The Hydrophobic Interaction

Modeling Hydrophobic Interactions and Aggregation of Non-Polar Particles in Aqueous Solutions

Thèse de doctorat

présentée à la
Faculté des Sciences de
l'Université de Lausanne
par

Susanne Mölbert

Diplômée en Physique
Université de Fribourg

Jury

Prof. Jean Hernandez, Président
Prof. Paolo De Los Rios, Directeur de Thèse
Prof. Giovanni Dietler, Rapporteur
Dr. Chao Tang, Expert
Dr. Pierpaolo Bruscolini, Expert

Lausanne
2003
Acknowledgments

I have benefited from long-term collaborations, discussions, and exchange of ideas and results with a large number of people. Most importantly, I am indebted to Paolo De Los Rios for his many ideas, useful suggestions, and interesting comments. He provided the motivation to begin the work, guidance in the methods to use, and constant reality checks for my results. I wish to thank all my collaborators at the Institute of Theoretical Physics at the University of Lausanne for their support, in particular Shin Miyahara and François Vernay, not least for their useful comments, valuable input, and help with specific problems.

I am grateful to Chao Tang at the NEC Research Institute in Princeton, NJ (USA), in whose group the study of hydrophobicity in proteins was performed. I also wish to thank Eldon Emberly and Jonathan Miller for many helpful discussions. This piece of research was supported in part by The Swiss Study Foundation.

I am particularly indebted to Bruce Normand for his encouragement, enlightening discussions, and constructive criticism, and especially for sacrificing so much of his time to the careful reading of this thesis.

This work was supported financially by the Swiss National Science Foundation through grants FNRS 21-61397.00 and 2000-67886.02.
Abstract

Hydration of hydrophobic solutes in water is the cause of different phenomena, including the hydrophobic heat-capacity anomaly. We use a simple yet powerful mixture model for water, an adapted two-state Muller-Lee-Graziano model, to describe the energy levels of water molecules as a function of their proximity to non-polar solute molecules. The model is shown to provide an appropriate description of many-body interactions between the hydrophobic solute particles. The solubility and aggregation of hydrophobic substances is studied by evaluating detailed Monte Carlo simulations in the vicinity of the first-order aggregation phase transition. A closed-loop coexistence curve is found, which is consistent with a mean-field calculation carried out for the same system.

In addition, we have studied the aggregation of hydrophobic particles in aqueous solutions in the presence of cosolvents. Here we have demonstrated that the important features of the effect of cosolvents on hydrophobic aggregates may be described within the same two-state model, with adaptations to focus on the ability of such substances to alter the structure of water. The relevant phenomena include a significant change in solubility of non-polar solute particles and preferential binding or exclusion of such substances to solute molecules.

We have further adapted the MLG model to include the solvation of amphiphilic solute particles in water. By allowing different distributions of hydrophobic regions at the molecular surface, we have found aggregation of the amphiphiles, and formation of various types of micelle as a function of the hydrophobicity pattern. We demonstrate that the essential features of micelle formation, usually attributed to the amphiphilic nature of the solute particles, are primarily solvent-induced.

Hydrophobicity remains a controversial quantity also in protein science. Statistical studies have shown that some amino-acids are found preferentially either in the core or on the surface of native folds, whereas many are present at all positions with equal probability. Based on these results we have presented the average surface-accessibility scale, which may lead to an improvement in the comparison of experimental data with the results from theoretical HP models. We verify the validity of the new scale on secondary-structure elements.
# Contents

<table>
<thead>
<tr>
<th>1 Introduction</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Aqueous Solutions</td>
<td>9</td>
</tr>
<tr>
<td>1.2 Hydrophobic-Polar Model</td>
<td>10</td>
</tr>
<tr>
<td>1.3 Cosolvent Effects</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Solutions of Amphiphilic Molecules</td>
<td>15</td>
</tr>
<tr>
<td>1.5 Hydrophobicity in Protein Folding</td>
<td>18</td>
</tr>
<tr>
<td>1.6 Overview</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 Model and Methods</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Energy Levels</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Coarse-Grained Model</td>
<td>25</td>
</tr>
<tr>
<td>2.3 Hydrophobic Interaction</td>
<td>27</td>
</tr>
<tr>
<td>2.4 Mean-Field Calculation</td>
<td>28</td>
</tr>
<tr>
<td>2.5 Molecular-Level Simulations</td>
<td>29</td>
</tr>
<tr>
<td>2.5.1 Monte Carlo Simulation</td>
<td>29</td>
</tr>
<tr>
<td>2.6 Pair Approximation</td>
<td>30</td>
</tr>
<tr>
<td>2.6.1 Cluster-Variation Method</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 Solution of Hydrophobic Particles in Water</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Mean-Field Calculation</td>
<td>35</td>
</tr>
<tr>
<td>3.2 Molecular-Level Simulations</td>
<td>40</td>
</tr>
<tr>
<td>3.3 Discussion</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 Chaotropic Effect</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Model including Chaotropic Cosolvent</td>
<td>49</td>
</tr>
<tr>
<td>4.2 Mean-Field Approximation</td>
<td>54</td>
</tr>
<tr>
<td>4.2.1 Single Mean Cosolvent Concentration</td>
<td>54</td>
</tr>
<tr>
<td>4.2.2 Distinct Mean Bulk and Shell Concentrations</td>
<td>56</td>
</tr>
<tr>
<td>4.2.3 Results</td>
<td>57</td>
</tr>
<tr>
<td>4.3 Molecular-Level Simulations</td>
<td>60</td>
</tr>
<tr>
<td>4.4 Pair Approximation</td>
<td>65</td>
</tr>
<tr>
<td>4.4.1 Model in Pair Approximation</td>
<td>66</td>
</tr>
</tbody>
</table>
Contents

4.4.2 Results .................................................. 69
4.5 Discussion ............................................... 71

5 Kosmotropic Effect ........................................... 75
  5.1 Model including Kosmotropic Cosolvent ................. 76
  5.2 Methods .................................................. 78
  5.3 Results and Discussion .................................. 79

6 Micelle Formation ........................................... 85
  6.1 Hydrophobic-Polar Model for Amphiphiles ............... 86
      6.1.1 Method ............................................ 89
      6.1.2 Results ........................................... 90
  6.2 Surfactant Model ....................................... 95
      6.2.1 Methods ........................................... 99
      6.2.2 Results ........................................... 100
  6.3 Discussion ............................................... 102

7 Hydrophobicity in Protein Folding ......................... 109
  7.1 Protein-Structure Classification ......................... 112
  7.2 Testing Hydrophobicity Scales .......................... 114
  7.3 Surface-Accessibility Scale ............................. 118
  7.4 Hydrophobic-Polar Model ............................... 122
  7.5 Secondary-Structure Analysis ......................... 125
  7.6 Discussion ............................................. 128

8 Summary ..................................................... 133

Résumé en français ............................................ 137

Bibliography ................................................. 139
Prologue

Life as we know it originated in water, and could not exist in its absence. It has been conditioned by the abnormal properties of water, because water was present on this planet long before the evolution of life. Water forms a necessary constituent of the cells of all animal and plant tissues, and is essential for the maintenance of organic life. Natural processes are characterized by the economy with which energy (matter) is used. However, organisms consist of up to 95% water, and it seems permissible to conclude that this liquid must fulfil a function other than that of an inert substrate. Only little is known about the manner in which water acts in the formation of organized biological structures at the subcellular, cellular, and multicellular level, while at the molecular level the role of water in the stabilization of native conformations of biopolymers has only recently begun to receive attention.

The importance of the unique physical properties of water, and in particular of its unique solvent power, has been cited frequently [1, 2, 3, 4]. However, the equally great importance of its unique lack of solvent power for many non-polar substances has attracted far less attention. Many of these non-polar molecules are soluble in alcohol and other solvents, but not in water. This poor solubility of non-polar substances in water, which is known as the hydrophobic (from the Greek “hydro” = “water” and “phobia” = “dread”) effect, is perhaps the most important single factor in the organization of the constituent molecules of living matter into complex structural entities such as cell membranes and organelles. Specifically, particles of dual nature, consisting of one part which is soluble in water and another part which is expelled from it, are forced by their duality to adopt unique orientations with respect to the aqueous medium, and to form suitably organized structures. The self-assembling properties of such molecules are crucial in the formation of living matter. The cell membrane, which in effect defines the living cell and allows it to exist as a coherent entity, is perhaps the best example [4, 5]. There can be little doubt that its formation is spontaneous, dependent only on the fact that its constituent molecules are partly hydrophilic (“water-loving”) and partly hydrophobic (“water-fearing”).

Despite its importance for life, the physical mechanisms underlying the hydrophobic effect and the resulting hydrophobic interaction, which is the interaction of non-polar particles in aqueous media, are at present not well understood and require further investigation. The aim of this study is to clarify the origin of the hydrophobic
interaction using a simple yet powerful model of water, which includes the essential features of aqueous solutions of hydrophobic molecules and reproduces various aspects of their behavior.
Chapter 1

Introduction

1.1 Aqueous Solutions

Hydrophobic substances are defined as those which are readily soluble in many non-polar solvents, but only barely soluble in water, in contrast to substances which form solids with strong intermolecular cohesion, and thus generally exhibit low solubility in all solvents. This distinction is especially important from the biological point of view, because it means that molecules expelled from water as a result of their hydrophobicity will tend to remain in a fluid, deformable state.

The existence of hydrophobic substances, and of dual-nature organic molecules containing polar and non-polar portions, has been known for a long time, but the mechanism underlying the hydrophobic effect remains controversial. For a considerable period, the association between hydrocarbon chains in the formation of micelles was believed to arise from their “like to like” attraction [6, 7]. However, it has been recognized more recently that the attraction of non-polar groups (such as hydrocarbon chains) plays only a minor role in the hydrophobic effect [8, 9]. The effect rather arises primarily from the strong attractive forces between isotropically arranged water molecules. These arrangements must be disrupted or distorted when any solute is dissolved in the water. If the solute particles are ionic or strongly polar, they can form strong bonds to water molecules, which more than compensate for the disruption or distortion of the bonds existing in pure water. These substances thus tend to be easily soluble in water. No such compensation occurs with non-polar groups, and their solution in water is accordingly resisted.

An initial interpretation of the hydrophobic effect, assigning the predominant role to the properties of water per se, was provided by Frank and Evans [8] in their interpretation of the thermodynamic properties of aqueous solutions of all kinds of hydrophobic and partially hydrophobic substances. They noted the ability of water to form strong hydrogen-bonded networks as illustrated in Fig. 1.1, but realized that the existence of long-lived structures in liquid water is most unlikely. Kauzmann [10],
in his analysis of the forces stabilizing the native structure of proteins, discussed the role of water in determining protein conformation and denaturation, and emphasized the unique nature of the solvent medium in which the processes of life take place.

