic ions*, Biophysical Journal **97** (2009), no. 2, 590.

[93] Amar B Pawar and Ilona Kretzschmar, *Fabrication, assembly, and application of patchy particles*, Macromolecular Rapid Communications **31** (2010), no. 2, 150.

[94] Oliver Penrose, *Foundations of statistical mechanics*, Reports on Progress in Physics **42** (1979), no. 12, 1937.

[95] Andrei V Petukhov, Remco Tuinier, and Gert Jan Vroege, *Entropic patchiness: Effects of colloid shape and depletion*, Current Opinion in Colloid & Interface Science **30** (2017), 54.

[96] R Piazza, S Iacopini, and M Galliano, *Blga protein solutions at high ionic strength: Vanishing attractive interactions and "frustrated" aggregation*, Europhysics Letters **59** (2002), no. 1, 149.

[97] Kenneth S Pitzer, *Corresponding states for perfect liquids*, The Journal of Chemical Physics **7** (1939), no. 8, 583.

[98] Florian Platten, Néstor E Valadez-Pérez, Ramón Castañeda-Priego, and Stefan U Egelhaaf, *Extended law of corresponding states for protein solutions*, Journal of Chemical Physics **142** (2015), no. 17, 05B6021.

[99] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.“*, Journal of Molecular Biology **7** (1963), 95.

[100] Flavio Romano, Eduardo Sanz, Piero Tartaglia, and Francesco Sciortino, *Phase diagram of trivalent and pentavalent patchy particles*, Journal of Physics: Condensed Matter **24** (2012), no. 6, 064113.

[101] Felix Roosen-Runge, Benjamin S Heck, Fajun Zhang, Oliver Kohlbacher, and Frank Schreiber, *Interplay of ph and binding of multivalent metal ions: charge inversion and reentrant condensation in protein solutions*, The Journal of Physical Chemistry B **117** (2013), no. 18, 5777.

[102] Felix Roosen-Runge, Marcus Hennig, Fajun Zhang, Robert MJ Jacobs, Michael Sztucki, Helmut Schober, Tilo Seydel, and Frank Schreiber, *Protein self-diffusion in crowded solutions*, Proceedings of the National Academy of Sciences **108** (2011), no. 29, 11815.

[103] Felix Roosen-Runge, Fajun Zhang, Frank Schreiber, and Roland Roth, *Ion-activated attractive patches as a mechanism for controlled protein interactions*, Scientific Reports **4** (2014), no. 1, 1.

[104] Charles M Roth, Brian L Neal, and Abraham M Lenhoff, *Van der waals interactions involving proteins*, Biophysical Journal **70** (1996), no. 2, 977.

[105] John A Rupley, Enrico Gratton, and Giorgio Careri, *Water and globular proteins*, Trends in Biochemical Sciences **8** (1983), no. 1, 18.

[106] William Bailey Russel, WB Russel, Dudley A Saville, and William Raymond Schowalter, *Colloidal dispersions*, Cambridge university press, 1991.

[107] J Russo, JM Tavares, PIC Teixeira, MM Telo da Gama, and Francesco Sciortino, *Re-entrant phase behaviour of network fluids: A patchy particle model with temperature-dependent valence*, The Journal of Chemical Physics **135** (2011), no. 3, 034501.

[108] Masahide Sato, *Clusters formed by dumbbell-like one-patch particles confined in thin systems*, Scientific Reports **11** (2021), no. 1, 18078.

[109] Andrea Sauter, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Robert MJ Jacobs, and Frank Schreiber, *Real-time observation of nonclassical protein crystallization kinetics*, Journal of the American Chemical Society **137** (2015), no. 4, 1485.

[110] Frank Schreiber, Fabio Zanini, and Felix Roosen-Runge, *Virial expansion–a brief introduction*, available at: www. soft-matter. uni-tuebingen. de/teaching/Tutorial_Virial_Expansion. pdf (accessed 5 June 2018) (2011).

[111] Nadine Schwierz, Dominik Horinek, Uri Sivan, and Roland R Netz, *Reversed hofmeister series—the rule rather than the exception*, Current Opinion in Colloid & Interface Science **23** (2016), 10.

[112] Richard P Sear, *Nucleation: theory and applications to protein solutions and colloidal suspensions*, Journal of Physics: Condensed Matter **19** (2007), no. 3, 033101.

[113] Richard P Sear and Daan Frenkel, *Phase behavior of colloid plus polydisperse polymer mixtures*, Physical Review E **55** (1997), no. 2, 1677.

[114] Maximilian D Senft, Ralph Maier, Anusha Hiremath, Fajun Zhang, and Frank Schreiber, *Effective interactions and phase behavior of protein solutions in the presence of hexamine cobalt (iii) chloride*, The European Physical Journal E **46** (2023), no. 12, 119.

[115] Frank Smallenburg, Ludwik Leibler, and Francesco Sciortino, *Patchy particle model for vitrimers*, Physical Review Letters **111** (2013), no. 18, 188002.

[116] Shigetoshi Sugio, Akiko Kashima, Shinichi Mochizuki, Mariko Noda, and Kumiko Kobayashi, *Crystal structure of human serum albumin at 2.5 å resolution*, Protein Engineering **12** (1999), no. 6, 439.

[117] Shahar Sukenik, Liel Sapir, and Daniel Harries, *Balance of enthalpy and entropy in depletion forces*, Current Opinion in Colloid & Interface Science **18** (2013), no. 6, 495.

[118] Furio Surfaro, Ralph Maier, Kai-Florian Pastryk, Fajun Zhang, Frank Schreiber, and Roland Roth, *An alternative approach to the osmotic second virial coefficient of protein solutions and its application to liquid–liquid phase separation*, The Journal of Chemical Physics **158** (2023), no. 16.