![Schematic diagrams of ice structure, ordered liquid water, and disordered liquid water.](image)

Figure 1.1: Schematic, two-dimensional projection of the structure of crystalline ice (left), ordered liquid water (middle), and disordered liquid water (right). Disordered liquid water is characterized by extremely bent and weak hydrogen bonds, and by the presence of many unbonded molecules. Ordered liquid water has rather straight, ice-like bonds which are therefore stronger than those in disordered water. The density of disordered water is higher than that of ordered water. Oxygen atoms are shown in red, hydrogen atoms in white and hydrogen bonds in green.

Many recent studies have focused on the surprising thermodynamic properties associated with the hydration of hydrophobic substances in aqueous solutions [4, 11, 12]. Indeed, the heat capacity for transfer of non-polar solutes to water is very large at room temperature. This behavior, known as the hydrophobic heat-capacity anomaly, stands in sharp contrast to the observations made for hydrophilic solutes in water [2]. Experiments confirm that the solubility of small hydrocarbons in water decreases when the solution is heated near room temperature. It is generally believed that these thermodynamic phenomena are associated with a structural change in the solvent, and the hydrophobic effect is therefore considered as directly related to the anomalous properties of liquid water [3].

### 1.2 Hydrophobic-Polar Model

Water molecules have the ability to form strong, intermolecular hydrogen bonds. Pure, liquid water forms extended hydrogen-bonded networks, and is thus highly ordered (Fig. 1.1). Although the insertion of a hydrophobic molecule leads to a destruction of local hydrogen-bond structure and hence to considerable entropy gain and
enthalpy loss, water molecules are found to rearrange in a cage-like structure around the solute molecule formed by even slightly stronger hydrogen bonds (Fig. 1.2), resulting in a net reduction of enthalpy during the insertion process [3, 8, 13, 15]. This theory is generally accepted, and confirmed experimentally for amino-acids [14], with the result that the hydrophobic effect is widely believed to be a consequence of changes in the ordering of water molecules, rather than being explained by water-solute interactions. Moreover, the hydrophobic interaction between non-polar solute molecules is mainly solvent-induced, through a minimization of the total surface exposed to water as two molecules approach each other, which reduces the local restructuring of water and thus increases the entropy. The free energy of the system therefore decreases during aggregation of hydrophobic particles if the enthalpy change for the process is suitably small. Aggregation of non-polar particles in aqueous solutions can thus be attributed to the hydrophobic interaction between the solute molecules.

At room temperature the solubility of small, hydrophobic solute particles in water increases when the system is cooled. At sufficiently low temperatures a homogeneous mixture is found, provided that none of the components crystallizes. Heating the mixture produces a decrease in solubility, which can result in a phase transition to a state with two phases of different solute densities. This transition temperature is called the Lower Critical Solution Temperature, (LCST) and has been measured for different solutions of hydrophobic particles in water [16, 17, 18, 19]. After reaching a minimum, the solubility increases steeply at higher temperatures, and if a coexistence of two phases is found near room temperature, another phase transition to one homogeneous phase occurs at the Upper Critical Solution Tem-
Introduction

Hydrophobic-Polar Model

Temperature (UCST), on condition that the boiling point of the solution is not reached [20].

Figure 1.3: Schematic representation of the temperature dependence of the solubility of hydrophobic particles in water. Between the LCST and the UCST the particles aggregate, while below the LCST and above the UCST the system is a homogeneous solution. Corresponding configurations on the molecular level are illustrated on the right side.

Aqueous solutions of non-polar particles which show a LCST also have an UCST, and therefore a closed-loop coexistence curve in the phase diagram, provided that none of the liquids undergoes a phase transition to a gaseous phase before that temperature is reached. In fact the rising temperature increases the entropy, which promotes a homogeneous mixing of the components. In addition, the number of hydrogen bonds formed, which is responsible for the phase separation, decreases with increasing temperature. This closed-loop miscibility curve has been found in different binary solutions, including nicotine/water and poly(ethylene glycol)/water [16, 17].

The hydrophobic interaction of non-polar particles in aqueous solutions can be explained by the concept the solvent forming two physically distinct types of state: shell (hydration) water and bulk water [9, 21]. This notion is described successfully by a simple form of hydrophobic polar (HP) model due to Muller, Lee, and Graziano (MLG) [22, 23], which focuses on changes in structural arrangements of liquid water in the presence of hydrophobic solute particles.
A complete understanding of the molecular mechanisms underlying the hydrophobic effect is essential in order to explain a variety of phenomena, including the aggregation of hydrophobic solutes in water and the destabilization, denaturation and aggregation of proteins. These important biological processes are responsible, among other things, for different conditions such as Parkinson’s disease, Alzheimer’s disease, Creutzfeld-Jacob disease, and sickle-cell anemia [24, 25, 26].

1.3 Cosolvent Effects

Most processes in living organisms are adjusted to function in a rather limited range of physiological conditions, and important deviations, such as high concentrations of dangerous substances, are generally expected to preclude life. Nevertheless, many living systems survive stresses of this kind and exist in hostile environments. One common adaptation strategy is to modify the properties of the solvent, which is usually water, in such a way as to exclude the undesirable solutes from solution [27]. Water is modified by relatively high concentrations of stabilizing solutes, which are therefore often referred to as cosolvents, and which remain compatible with the metabolism of the cell even at very high concentrations (therefore they are also referred to as ‘compatible osmolytes’) [28, 29]. These cosolvents neutralize dangerous solutes by decreasing their solubility and enhancing the formation of their aggregates. Such cosolvents are known as promoters of the water structure and are therefore referred to as kosmotropes (‘kosmo-trope’ = order maker). Many recent investigations have focused on the ability of cosolvents to influence the solubility of hydrophobic solute particles in aqueous media [9, 21]. However, the exact physical mechanism for the changes in water structure underlying the stabilizing function of kosmotropic cosolvents is at present not fully understood.

Recently there has also been growing interest in the mechanisms underlying the ability of chaotropic cosolvents to increase the solubility of hydrophobic solutes in aqueous solutions [30, 31, 32, 33, 34, 35, 36, 37]. In fact, in some cases the solubility can be enhanced by several orders of magnitude, and the properties of hydrophobic molecules are affected in a way that destabilizes their aggregation [38, 39, 40]. A remarkable number of substances, including urea, decrease the stability of proteins in water, and can even result in a complete denaturation [41]. A highly concentrated solution of urea is therefore often used as a protein denaturant. In addition, decreased micelle formation has been observed in aqueous urea solutions [42, 43]. The solution must contain a high concentration (typically 0.5-10 M) of destabilizing agent in order to show such an effect.

The underlying cause of this process, known as the chaotropic effect, is generally believed to be a decrease in the order of the water structure (‘chao-trope’ = disorder
maker), thus indirectly increasing the solubility of non-polar solutes [44, 45]. However, different attempts to discover how chaotropic agents perform this function have not yet been able to explain satisfactorily the exact mechanism of hydrogen-bond disruption which stabilizes the aggregate.

The addition of a cosolvent to a solution of hydrophobic particles in water can affect the solubility, and consequently the stability of aggregates, indirectly through its effects on the solvent. Chaotropic substances, which cause an increase in the solubility of hydrophobic particles in water decreasing the extent of the closed-loop coexistence curve, have been found preferentially at the surfaces of solute molecules in experiment [33]. In contrast, kosmotropic substances, which decrease the solubility of hydrophobic particles and stabilize their aggregates, are excluded from the immediate surroundings of non-polar solute particles [46]. This preferential exclusion leads to a higher number of water molecules in the solvation shell of solutes, and is therefore also known as preferential hydration. The net repulsion between hydrophobic solute and solvent increases, which causes the system to minimize the interface between them. Kosmotropic cosolvents thus lead indirectly to a stabilization of aggregates or of native protein structures, because their exposed surface is minimal compared to that of any disaggregated or denatured state. A schematic representation of solvent and cosolvent distributions around solute particles is shown in Fig. 1.4.

By contrast, for chaotropic substances a higher local cosolvent concentration is contained in the solvation shell of hydrophobic particles than in the bulk solvent. This effect, known as preferential binding (although the cosolvents are pushed from the bulk water into the surrounding shell of the solute, rather than being bound to the latter), results in a smaller number of water molecules in contact with the surface of non-polar solute particles. This leads to a weaker net interaction between the solute molecules and the solvent, such that a larger interface between solvent and solute becomes favorable. The addition of chaotropic cosolvents to aqueous solutions therefore results in an increase in solvent-accessible surface area which destabilizes aggregates and native protein states [30, 31, 32, 33, 47, 48].

Such preferential interactions with hydrophobic particles appear to arise primarily from the properties of the solvent, rather than from interactions between cosolvent and particles [49]. The free-energy gain favoring the process may be attributed primarily to a disruption of the cage-like structure around the hydrophobic particle by the chaotropic cosolvent, resulting in preferential binding. The effect is associated with a decrease in the average number of intact hydrogen bonds, rather than with a change in hydrogen-bond strength. In the case of preferential hydration, the ordered arrangement of water molecules in the solvation shell is energetically favorable, and the kosmotropic cosolvent is excluded from the water-solvent interface. However, despite its importance for molecular biologists, the physical mechanisms underlying the interactions between chaotropic cosolvents and hydrophobic solutes in aqueous
Introduction

Solutions of Amphiphilic Molecules

Figure 1.4: Schematic representation of solvent and cosolvent distribution around a solute particle. The solvation shell of the solute particle is shown in yellow. Preferential binding occurs for chaotropic cosolvents (left), leading to an increase in solubility of the solute particle, whereas preferential exclusion is found for kosmotropic cosolvents (right), resulting in decreased solubility and the stabilization of aggregates.

solutions remain controversial.

In this study we adapt an MLG-type model for non-polar solute particles in water to include cosolvent effects, to capture both the stabilizing effects of kosmotropic cosolvents and their preferential exclusion from the hydration shell, and the opposing aggregate-destabilizing properties of chaotropic cosolvents with the corresponding preferential binding to the solute particle.

1.4 Solutions of Amphiphilic Molecules

The solubility of a hydrophobic compound in water contrasts sharply with that of strongly polar species. If both characteristics occur simultaneously in the same molecule, a rich variety of solvation properties may emerge. The polar parts tend to favor contact with similarly polar water molecules, whereas the hydrophobic parts avoid the proximity of the aqueous solvent. Substances which are characterized by two distinct parts, one polar and one hydrophobic, are generally referred to as amphiphiles ('amphi-philic' = loving both). Typical amphiphilic molecules are
composed of a polar or ionic group, which is usually called the head, and one or more long hydrocarbon chains called the tail. Fig. 1.5 shows several representations of a typical amphiphilic molecule. Solvent-solute interactions are relatively short-ranged, and the total interaction of an amphiphilic molecule with an aqueous solvent may therefore be divided into the sum of nearly independent contributions from the head and the tail. The ordering of water molecules around the polar head results in a strong attraction due to hydrogen bonds, whereas the contribution of one CH₂ group in the tail is relatively weak, because the only interaction in which it may be involved is of the Van-der-Waals type. With increasing tail-length, however, the repulsion between water and the hydrophobic tail becomes important, and the solubility of the solute particle decreases.

Figure 1.5: Representations of the amphiphilic molecule sodium lauryl sulfate. a) cross-section of a space-filling model, b) chemical structure, c) simplified picture of the relative size of head and tail.

In an aqueous solution, the hydrophobic parts of amphiphilic molecules tend to separate themselves from water molecules by forming aggregates (organic phase), such as micelles and microemulsion droplets. The simplest possible structure occurs if the polar and hydrophobic parts of the amphiphilic molecule are well separated into head and tail regions. Such amphiphiles are generally referred to as surfactants. For this type of molecule, micelles are formed consisting of a hydrophobic core containing all the tails (Fig. 1.6). The surface of a micelle is formed by the polar heads of the amphiphiles, which shield the hydrophobic tails from contact with water, and which maintain the high solubility of the entire micelle. Micelles may occur with various shapes and sizes, including spheres, ellipsoids, long cylinders, and bilayers (Fig. 1.7). The latter are formed by two parallel layers of amphiphilic molecules with their hydrophobic tails directed to the inside of the bilayer. Such bilayer micelles
Introduction

Solutions of Amphiphilic Molecules often form spherical vesicles, producing a solvent-filled cavity, and as such form the basic constituents of biological membranes around organelles and living cells. They are therefore essential for life in general allowing cells to maintain physiological conditions on the inside in varying environments by separating the content of the cell from its surroundings [4].

Single layers may occur in mixed solutions containing water, surfactants, and completely hydrophobic particles. In that case, small droplets (oil bodies) appear, and the surfactants form a shield between the purely hydrophobic core and the surrounding water in which the droplets are dissolved. Diverse organisms such as plant seeds, pollens, spores, and yeast store lipids in such intracellular oil bodies as food reserves which can be mobilized during an active period of metabolism [50]. Such oil bodies are very resistant to changes in external conditions, and can therefore protect the lipid over long periods of time. Some organisms, including spores and many plant seeds, are thus capable even of surviving complete dehydration. Single layers of surfactants occur also at water surfaces, thus reducing the surface tension of the liquid.

Figure 1.6: Schematic representation of a small, spherical micelle in water.

Different local structures and phases may be formed by surfactants depending on the water content of the solution, as shown in Fig. 1.7. At very high amphiphile concentrations, inverted micelles are formed where small amounts of water are enclosed in spherical regions.

Amphiphilic molecules whose polar and hydrophobic regions are distributed over the entire molecule, rather than being clearly separated, are unable to form well-defined, strictly organized micelles. Instead, they may aggregate to form assemblies which minimize the hydrophobic area per molecule exposed to the aqueous
Introduction

Hydrophobicity in Protein Folding

Figure 1.7: Schematic representation of different phases formed by surfactants in water. At very low surfactant concentration small micelles are formed, while at very high concentrations an inverse micellar solution is found.

phase under geometrical conditions determined by the nature of the amphiphiles. Poly(N-isopropylacrylamide), which belongs to this class of polymers, exhibits a phase transition at the LCST from a homogeneous solution, where the polymers are completely soluble, to a system of two separated phases [51]. At the UCST the organic phase disaggregates, and above this temperature the amphiphilic molecules are again soluble due to entropic effects [52].

The UCST and LCST depend on the net interaction resulting from the forces between hydrophobic and polar parts of the amphiphilic molecules, and on the interactions of these regions with water. One means of probing the nature of these interactions would be by systematic alteration of the polarity of the amphiphilic polymers in solution. Replacement of hydrophobic monomers by polar ones within the polymer chain may therefore lead to changes in the phase diagram. Many studies have been carried out to describe the various aggregation phases of amphiphilic molecules in aqueous solutions [55, 56, 4, 57]. However, the mechanism underlying the process of self-aggregation, and the dependence of the aggregation phase diagram on the distribution of polar groups in amphiphiles, has received far less attention. We will investigate the aggregation of amphiphilic molecules in aqueous solution by adapting the two-state MLG-derived model for water in which the hydrophobic interaction is included implicitly.

1.5 Hydrophobicity in Protein Folding

Proteins are the major functional molecules of life, being involved in most processes in every living organism. Through selective pressure, they have evolved to perform specific functions, which depend upon their three-dimensional, geometrical structures, known as *folds*. Changes in these structures can cause them to lose their
Introduction Hydrophobicity in Protein Folding

initial function. Indeed, in order to become biologically functional, proteins in a cell must fold into their unique, native configuration. However, small mutations of amino-acids (the monomers of proteins, also known as residues) can lead to misfolding which results in improper function and possible aggregation of the proteins, thus giving rise to a variety of conditions including Alzheimer’s disease, Parkinson’s disease, and Creutzfeld-Jacob disease, as well as sickle-cell anemia and other prion diseases. These conformational diseases consist typically of significant variations in size and shape compared to the native conformation, which can cause aggregation and deposition of the affected proteins. In order to understand the origin of these diseases and the biological function of proteins, the exact connection between the amino-acid sequence of a protein and its three-dimensional structure (and thus its function) has been investigated by many research groups in recent decades. However, the problem of predicting a protein fold on the basis of its sequence is still unsolved, and remains one of the most persistent challenges in modern molecular biology [64].

The difficulty of the protein folding problem is the fact that a protein is a long polymer chain formed by 20 different amino-acids, which can be combined to an extremely large number of different proteins, each consisting typically of hundreds of monomers. One way of characterizing the 20 amino-acids may be according to their hydrophobicity, which is thought to represent the essential property determining whether an amino-acid is more likely to be found in the core or on the surface of a folded protein. Because the hydrophobicity is believed to represent the main force driving the folding of proteins [10], certain models, also usually referred to as HP models, classify the amino-acids into hydrophobic (H) and hydrophilic (P for polar), sometimes including intermediates. To compare the resulting observations with experimental data, the latter must also be analyzed by classifying the amino-acids into a corresponding number of hydrophobicity groups. Commonly used hydrophobicity scales for amino-acids frequently disagree with each other, and often fail to predict the degree of surface exposure of a residue in the folded protein structure.

We consider a classification of the amino-acids according to their surface exposure in all known protein structures. A comparison with known hydrophobicity scales indicates that amino-acids must be considered in their protein environment to provide an accurate description of their tendency to occur at the surface of folded proteins. The new scale contains all possible forces, because for its determination no assumptions are required concerning the nature of these forces or the mechanism of the folding process. We analyze the accuracy of the new scale in the description of secondary-structure elements. Calculations using an off-lattice HP model in three dimensions allows us to test the features of the scale. Because of its model-independence, it represents a good measure of the degree to which an amino-acid is driven to the surface of native proteins, and may lead to an improved agreement between models and experiments.
1.6 Overview

In this study, we investigate the hydrophobic effect in aqueous solutions, using a two-state model for water which includes implicitly the hydrophobic interaction. The model focuses on the ability of water to form structurally ordered arrangements around solute molecules, and the formation of two distinct types of solvent, namely bulk and shell water. Chapt. 2 introduces a detailed description of the MLG model, and its adaptations to incorporate the phenomena discussed above. Also presented here are the different techniques used for our analysis, which include mean-field approximations, Monte Carlo simulations, and a pair approximation using the cluster-variation method. In Chapt. 3 we analyze the properties of hydrophobic particles in water, and determine the aggregation phase diagram including the upper and lower critical solution temperatures. We extend the model to describe the solvation properties of hydrophobic particles in a solution of water and a chaotropic cosolvent, and the destabilizing effect of such substances on aggregates of non-polar particles is investigated in Chapt. 4. The decrease in solubility of hydrophobic particles in water and the corresponding stabilizing effect on aggregates in the presence of kosmotropic substances is discussed in Chapt. 5. In Chapt. 6 we study solutions of amphiphilic molecules in water within two models, a simplified HP model and a much more refined treatment of surfactant molecules, to characterize the aggregation of such molecules as a function of the distribution of their hydrophobic regions. This enables us to analyze the formation of micelles from surfactants in aqueous solutions. Finally, in Chapt. 7 we consider the role of hydrophobicity in determining protein structure, by studying the correspondence between hydrophobicity and the surface exposure of amino-acids in proteins. Chapt. 8 contains a summary and our conclusions.
Chapter 2

Model and Methods

In order to describe the interaction between hydrophobic solutes and water, a simple model which describes the essential physics of the system is required. It is generally believed that the driving force in the aggregation process is the effective repulsive hydrophobic interaction between the polar water and non-polar solute [10]. As early as 1945, Frank and Evans [8] noted that this interaction arises from a cage-like arrangement of water molecules around the solute, which allows them to optimize their mutual hydrogen bonding and thus to minimize their energy with respect to bulk liquid water. Entropically, however, this strict ordering is unfavorable compared with the disordered configurations in bulk water.

The minimal model which contains the essential features of the physics of water as an aqueous solvent is the bimodal description of Muller, Lee, and Graziano (MLG) [22, 23]. The model describes liquid water by dividing it into two different populations based on the number of hydrogen bonds formed. Water molecules which are highly hydrogen-bonded to their neighbors have fewer rotational degrees of freedom and thus a lower multiplicity of degenerate configurations (lower entropy) than unbonded molecules. These are therefore denoted as “ordered”, while water molecules with many broken hydrogen bonds are considered to be “disordered”. The presence of a non-polar solute alters the enthalpy and entropy of water molecules in the solvation shell of each solute particle, and so a further distinction is required between “shell” and “bulk” water [9]. More details of the physical processes underlying these considerations are provided below. This simple model has been justified recently from molecular simulations of water [65].

In this form the MLG model reproduces correctly the ordering and strengthening of the hydrogen bonds in the first solvation shell of an added non-polar solute molecule at low temperatures, as well as the opposite behavior at high temperatures. It has been shown that the model provides an adequate description of the heat-capacity anomaly [66], and it has been extended to reproduce consistently the important properties of protein solutions, including warm and cold denaturation [67, 68].
2.1 Energy Levels

In this analysis we use an adapted version of the MLG model. An appropriate description of the solvent in the vicinity of hydrophobic solute particles which may be much larger than individual water molecules is obtained by allowing each site of a lattice representing the system to be occupied by a group of water molecules. The bimodal nature of the MLG model is preserved in this coarse-grained version by specifying only two types of water cluster at each site, where now an ordered site is characterized by having most of the hydrogen bonds among the molecules in the cluster intact, while a disordered site is understood to have a number (but by no means all) of these bonds broken. While this is the simplest possible approximation to the continuous distribution of intact or broken bonds within a site, we will demonstrate that it retains the capability of describing all the primary physical properties observed in aqueous solutions of non-polar solutes. The energy and degeneracy parameters for the coarse-grained model are determined by the same processes as in the bare MLG description outlined above, which we now discuss in more detail.