[119] Dmitriĭ Ivanovich Svergun, Michel HJ Koch, Peter A Timmins, and Roland P May, *Small angle x-ray and neutron scattering from solutions of biological macromolecules*, vol. 19, OUP Oxford, 2013.

[120] Robert Swendsen, *An introduction to statistical mechanics and thermodynamics*, Oxford University Press, USA, 2020.

[121] Gregor Trefalt, Istvan Szilagyi, and Michal Borkovec, *Poisson–boltzmann description of interaction forces and aggregation rates involving charged colloidal particles in asymmetric electrolytes*, Journal of Colloid and Interface Science **406** (2013), 111.

[122] Shaghayegh Vafaei, Bruno Tomberli, and CG Gray, *Mcmillan-mayer theory of solutions revisited: Simplifications and extensions*, The Journal of Chemical Physics **141** (2014), no. 15.

[123] JJ Valle-Delgado, JA Molina-Bolívar, F Galisteo-González, and MJ Gálvez-Ruiz, *Evidence of hydration forces between proteins*, Current Opinion in Colloid & Interface Science **16** (2011), no. 6, 572.

[124] Johannes Diderik van der Waals, *Contribution to the theory of binary mixtures. vi. the plaitpoint line*, KNAW, Proceedings, vol. 10, 1907, p. 183.

[125] Johannes Diderik van der Waals, *Some remarks on the value of the volumes of the coexisting phases of a simple substance. i*, KNAW, Proceedings, vol. 13, 1910, p. 1910.

[126] Johannes Diderik van der Waals, *The law of corresponding states for different substances*, KNAW, Proceedings, vol. 15, 1913, p. 1912.

[127] Peter G Vekilov, *Nucleation*, Crystal Growth & Design **10** (2010), no. 12, 5007.

[128] Henk Verduin and Jan KG Dhont, *Phase diagram of a model adhesive hard-sphere dispersion*, Journal of Colloid and Interface Science **172** (1995), no. 2, 425.

[129] Nina Vlachy, Barbara Jagoda-Cwiklik, Robert Vácha, Didier Touraud, Pavel Jungwirth, and Werner Kunz, *Hofmeister series and specific interactions of charged headgroups with aqueous ions*, Advances in Colloid and Interface Science **146** (2009), no. 1-2, 42.

[130] G. A. Vliegenthart and H. N. W. Lekkerkerker, *Predicting the gas–liquid critical point from the second virial coefficient*, Journal of Chemical Physics **112** (2000), no. 12, 5364.

[131] Michael S Wertheim, *Fluids with highly directional attractive forces. i. statistical thermodynamics*, Journal of Statistical Physics **35** (1984), no. 1-2, 19.

[132] Michael S Wertheim, *Fluids with highly directional attractive forces. ii. thermodynamic perturbation theory and integral equations*, Journal of Statistical Physics **35** (1984), no. 1-2, 35.

[133] Michael S Wertheim, *Fluids with highly directional attractive forces. iii. multiple attraction sites*, Journal of Statistical Physics **42** (1986), no. 3-4, 459.

[134] Stephen Whitelam and Robert L Jack, *The statistical mechanics of dynamic pathways to self-assembly*, Annual Review of Physical Chemistry **66** (2015), 143.

[135] Marcell Wolf, Felix Roosen-Runge, Fajun Zhang, Roland Roth, Maximilian WA Skoda, Robert MJ Jacobs, Michael Sztucki, and Frank Schreiber, *Effective interactions in protein–salt solutions approaching liquid–liquid phase separation*, Journal of Molecular Liquids **200** (2014), 20.

[136] Gary Wulfsberg, *Principles of descriptive inorganic chemistry*, University Science Books, 1991.

[137] Fajun Zhang, José A Gavira, Geun Woo Lee, and Dirk Zahn, *Nonclassical nucleation—role of metastable intermediate phase in crystal nucleation: An editorial prefix*, Crystals **11** (2021), no. 2, 174.

[138] Fajun Zhang, Roland Roth, Marcell Wolf, Felix Roosen-Runge, Maximilian WA Skoda, Robert MJ Jacobs, Michael Stzucki, and Frank Schreiber, *Charge-controlled metastable liquid–liquid phase separation in protein solutions as a universal pathway towards crystallization*, Soft Matter **8** (2012), no. 5, 1313.

[139] Fajun Zhang, Maximilian WA Skoda, Robert MJ Jacobs, Richard A Martin, Christopher M Martin, and Frank Schreiber, *Protein interactions studied by saxs: effect of ionic strength and protein concentration for bsa in aqueous solutions*, The Journal of Physical Chemistry B **111** (2007), no. 1, 251.

[140] Fajun Zhang, Georg Zocher, Andrea Sauter, Thilo Stehle, and Frank Schreiber, *Novel approach to controlled protein crystallization through liganda-tion of yttrium cations*, Journal of Applied Crystallography **44** (2011), no. 4, 755.

[141] Hong Zhang, Xiong Ji, Pilong Li, Cong Liu, Jizhong Lou, Zheng Wang, Wenyu Wen, Yue Xiao, Mingjie Zhang, and Xueliang Zhu, *Liquid-liquid phase separation in biology: mechanisms, physiological functions and human diseases*, Science China Life Sciences **63** (2020), no. 7, 953.

[142] Yanjie Zhang and Paul S Cremer, *Interactions between macromolecules and ions: the hofmeister series*, Current Opinion in Chemical Biology **10** (2006), no. 6, 658.