An approximation in which a group of water molecules is considered as one entity is justified when the non-polar solute particle is relatively large compared with a single water molecule. In this case the formation of a complete cage around the solute particle is rather improbable because it must be formed rapidly in the presence of local thermal fluctuations, and may even be prevented sterically [69]. Partial cages may therefore be formed in the vicinity of a solute particle, rather than one complete cage. In addition, formation of a hydrogen bond promotes the formation of further hydrogen bonds, which are stronger than before due to the change in charge distribution on forming the first bond [70]. This mutual reinforcement is known as cooperativity of hydrogen bonds, and leads to the formation of chains or clusters of hydrogen-bonded water molecules, whose extent depends on the size and the shape of the solute particles.

Water molecules which participate in hydrogen-bonded clusters have a higher degree of order and fewer rotational degrees of freedom than those in regions with many unbonded molecules [69]. The number of possible configurations of such an ordered cluster is thus significantly smaller than that of a disordered group of water molecules for both shell and bulk water. For steric reasons, fewer hydrogen-bonded water configurations are possible around a non-polar solute particle which is unable to form hydrogen bonds. These shell water molecules are forced into a tangential orientation [71], whereas the molecules in bulk water may also form radially oriented hydrogen bonds with central water molecules replacing the solute [66]. The degeneracy, or total number of configurations, of a hydrogen-bonded cluster of shell water (ordered shell) is consequently smaller than that of a hydrogen-bonded cluster of bulk water (ordered bulk).

In contrast, fewer orientational configurations exist for unbonded water molecules
Model and Methods Energy Levels

in the bulk than next to a non-polar solute particle [66]. The geometrical reason for this result is that for a shell site no hydrogen bonds are possible in direction of the solute particle, unlike the bulk situation obtained on replacing the solute particle by water. All orientations in which water molecules form radial hydrogen bonds with the central water in the bulk case (contributing to the ordered bulk degeneracy) are therefore transformed into configurations with many broken hydrogen bonds when the central water is replaced by a non-polar solute particle. We take such sites to be disordered in the bimodal sense discussed above. The degeneracy of a group of water molecules with many broken hydrogen bonds is then higher in the shell (disordered shell) than in the bulk (disordered bulk).

In summary, these considerations lead to a distribution of the total number of states \( q \) according to the sequence \( q_{ds} > q_{db} > q_{ob} > q_{os} \). We emphasize that this ordering of degeneracies is crucial for the qualitative behavior of the system. In combination with the corresponding sequence of energy levels (below) it determines the existence of both aggregation and disaggregation transitions, and thus of the coexistence curve. Because many fewer configurations have intact than broken hydrogen bonds, the difference in degeneracy between ordered and disordered states is much higher than that between shell and bulk states for both types of site. In fact the difference in degeneracy between shell and bulk states depends primarily on the number of possible radial hydrogen bonds, which is much smaller than the total number. In much of what follows we employ the degeneracy factors \( q_{ds} = 49, q_{db} = 40, q_{ob} = 10, \) and \( q_{os} = 1 \).

These relative values are chosen to be qualitatively representative of the degeneracies expected for a system of small solute particles in water, with both experimental and theoretical [66] justification based on the above properties of water. We note here that, while it is known that a number of ordered states exists even for the cage-like water structures coordinating dissolved solute particles, the absolute multiplicities of the \( q \) factors contribute only an additive constant to the free energy and are irrelevant for the phenomena to be discussed below; thus we set the lowest degeneracy to \( q_{os} = 1 \).

Fig. 2.1 shows the schematic energy levels of a group of water molecules in the coarse-grained MLG-type model. The hydrogen-bond energy is optimized in the ordered, cage-like shell structure where strong, tangentially oriented hydrogen bonds are formed among water molecules in the cluster (ordered shell), and hence \( E_{os} \) is lowest. Direct experimental evidence for the tangential orientation of hydrogen bonds between water molecules in the solvation shell of non-polar solute particles has been provided recently in Refs. [14, 71]. When the solute particle is replaced by water, radial hydrogen bonds may be formed. However, clusters including radially oriented water molecules have on average a higher hydrogen-bond energy than those whose bonds are predominantly tangentially oriented, because for steric reasons a radial hydrogen bond precludes another good tangential hydrogen bond to a first-shell neighbor. This result was demonstrated in a model including oriented hydrogen bonds between water molecules [66]. In bulk water, both configurations are possible, and
Figure 2.1: Energy levels in the MLG model for water. The levels are arranged according to $E_{ds} > E_{db} > E_{ob} > E_{os}$ (see text). The configurations of the different energy levels at the molecular level are illustrated schematically.

thus the average energy in a cluster of hydrogen-bonded bulk water $E_{ob}$ (ordered bulk) is higher than $E_{os}$.

The energies $E_{ds}$ (disordered shell) and $E_{db}$ (disordered bulk) are relatively much higher than those of the respective ordered states, because breaking of hydrogen bonds is required in forming the disordered states. The average hydrogen-bond energy of a group of disordered shell water molecules decreases when the solute particle is replaced by water, because some radial hydrogen bonds broken by the solute may be formed, lowering the average energy of the group and ensuring that $E_{db} < E_{ds}$.

Specific determination of the energy values for a selected binary system would require structural calculations and molecular dynamics simulations [67, 72, 73, 74]. However, such refinement is not necessary for the general phenomena to be illustrated in the chapters to follow. We note here that the temperature scale $\beta^{-1}$ is defined by the energy scale. We use the parameter values $E_{ds} = 1.8$, $E_{db} = 1.0$, $E_{ob} = -1.0$, and $E_{os} = -2.0$, which are thought to be quite generic for aqueous systems, and furthermore agree closely with the energies used in a successful description of the thermodynamic behavior of biopolymers in water [75]. While the results of the calculations to follow are critically dependent on the sequence (Fig. 2.1) of the energy parameters, they are not particularly sensitive to the exact values of the differences between these parameters, and are of course independent of their absolute values. However, it remains possible to refine these parameters by comparison with
experiment to obtain good qualitative agreement with measured quantities for different solutions.

We note here that the coarse-grained model presents a natural extension of the well-established MLG framework, and as such its success is not entirely surprising. In the original model, hydrogen bonds are divided into two different populations, namely intact and broken, with energy levels $E_i$ and $E_{br}$, and degeneracies $q_i$ and $q_{br}$ ($q_i \ll q_{br}$). When considering sites containing a number of water molecules, the total energy $E_{n_0}(k)$ of a cluster which may form a maximum of $n_0$ hydrogen bonds, of which $k$ are intact, is given by $E_{n_0}(k) = kE_i + (n_0 - k)E_{br}$. This energy distribution is accompanied by a distribution of the number of different cluster configurations, $q_k = \frac{n_0!}{k!(n_0-k)!} q_i^k q_{br}^{n_0-k}$, which from the relative values of the bond degeneracies is strongly skewed towards high energies. However, the free energy of the system, which is proportional to the logarithm of the site partition function

$$\ln Z_{n_0} = \ln \sum_{k=0}^{n_0} \frac{n_0!}{k!(n_0-k)!} q_i^k e^{-\beta k E_i} q_{br}^{n_0-k} e^{-\beta (n_0-k) E_{br}}$$

$$= n_0 \ln (q_i e^{-\beta E_i} + q_{br} e^{-\beta E_{br}}), \quad (2.1)$$

is fully characterized only by the single-bond energies, and simply gains an overall multiplicative factor of $n_0$ (the same applies to the effective interactions defined in Secs. 2.2 and 2.3), and thus corresponds to the free energy of $n_0$ independent hydrogen bonds. For this reason the bimodal treatment of the pure MLG model is completely justified also in the coarse-grained system. Further, it is not necessary to specify the exact number of water molecules at a site, and this enters only in the relative values of the energy and degeneracy parameters.

### 2.2 Coarse-Grained Model

On a cubic lattice the energy of a system of $N$ sites, occupied either by particles ($n_i = 0$) or by water ($n_i = 1$), is given by the Potts-like Hamiltonian

$$H[\{n_i\}, \{\sigma_i\}] = \sum_{i=1}^{N} n_i [(E_{os} \tilde{\delta}_{i,os} + E_{ds} \tilde{\delta}_{i,ds})(1 - \lambda_i) + (E_{ob} \tilde{\delta}_{i,ob} + E_{db} \tilde{\delta}_{i,db}) \lambda_i], \quad (2.2)$$

where $\lambda_i$ is 1 if site $i$ is surrounded only by water, is 0 otherwise, and is defined as the product of the nearest-neighbor factors, $\lambda_i = \prod_{(j,i)} n_j$. Each water site $i$ can be in one of the $q$ different states which are divided among the 4 energy levels shown in Fig. 2.1. Therefore, $\tilde{\delta}_{i,os}$ is 1 if site $i$ is occupied by water in one of the $q_{os}$ ordered shell states and 0 otherwise, and $\tilde{\delta}_{i,ds}$ is 1 if it is occupied by water in one of the $q_{ds}$ disordered shell states and 0 otherwise. Analogous considerations apply for the bulk states.
An important observation is that this model does not include interactions between different water sites. Hence it is valid as long as water is liquid, and it neglects long-range effects arising from extended hydrogen-bonded networks.

Hamiltonian (2.2) leads to the partition function

$$Z = \sum_{\{n_i\}, \{\sigma_i\}} e^{-\beta H[\{n_i\}, \{\sigma_i\}]}$$

(2.3)

where every term of the sum represents the statistical weight of the corresponding configuration ($\beta^{-1} = k_B T$). Performing the sum over the configurations of the states of each water site, including the number of states of the respective energy levels, gives

$$Z = \sum_{\{n_i\}} \prod_i e^{-\beta [S_n(\lambda_i) + Sn(1-\lambda_i)]}$$

(2.4)

where we define

$$S = -\frac{1}{\beta} \ln [q_{os} e^{-\beta E_{os}} + q_{ds} e^{-\beta E_{ds}}] \text{ and}$$

$$B = -\frac{1}{\beta} \ln [q_{ob} e^{-\beta E_{ob}} + q_{db} e^{-\beta E_{db}}].$$

(2.5)

(2.6)

The formal method for obtaining an effective Hamiltonian (and effective interactions) is to integrate over the degrees of freedom of the particles thought to be responsible for the interactions, which are the solvent molecules. The canonical partition function, i.e. the partition function for a fixed number of particles $N_p$ (whence the number of water sites is $N_w = N - N_p$), is

$$Z_{N_w} = \sum_{\{n_i\}} e^{-\beta H_{\text{eff}}[\{n_i\}]}$$

(2.7)

where the effective Hamiltonian $H_{\text{eff}}[\{n_i\}]$ is formally the free energy of the solvent at fixed particle configuration. In a system where the number of particles is not fixed, a chemical potential $\mu$ associated with the replacement of water sites by solute particles ($\mu$ is thus the difference between the chemical potentials of water and solute particles) is included, and the grand canonical partition function may be expressed as

$$\Xi = \sum_{N_w} e^{\beta \mu N_w} Z_{N_w} = \sum_{\{n_i\}} e^{-\beta H_{\text{eff}}^{gc}[\{n_i\}]}$$

where

$$H_{\text{eff}}^{gc}[\{n_i\}] = \sum_{i=1}^{N} [(S + \mu)n_i + (B - S)n_i \lambda_i].$$

(2.8)

$H_{\text{eff}}^{gc}[\{n_i\}]$ represents an effective Hamiltonian for single sites, and provides the first step in obtaining the effective interactions between particles.

This model, which focuses on the formation of two distinct populations of solvent and on changes in the structural arrangements of liquid water, contains the
necessary ingredients to describe the hydrophobic interaction, and may be expanded to include additional effects. In Chaps. 4 and 5 we adapt the model to include cosolvents by changing the number of disordered states of both shell and bulk water due to the presence of chaotropic and kosmotropic substances. Chaotropic cosolvents have the ability to reduce the number of intact hydrogen bonds between water molecules, and hence to decrease the number of possible structural arrangements of water, which reduces the number of ordered states and increases that of disordered states. This leads to a growth in solubility of hydrophobic particles and to a destabilization of their aggregates. In contrast, kosmotropic cosolvents enhance hydrogen bond formation in bulk water, thus increasing the number of ordered states with respect to the number of disordered states, leading to a suppressed solubility of solute particles with concomitant stabilization of aggregates. We show that changing the degeneracies of the different water states in the model is an appropriate means of reproducing such cosolvent effects on the solubility of hydrophobic particles.

The model may be extended to consider solutions of amphiphilic solute particles in water (Chapt. 6). We describe amphiphiles with varying distributions of polar and non-polar regions by representing the particles as cubes on which each face may be either polar or hydrophobic. The state of a neighboring water site is determined by whether it is facing a polar or a hydrophobic side. We show that there exists a rich variety of micellar structures exists depending on the polarity distribution of the amphiphilic particles in solution. In a further refinement of the approach, we describe surfactant molecules of varying length in water by distinguishing between the sides and the tip of their hydrophobic tail.

2.3 Hydrophobic Interaction

To describe the consequences for hydrophobic particles in the solution, we replace the number of water molecules $n_i$ by the number of particles $m_i$,

$$n_i = 1 - m_i,$$  \hspace{1cm} (2.9)

where $m_i$ is 1 if the site is occupied by a particle and 0 otherwise. With this substitution, $\lambda_i$ becomes the product over the nearest neighbors, $\lambda_i = \prod_{j \neq i} (1 - m_j)$, and takes the value 1 only if site $i$ is completely surrounded by water molecules, or 0 otherwise. Introducing an effective interaction $J_{\text{eff}} = (B - S)$ between particles, and the effective chemical potential $\mu_{\text{eff}} = (S + \mu)$, the effective Hamiltonian for the particles becomes

$$H_{\text{eff, p}}[\{m_i\}] = K - \sum_{i=1}^{N} [\mu_{\text{eff}} m_i + J_{\text{eff}} (m_i - 1) \lambda_i],$$  \hspace{1cm} (2.10)

where $K = N(S + \mu)$. In this formulation it can be seen that the interactions are not limited to two-body terms, but include many-body interactions through the last term.
Because $S$ is negative and decreases continuously, $\mu_{\text{eff}}$ decreases as temperature increases, and the solubility drops. If $\mu > -S$, $\mu_{\text{eff}}$ is positive at low temperatures and the solubility is high. In the ground state ($T = 0$), this results in a solid phase for $\mu > -S$ and pure water for $\mu$ large and negative.

At low temperatures, $J_{\text{eff}}$ is positive and therefore the interaction term is minimized by $\lambda_i = 1$, which means that there is a repulsive force between particles. At high temperatures, however, increasing entropy effects cause $J_{\text{eff}}$ to become negative, and the minimal interaction energy is obtained for $\lambda_i = 0$, resulting in an attractive force between particles. These different effective forces result from the interplay of entropic and enthalpic effects, and give rise to the complex properties of solutions of hydrophobic particles.

### 2.4 Mean-Field Calculation

The mean-field approximation is based on the assumption that spatial fluctuations of the order parameter are insignificant, and is very useful for qualitative predictions. In the mean-field treatment, the average occupancy of a site is simply the density, $\langle n_i \rangle = \rho$, which is taken to be constant throughout the system. (Because of the nature of the model, $\rho$ is connected with the number of water molecules, and therefore the particle density, which is the variable used in the graphs to follow, becomes $\rho_p = 1 - \rho$.)

The grand canonical mean-field free energy of a particle is given by

$$f = (B - S)\rho^{z+1} + (S + \mu)\rho + \beta^{-1}\rho\ln\rho + (1 - \rho)\ln(1 - \rho).$$

The density minimizing the free energy at a given temperature and chemical potential corresponds to the equilibrium concentration for these values.

The number $z$ of nearest neighbors of a site is defined by the lattice. In the fully aggregated phase a solute particle is in contact with $z$ other solute particles. Mean-field calculations, however, are independent of a lattice, and thus $z$ in Eq. (2.11) has the meaning only of an effective coordination number. A solute particle is on average in contact with $z_{\text{eff}}$ other solute particles, and it is this number which is important in determining the effective interparticle interaction. The critical particle density at the LCST and UCST, where the extent of the aggregated phase is maximal, will be calculated analytically in Sec. 3.1. The average number of solute particles interacting with each other depends on both solvent and solute species. $z_{\text{eff}}$ may be determined from experiment for different types of solute particle.
2.5 Molecular-Level Simulations

Because in the mean-field approach the local densities are replaced by their average value, the effects of fluctuations are completely neglected. These statistical fluctuations may be included by studying the system in more detail on the molecular level, for which an appropriate method is Monte Carlo simulation.

2.5.1 Monte Carlo Simulation

We perform Monte Carlo simulations in real space to probe the microscopic properties of a system described by the Hamiltonian in Eq. (2.8). We study a three-dimensional system of \( N = 27000 \) sites with a random initial distribution of particles and water molecules. Periodic boundary conditions are used to eliminate boundary effects. With regard to finite-size effects, we have found that our results are robust with respect to changes in the system size. Because the system has a large number of degrees of freedom, a representative sampling of the high-dimensional phase space is necessary to estimate thermal averages in the equilibrium state.

In a system of fixed density (i.e. in the canonical ensemble), every possible configuration \( \{n_i\} \) has the statistical weight

\[
w_c(\{n_i\}) = \frac{e^{-\beta H_n[\{n_i\}]}}{Z_N},
\]

where the partition function \( Z_N \) of a system of \( N \) particles is given by Eq. (2.7). At equilibrium the system must satisfy the detailed-balance condition

\[
w_c(\{n_i\})P_{n \rightarrow n'} = w_c(\{n_i'\})P_{n' \rightarrow n},
\]

where \( P_{n \rightarrow n'} \) is the transition rate from configuration \( \{n_i\} \) to a new one \( \{n_i'\} \). The relative probability to produce configuration \( \{n_i'\} \) from the previous one \( \{n_i\} \) thus becomes the ratio of the two weights,

\[
r = \frac{w_c(\{n_i'\})}{w_c(\{n_i\})} = e^{-\beta (H_n[\{n_i'\}] - H_n[\{n_i\}] )},
\]

and depends only on the difference in free energy between them. This transition probability is used in the Metropolis algorithm to generate new configurations from previous ones. Specifically, the new configuration \( \{n_i'\} \) is accepted if \( r > 1 \), or if \( r < 1 \) but larger than a random number uniformly distributed in the interval \([0, 1]\).

If the new configuration is completely different from the previous one, the acceptance probability is rather low. The method therefore sweeps randomly through the system considering configurations which differ from the previous one only by single-site exchanges of particles and water. After a number of thermalization sweeps,
during which the system relaxes towards equilibrium and no observables are calculated, the system is taken to be in equilibrium with only thermodynamic fluctuations present. Thermodynamic quantities are estimated by averaging over the configurations which are kept during a subsequent number of steps which is sufficiently large that a considerable portion of the total phase space is sampled. The decorrelation time of successive configurations in equilibrium is found to be lower than 10 Monte Carlo steps (one Monte Carlo step corresponds to the consideration of every site in the system once), both in the coexistence phase and in the homogeneous phase. Measurements are taken only every 50 Monte Carlo steps over a period of 500 000 steps after 100 000 initial relaxation steps. This process is repeated for 10 different random initial configurations and the observations are averaged over these independent simulations.

In the crystal phase (below) the decorrelation time of consecutive configurations is longer because of the very low temperature, and measurements are averaged over a larger number of independent simulations.

In order to find the equilibrium density for a fixed temperature and chemical potential, a grand canonical sampling (i.e. the number of particles is not constant) of the phase space is performed. The procedure is the same as in the canonical case, except that the weight of a configuration \( \{n_i\} \) in the grand canonical ensemble is

\[
\begin{equation}
 w_{gc}(\{n_i\}) = \frac{e^{-\beta H_{gc}^{ne}(\{n_i\})}}{\Xi},
\end{equation}
\]

where the grand canonical partition function \( \Xi \) is given by Eq. (2.8). This leads to a relative transition probability for configuration \( \{n'_i\} \) from a previous one \( \{n_i\} \) which depends on the difference in free energy of the two configurations

\[
\begin{equation}
 r = e^{-\beta H_{gc}^{ne}(\{n'_i\}) - H_{gc}^{ne}(\{n_i\})}. \tag{2.16}
\end{equation}
\]

The behavior of a system which is heated at constant density can be analyzed by first determining the equilibrium density using grand canonical sampling at a chosen starting temperature, and then raising the temperature continuously while applying a canonical sampling procedure in which the number of particles remains constant but particles and water may exchange sites. Most of the simulations to follow are grand canonical except where indicated otherwise.

### 2.6 Pair Approximation

The mean-field approximation to be discussed in detail in Sec. 3.1 provides confirmation of hydrophobic aggregation, and in Sec. 4.2 of the fact that chaotropic cosolvents increase the solubility of non-polar solutes, leading to disaggregation. However, no conclusions may be drawn concerning the molecular distribution in the ternary system, because local densities are replaced by their averages and spatial
fluctuations are thus neglected. The cluster-variation method offers an accurate approximation technique for the detailed study of lattice systems on the molecular level [76, 77, 78]. Phase diagrams for a variety of systems have been determined by this method, which is based on a variational approach. If the basic clusters under consideration consist of two neighboring lattice sites (pair approximation), it is equivalent to the Bethe approximation [79]. We have implemented this pair approximation for the solute-solvent-cosolvent system to obtain additional information about the local concentration of the cosolvent and about intersite correlations.

2.6.1 Cluster-Variation Method

The free energy per lattice site may be expanded as an infinite series, and the exact cluster expansion coefficients determined by the cluster variation method up to a certain order, which is given by the “maximal clusters” \( \nu_i \). These maximal clusters are chosen according to the interactions in the system, such that the Hamiltonian is expanded as

\[
H = \sum_{\alpha \in \Lambda} h_\alpha (n_\alpha, i \in \alpha),
\tag{2.17}
\]

where \( \alpha \) is a subset of \( n_\alpha \) lattice sites which is included in the set of maximal clusters \( \Lambda \), and \( h_\alpha = 0 \) if \( \alpha \not\in \Lambda \). In this sum, the terms \( h_\alpha (n_\alpha, i \in \alpha) \) denote the interactions between the sites \( i \) of cluster \( \alpha \). The set \( \Lambda \) is defined to contain all subclusters of \( \alpha \) if \( \alpha \) belongs to it. In addition, the maximal clusters \( \nu_i \), which define the set \( \Lambda \) uniquely, are not included in any other cluster belonging to the set.

In the canonical ensemble, the density operator \( \hat{\rho}_\Lambda \) in the set \( \Lambda \) can be written in an arbitrary basis as

\[
\hat{\rho}_\Lambda = \frac{e^{-\beta H}}{Tr[e^{-\beta H}]},
\tag{2.18}
\]

because the canonical partition function \( Z \) is given by

\[
Z = \sum_n e^{-\beta E_n} = Tr[e^{-\beta H}],
\]

where \( E_n \) are the eigenvalues of the Hamiltonian. Knowledge of the diagonal, normalized \( (Tr[\hat{\rho}_\Lambda] = 1) \) density operator allows the determination of any observable, whose average is given by

\[
\langle A \rangle = Tr[\hat{\rho}_\Lambda A] = \frac{Tr[e^{-\beta H} A]}{Tr[e^{-\beta H}]}.
\tag{2.19}
\]

Thus, for a chosen configuration \( \{n_i\} \), the internal energy \( U \) is

\[
U = Tr[\hat{\rho}_\Lambda H[\{n_i\}]] = \frac{Tr[e^{-\beta H} H[\{n_i\}]]}{Tr[e^{-\beta H}]},
\tag{2.20}
\]

and the entropy takes the form

\[
S = \langle -k_B \ln \hat{\rho}_\Lambda \rangle = -k_B Tr[\hat{\rho}_\Lambda \ln \hat{\rho}_\Lambda].
\tag{2.21}
\]
Finally, the free energy can be obtained from these expressions as

\[ F[\rho_\Lambda] = Tr[\hat{\rho}_\Lambda H] + \beta^{-1}Tr[\hat{\rho}_\Lambda \ln \hat{\rho}_\Lambda]. \tag{2.22} \]

The reduced density operator for a cluster \( \alpha \in \Lambda \) is defined as the partial trace over all clusters in \( \Lambda \) excluding \( \alpha \),

\[ \rho_\alpha = Tr_{\Lambda\setminus\alpha}[\rho_\Lambda]. \tag{2.23} \]

The same holds for all subclusters \( \beta \) of cluster \( \alpha \), and \( \rho_\beta \) may be written as \( \rho_\beta = Tr_{\alpha\setminus\beta}[\rho_\alpha] \). In particular, it follows from the normalization of the density matrix that \( Tr_\alpha[\rho_\alpha] = 1 \). Thus the entropy of a cluster \( \alpha \) may be defined as

\[ S_\alpha = -k_B Tr[\hat{\rho}_\alpha \ln \hat{\rho}_\alpha], \tag{2.24} \]

and may further be expanded as a cumulant sum over all the subclusters \( \beta \) of cluster \( \alpha \),

\[ S_\alpha = \sum_{\beta \subseteq \alpha} \tilde{S}_\beta. \tag{2.25} \]

The cumulants \( \tilde{S}_\beta \) are defined using a Möbius inversion [80]. In fact, in a partially ordered set \( \Lambda \), its elements can be ordered using the relation \( \subseteq \), and the Möbius function \( \nu \) of this partially ordered set is then defined through the relation

\[ \sum_{\alpha \subseteq \beta \subseteq \gamma} \zeta(\alpha, \beta) \nu(\beta, \gamma) = \delta(\alpha, \gamma), \tag{2.26} \]

where \( \delta \) is the Kronecker function and \( \zeta \) is one if \( \beta \subseteq \alpha \) or zero otherwise. The Möbius function is defined uniquely by Eq. (2.26). On the lattice, each cluster \( \alpha \) contains \( n_\alpha \) sites, and the clusters may be partially ordered using the relation \( \beta \subseteq \alpha \), if \( \beta \) is contained in \( \alpha \). Hence, the set \( \Lambda \) is by definition partially ordered, and the Möbius function is given by [80]

\[ \nu(\alpha, \beta) = (-1)^{n_\beta - n_\alpha}. \tag{2.27} \]

Applying this Möbius inversion to the cumulants of the cluster entropy given in Eq. (2.25) yields

\[ \tilde{S}_\beta = \sum_{\alpha \subseteq \beta} (-1)^{n_\beta - n_\alpha} S_\alpha. \tag{2.28} \]

The total entropy of the system may be expressed as a sum of all subcluster entropies, which are known. Because the Hamiltonian is limited to short-range interactions, clusters larger than the correlation length may be neglected. The total entropy of the system may therefore be approximated by the sum of the cluster entropies of the set \( \Lambda \),

\[ S = \sum_{\alpha \in \Lambda} \tilde{S}_\alpha = \sum_{\alpha \in \Lambda} \sum_{\beta \subseteq \alpha} (-1)^{n_\alpha - n_\beta} S_\beta \]

\[ = \sum_{\alpha \in \Lambda} \sum_{\beta} (-1)^{n_\alpha - n_\beta} \zeta(\alpha, \beta) S_\beta = \sum_{\beta \in \Lambda} a_\beta S_\beta, \tag{2.29} \]
where $a_{\beta} = \sum_{\beta \in \Lambda} (-1)^{n_{\alpha} - n_{\beta}} \zeta(\alpha, \beta)$. As can be seen from Eq. (2.26), the sum over all clusters $\beta$ which contain the subclusters $\alpha$ and are members of the set $\Lambda$ is

$$\sum_{\beta \subseteq \alpha \in \Lambda} a_{\beta} = 1. \quad (2.30)$$

Finally, the total free energy of the system becomes

$$F[\rho_{\alpha}] = \sum_{\alpha \in \Lambda} Tr[\rho_{\alpha} H_{\alpha} + a_{\alpha} \rho_{\alpha} \ln \rho_{\alpha}]. \quad (2.31)$$

The free energy per lattice site can be expanded as an infinite series, and the exact cluster expansion coefficients determined by the cluster-variation method up to a certain order as discussed above. Because we are most concerned with nearest-neighbor correlations, and discard the effects of long-range order in the system, we neglect terms of third and higher order to focus on clusters within the ternary system which include only nearest-neighbor pairs.
Chapter 3

Solution of Hydrophobic Particles in Water

We begin by considering solutions of simple hydrophobic particles in pure water. As discussed in Chapt. 1, the anomalous properties of aqueous solutions of hydrophobic molecules, and the hydrophobic effect itself, are believed to be a consequence of changes in the structure of water, rather than being explained by water-solute interactions. Pure liquid water is highly ordered due to strong intermolecular hydrogen bonds. Addition of a non-polar solute unable to form hydrogen bonds causes some of these bonds to break, leading to an increase in enthalpy and a considerable gain in entropy due to the disruption of the structural arrangement. However, at low temperatures water molecules rearrange in a new, cage-like structure, resulting in a recovery of the lost hydrogen bonds which are even slightly stronger than before. Although the new arrangement is enthalpically favorable, it causes a decrease in entropy as a consequence of the increase in local order. Measurements of the changes in entropy and enthalpy for different solutes suggest that this local ordering of water molecules around the solute is not unique, but that a number of different organizations is possible. At higher temperatures hydrogen bonds are broken due to thermal agitation, and the ordered structure around solute molecules is disrupted. The solute particles aggregate to minimize their total exposed surface area. At even higher temperatures entropy effects become dominant and the particles are again soluble. In the following, we will analyze this behavior on a molecular level.

3.1 Mean-Field Calculation

On a three-dimensional square lattice every site has $z = 6$ nearest neighbors, which can be occupied either by a particle or by water.

As described in Sec. 2.4, the mean occupancy of a site is simply the density, $\langle n_i \rangle = \rho$, which is taken to be constant throughout the system. This approximation leads to
the grand canonical mean-field free energy per site

\[ f = (B - S)\rho^{\tau+1} + (S + \mu)\rho + \beta^{-1}[\rho \ln \rho + (1 - \rho) \ln(1 - \rho)]. \tag{3.1} \]

Minimization of this free-energy density gives the mean particle density of the equilibrium configuration. Within the mean-field framework, \( z \) may be interpreted as an effective coordination number \( z^{\text{eff}} \), as described in Sec. 2.4.

The solubility of non-polar solute particles in water may be quantified by calculating the inverse solubility product \( K_{\text{sp}}^{-1} \), which is the equilibrium constant for the dissolution reaction starting with aggregates of non-polar solute particles in pure water. The choice of sign for \( \Delta F \) dictates that the solubility product is defined by

\[ K_{\text{sp}} = e^{-\frac{1}{T} \Delta F} = e^{-\frac{1}{T} \Delta U + \frac{k}{T} \Delta S}, \tag{3.2} \]

where \( R \) is the universal gas constant. Because of the exponential nature of Eq. (3.2), if \( K_{\text{sp}} < 1 \) the equilibrium lies deep in the aggregation regime, resulting in extensive aggregation and little remaining dissolved solute. Conversely, if \( K_{\text{sp}} \) exceeds 1 the equilibrium state is far in the dissolution regime, solution of the solute is almost complete, and aggregation is effectively absent. Consistent with the definition of \( \beta^{-1} \), for the calculations to follow we set \( R \) to 1.

The thermodynamic properties of the system, and specifically the propensity towards aggregate formation, may be illustrated as follows. The free energy per particle of a homogeneously mixed system with a given solute density is compared to that of one consisting of pure water (with the same number of water and particle sites) in contact with but entirely separated from the particles, which form a solid phase. As shown in Fig. 3.1 for systems of density \( \rho_p = 0.1 \) (and fixed chemical potential \( \mu = 0 \)), the homogeneous mixture is energetically more favorable at low and at high temperature, indicating good solubility of the solute in water. However, at intermediate temperatures the separated phase is preferred, and in this regime one would expect solute aggregation. Two phase transitions then occur, at temperatures \( T_L \) and \( T_U \), where the differences in enthalpy and in entropy (multiplied by \( T \)) are equal. This behavior is also illustrated by the solubility product \( K_{\text{sp}} \), which is smaller than 1 between \( T_L \) and \( T_U \), and exceeds 1 below \( T_L \) and above \( T_U \), where the solubility of the non-polar solute particles is thus high.

In this simple example, where the system is forced into a homogeneous state and \( \mu = 0 \) at constant \( \rho_p \), the resulting free energy is larger than that of the equilibrium state obtained by minimizing \( f(\rho) \) at a finite value of \( \mu \) chosen to produce the same density \( \rho_p \). In addition, the comparison neglects the small water-solute surface of the fully separated phase, as only the free energy of the water sites is considered. Thus the immiscible region for the true mean-field equilibrium states would be larger than that in Fig. 3.1. All of calculations to follow are performed by minimizing \( f(\rho) \) in Eq. (3.1), which provides the best approximation available at the mean-field level.
Figure 3.1: Differences in free energy $F$, enthalpy $U$, and entropy $S$ (multiplied by $T$) per particle between a completely mixed system of particle density $\rho_p = 0.1$ and pure water (completely separated from solute particles). The system with density $\rho$ is forced into a homogeneous state in which the particles are in solution. Between the critical temperatures $T_L$ and $T_U$, pure water is energetically favorable and the system separates into two phases. This aggregation is characterized by low solubility through a solubility product $K_{sp}$ which is smaller than 1. At temperatures below $T_L$ and above $T_U$, the three components are miscible, and the solubility product exceeds 1.

Fig. 3.2 represents the form of the free-energy density as a function of the particle density at temperatures near the UCST. At a temperature $T_L < T_t < T_U$ the free energy is minimal for the lower and upper coexistence densities $\rho_{c1}$ and $\rho_{c2}$ indicated by the $\circ$ symbols. Heating the system from below the transition temperature at fixed chemical potential $\mu$ (corresponding to a vertical line in Fig. 3.5) results in a discontinuous jump in density from $\rho_{c1}$ to $\rho_{c2}$ at $T_t$.

Based on these equilibrium values, a phase diagram as a function of $\rho$ and $T$ may be obtained as a function of temperature, as shown in Fig. 3.3. The outer line represents the closed-loop coexistence curve $T_{co}(\rho)$, outside which the system is in a homogeneous state. Because we find aggregation by heating at low temperatures, and at high temperatures the entropy of solvation should dominate, a closed-loop solubility curve showing a LCST and an UCST [20, 81, 82] is to be expected.
Figure 3.2: Schematic representation of the free-energy density of an aqueous solution of hydrophobic particles as a function of the density $\rho$ near the critical point for temperatures $T_1$ close to $T_U$. Circles indicate the equilibrium values and stars show the inflection points.

The inner line is the spinodal curve $T_{sp}(\rho)$, which is important when quenching a system of constant density $\rho_0$ from the homogeneous phase ($T_0 > T_U$ or heating from $T_0 < T_L$) into the coexistence region. The spinodal curve is given by the inflection points of the free energy, indicated by symbols in Fig. 3.2 and determines the transition from a metastable region (between $T_{co}(\rho)$ and $T_{sp}(\rho)$) with respect to phase separation to an unstable region (inside $T_{sp}(\rho)$). For quenches into the metastable region ($T_1$ in Fig. 3.3), the free energy has two local minima separated by a barrier $\Delta f$, while quenches into the spinodal region ($T_2$) render the system globally unstable, because the free-energy barrier disappears and no longer prevents the system from attaining the global minimum (see Fig. 3.4).

Microscopically, the process of phase separation is different in the two regions, as can be seen using Monte Carlo simulations. In both cases, at long times after the quench, the system will be completely separated into two phases (of densities $\rho_{c1}$ and $\rho_{c2}$) in a ratio depending on the initial density $\rho_0$. The fraction of the total volume occupied by phase $i$ is $V_i = \frac{\rho_i - \rho_j}{\rho_i - \rho_j}$, where $j$ denotes the other phase. In the spinodal region the system aggregates spontaneously in a process known as spinodal decomposition, whereas in the metastable region small droplets of solute particles...
Figure 3.3: $\rho$-$T$ phase diagram for an aqueous solution of non-polar solute particles obtained by mean-field calculation. The outer line represents the closed-loop coexistence curve, while the inner line marks the spinodal curve. The arrows of decreasing $T$ demarcate quenches into the metastable ($T_1$) and spinodal ($T_2$) regions (see text).

“evaporate” before condensing to larger, growing nucleation domains [83].

The mean-field $\mu$-$T$ phase diagram in Fig. 3.5 shows a first-order transition line bounded by two critical points characterized by the same critical solvent density $\rho^*$. Analytically, $\rho^*$ may be obtained by imposing that the two local minima of the free energy $f$, as well as the inflection points of $f$, coincide at the critical points $\rho^*, T^*$ and $\mu^*$. The first, second and third derivatives of the free energy with respect to the density must therefore vanish at the critical points. We calculate simultaneously

$$\frac{\partial f}{\partial \rho} = (z + 1)(B - S)\rho^z + (S + \mu) + \beta^{-1}\ln\left(\frac{\rho}{1 - \rho}\right) = 0,$$  \hspace{1cm} (3.3)

$$\frac{\partial^2 f}{\partial^2 \rho} = z(z + 1)(B - S)\rho^{z-1} + \beta^{-1}\frac{1}{\rho(1 - \rho)} = 0, \text{ and}$$  \hspace{1cm} (3.4)

$$\frac{\partial^3 f}{\partial^3 \rho} = z(z + 1)(z - 1)(B - S)\rho^{z-2} + \beta^{-1}\left(\frac{1}{(1 - \rho)^2} - \frac{1}{\rho^2}\right) = 0,$$  \hspace{1cm} (3.5)

and simplify Eq. (3.4) to the form

$$\beta^{-1} = -z(z + 1)(1 - \rho)(B - S)\rho^z.$$  \hspace{1cm} (3.6)
Figure 3.4: Free-energy density for $T$ near $T_i$ at a fixed chemical potential $\mu = 3.0$. Filled circles indicate the stable minima for different temperatures, while open circles indicate metastable minima which occur for $T_{sp} < T < T_i$.

Introducing Eq. (3.6) into Eq. (3.5) leads to the critical density $\rho^* = \frac{z}{z+1}$ (i.e. critical particle density $\rho^*_p = \frac{1}{z+1}$), which depends only on the effective coordination number $z$. Inserting $\rho^*$ into Eq. (3.6) provides the LCST $T^*_L = 0.518$ and the UCST $T^*_U = 2.79$ for the parameter values chosen. Finally, Eq. (3.3) gives the corresponding lower and upper critical chemical potentials $\mu^*_L = 1.69$ and $\mu^*_U = 7.53$. These values are shown in the $\mu$-$T$ phase diagram in Fig. 3.5.

### 3.2 Molecular-Level Simulations

Because the mean-field approximation neglects all local density fluctuations, we proceed to perform Monte Carlo simulations described in Sec. 2.5. We study a three-dimensional system of $N = 27,000$ sites with a random initial distribution of particles and water molecules. We have implemented a Metropolis algorithm for sampling of the configuration space, using the effective grand canonical Hamiltonian in Eq. (2.8) for determination of the statistical weights of different configurations $\{n_i\}$. After a sufficiently large number of relaxation steps, the system achieves thermal equilibrium and averages are taken over a considerable number of measurements to
estimate thermodynamic quantities.

![Graph showing phase diagrams for hydrophobic particles in water](image)

Figure 3.5: Comparison of $\mu$-$T$ phase diagrams for hydrophobic particles in water obtained by mean-field calculation (solid line) and by Monte Carlo simulations (○). The results agree within the simulation error below the UCST of the Monte Carlo simulation $T_{U,MC}^*$. The mean-field calculation results in a higher UCST than do Monte Carlo simulations (see text).

Fig. 3.5 shows the phase diagram of the system as a function of $\mu$ and $T$. On heating at constant $\mu$ (vertical lines) a phase transition is found at a temperature $T_L^* = 0.54 < T_1(\mu) < T_U^* = 2.07$ (for $\mu_L^* = 1.8 < \mu < \mu_U^* = 5.95$), where the particle density of the system jumps discontinuously from a value $\rho_{c2}$ to $\rho_{c1}$. At a constant temperature, such as $T_1$ in Fig. 3.3, the free-energy density $f_1(\rho)$ is minimal at values $\rho_{c1}$ and $\rho_{c2}$. Heating the system from below the transition temperature at fixed chemical potential $\mu$ results in a discontinuous jump in density from $\rho_{c1}(\mu)$ to $\rho_{c2}(\mu)$ at $T_t$.

Fig. 3.6 displays the $\rho$-$T$ phase diagram obtained from Monte Carlo simulations, where each data point represents the average of 10 independent simulations. The system shows a phase transition from a homogeneous state to a two-phase aggregation state at a lower transition temperature $T_L$, and a disaggregation at an upper transition temperature $T_U$, i.e. a closed-loop coexistence curve is found as expected from the mean-field description.
Figure 3.6: $\rho$-$T$ phase diagram for an aqueous solution of hydrophobic particles obtained by Monte Carlo simulations, illustrating the closed-loop coexistence curve. $\rho_p$ indicates the particle density. The system size is $N = 27000$. The dotted line represents a system of density $\rho_p = 0.5$ which is heated from a starting temperature $T_0$. At $T_L$, a phase transition occurs from the homogeneous state to an aggregated phase, while further heating results in a second transition at $T_U$ where the system disaggregates and the particles are dissolved again.

After rapid cooling at constant density $\rho_0$ from a temperature $T_0 > T_U$ to a fixed temperature $T < T_U$ in the coexistence region, the system develops a clear phase separation into a phase with density $\rho_{c2}$ and nearly pure water (density $\rho_{c1}$). All of the particles aggregate to form clusters, and after a certain period one single cluster of density $\rho_{c2}$ remains, which occupies a fraction $V_{c2} = \frac{\rho_{c2} - \rho_{c1}}{\rho_{c2} - \rho_{c1}}$ of the total volume.

An analytical solution is possible for the ground state of the system at vanishing temperature, which provides a test for the Monte Carlo simulations. The calculation, which uses the fact that $E_{ds} > 0$ and $E_{os} < 0$, results in

$$
S(T \to 0) = \lim_{\beta \to \infty} \left(-\frac{1}{\beta} \ln[q_{os}e^{-\beta E_{os}} + q_{ds}e^{-\beta E_{ds}}]\right)
= \lim_{\beta \to \infty} \left(-\frac{1}{\beta} \ln(q_{os}) + \ln(e^{-\beta E_{os}})\right)
= E_{os}
$$

(3.7)
and, analogously, \( B(T \to 0) = E_\phi \). At \( T = 0 \), minimizing the free energy is equivalent to minimizing the Hamiltonian in Eq. (2.8), which leads to three different phases depending on the chemical potential \( \mu \). For \( \mu > 2.0 \) the free energy is minimized by a solid phase \( (\rho_p = 1 \text{ and } H = 0) \), while for \( \mu < -5.0 \) no particles are dissolved and pure water is found \( (\rho_p = 0 \text{ and } H = B + \mu) \). Between these limiting cases, the system forms a dispersed crystal structure in which the particles are arranged in such a way that every water molecule is the neighbor of exactly one particle, which leads to a particle density \( \rho_p = \frac{1}{z+1} = 1/7 \) and to \( H = z(S + \mu)/(z+1) \). The different phases of a 2D system at \( T = 0 \) are illustrated in Fig. 3.7.

Monte Carlo simulations of the above system at sufficiently low temperatures confirm this behavior, as shown in the phase diagram in Fig. 3.8. The dot-dashed line shows the evolution of a system which is heated at constant particle density \( \rho_p = \frac{1}{z+1} \), which corresponds to the critical particle density \( \rho_p^* \) found in the mean-field calculation, starting in the dispersed crystal phase. The critical particle densities \( \rho_p^* \) determined by mean-field approximation and by Monte Carlo differ very slightly, but the discrepancy lies within the error of the Monte Carlo simulations.

### 3.3 Discussion

We have studied the aggregation of hydrophobic solute particles in water using a variant of the bimodal MLG model, which we have adapted to describe a coarse-grained system where each site may be occupied by one or more molecules. One of the objectives of the analysis was to reproduce the thermodynamic properties associated with the hydration and aggregation of non-polar solutes in water. For this purpose, Monte Carlo simulations were conducted to establish the phase diagram for
Figure 3.8: $\mu$-$T$ phase diagram for a system of non-polar solute particles in water, including the crystal phase appearing at low temperatures, obtained by Monte Carlo simulations for a system of size $N = 27\,000$. The lines show a system heated at constant density $\rho_p = 1/(z + 1)$ (---) and at $\rho_p = 0.5$ (\ldots{}), corresponding to the vertical lines in Fig. 3.6. Although the critical density $\rho_p^* = 1/(z + 1)$ determined by mean-field calculation differs very slightly from the value found by Monte Carlo simulations, it lies within the numerical error and no visible deviation is found from the expected behavior.

As expected, both methods display clear phase transitions at a LCST and an UCST within a range of densities, and thus define a closed-loop coexistence region. Outside this region the system appears as a homogeneous particle-solvent mixture, while inside it a separation occurs into two phases of fixed (upper and lower) coexistence densities. The exact time evolution of the system after a quench into both the metastable and the spinodal regions at constant density has not yet been studied, but would offer interesting insight into liquid-liquid phase-ordering kinetics [83, 84].

At low temperatures ($T_0$ in Fig. 3.6) the system is in the homogeneous region where the solubility of the solute is high and the solvent-induced effective interaction between solute particles repulsive. On raising the temperature at constant density, the system shows a sharp transition to an aggregation state at the LCST $T_L$. In the
equilibrium state at this temperature, clusters of aggregated, hydrophobic solute molecules with density $p_{c2}$ are suspended in nearly pure water (density $p_{c1} \approx 0$). This aggregation allows the solute particles to minimize their exposed surface and to reduce the structural enhancement of the surrounding water, thus causing a positive entropy change. The effective interaction between non-polar particles becomes attractive above $T_L$, and therefore the formation of aggregates is preferred over the solution of single solute particles. Further heating of the system results in a second phase transition at the temperature $T_U$ where the hydrophobic particles disaggregate, and above this temperature one homogeneous phase is found. This last process is dominated by the favorable entropy change of solvation in the whole system.

As can be seen from the $\mu$-$T$ phase diagram in Fig. 3.5, the transition line for the Monte Carlo simulations and for the mean-field calculation correspond rather well. They agree closely on the lower critical point, as well as on the densities above and below the transition temperatures. The upper critical point determined by Monte Carlo, however, lies at a lower temperature than the mean-field one. This result is not surprising, because it is well known that mean-field calculations, which neglect fluctuation effects, generally overestimate transition temperatures. The good agreement at low temperatures is rather a signature of the predominance of local effects which do not involve large fluctuations from site to site.

Apart from this difference, the coexistence curves from Monte Carlo and the mean-field calculations are qualitatively similar. At very low temperatures, however, the former show an additional crystal phase which cannot be explained by the latter, because it arises from spatial ordering of the hydrophobic solute and is therefore neglected by the present mean-field considerations, although this phase could be recovered within a more refined mean-field approximation. The appearance of this phase confirms the prediction of our analytical calculation at $T = 0$.

Encouraged by the qualitative and quantitative agreement of the mean-field calculation with Monte Carlo simulations, we have taken experimental values for the critical density $\rho_p^c$ in different systems to adapt the parameters of our model. Analytically, we have obtained the relation $\rho_p^c = \frac{1}{1+z}$ by mean-field calculations, and have confirmed this numerically by Monte Carlo simulations. Thus from an experimental value for $\rho_p^c$ we may extract an effective coordination number $z_{\text{eff}}$, which is then introduced in the calculations. The critical density for the system nicotine/water is $\rho_p^c = 0.4$ [16], which results in $z_{n/w}^{\text{eff}} = 1.5$. This coordination number $z_{\text{eff}}$ may be interpreted as the average number of hydrophobic solute molecules surrounding any chosen solute particle, which is relevant for the net effective hydrophobic interaction leading to attraction or repulsion. After changing the values for the energy levels to $E_{ds} = 3.4$, $E_{db} = 3.3$, $E_{cb} = -3.3$, and $E_{os} = -4.1$, the closed-loop coexistence curve is in good agreement with the experimental curve (Fig. 3.9). The UCST is higher than the measured one, as expected because of the mean-field nature of the
calculation. The ratio between the mean-field result $T_{U,\text{mf}}^*$ and the experimental value $T_{U,\text{exp}}^*$ is $T_{U,\text{mf}}^*/T_{U,\text{exp}}^* = 1.25$. In fact this value agrees quantitatively with the ratio $T_{U,\text{mf}}^*/T_{U,\text{MC}}^* = 1.35$ which we obtain by comparing the mean-field calculation with the more accurate Monte Carlo simulations when both are performed for $z = 6$ (cf. Fig. 3.5). Using this as an effective scaling factor to renormalize the mean-field results for different $\varepsilon_{\text{eff}}$ yields good agreement with the experimental results for the nicotine/water system (Fig. 3.9). We have repeated this procedure for the system poly(ethylene glycol) in water, which has a critical density of $\rho_p^* = 0.15$ and a much larger molecular weight $M_w = 3350$ [17], and find a similar agreement with the experimental curve.

Overall, we have shown that qualitative features of the liquid-liquid demixing process of hydrophobic aggregation may be explained successfully within a simple model for aqueous solutions of non-polar particles by including hydrophobic interactions only in terms of changes in water structure. Although the explicit terms of the model Hamiltonian describe solely the states of water molecules in solution, we have demonstrated that it contains implicitly both two- and even
many-particle interactions between hydrophobic solute molecules. The complete density-temperature phase diagram was established, by both analytical and numerical techniques, and illustrates the characteristic properties of hydrophobic aggregation.
Chapter 4

Chaotropic Effect

The addition of urea to an aqueous solution of hydrophobic molecules may affect the properties of the latter in a way that destabilizes aggregation of the non-polar solute [86]. In the case of protein solutions, this destabilization can result in a complete denaturation of the proteins, and even prevent their aggregation if the latter is due to hydrophobic interactions [87]. A highly concentrated solution of urea is therefore often used as a protein denaturant.

The underlying cause of this process, known as the chaotropic effect (Sec. 1.3), is generally believed to be a disruption of hydrogen bonds and hence a decrease in the order of the water structure (‘chao-trope’ = disorder maker), thus indirectly increasing the solubility of non-polar solutes [44, 88]. However, different attempts to discover how chaotropic agents perform this function have not yet been able to explain in a satisfactory manner the exact mechanism for disruption of the hydrogen bonds which stabilize the aggregate. In the remainder of this section we adapt the MLG-type model derived in Chapt. 2 by including chaotropic cosolvent molecules with the aim of demonstrating the chaotropic effect in our model system.

We first employ a mean-field approximation to show the destabilizing effect of chaotropic substances on aggregates of hydrophobic solutes. More accurate calculations are then performed within a pair approximation, and detailed Monte Carlo simulations of a three-dimensional system are presented to confirm preferential binding of the cosolvent to the non-polar particles.

4.1 Model including Chaotropic Cosolvent

The MLG model has been shown to provide an adequate description of the hydrophobic effect, including the heat-capacity anomaly, and to reproduce correctly both the ordering and strengthening of hydrogen bonds in the first solvation shell of an added non-polar solute particle at low temperatures, as well as the opposite
Chaotropic Effect Model including Chaotropic Cosolvent behavior at high temperatures [66]. It has been adapted successfully to provide a consistent account of the important properties of protein solutions, including warm and cold denaturation [67, 68]. Furthermore, the MLG model has been shown to yield a satisfactory account of hydrophobic interactions, because it contains implicitly the many-body interactions between non-polar solute particles, and to describe correctly the characteristic properties of hydrophobic aggregation, including the LCST and UCST (cf. Chapt. 3). Initial attempts to extend the model to include chaotropic cosolvent effects by taking into account purely geometric considerations, which cause an increase in solute solubility, have met with some success in highly dilute solutions [89, 90].

Chaotropic substances are in general those which are less strongly polar than water. In aqueous solutions of non-polar species they act to reduce the number of possible intact hydrogen bonds between water molecules, both in the solvation shell and in the bulk, compared to a pure water-solute mixture [91]. Within the adapted MLG framework, a straightforward approximation to the effect of a chaotropic cosolvent is to consider that its addition to strongly hydrogen-bonded, “ordered” clusters creates “disordered” clusters with additional broken hydrogen bonds and higher net energy. The creation of disordered states from ordered ones in the presence of a chaotropic cosolvent increases the degeneracy of the former at the expense of the latter (Fig. 4.1). The energy increase due to the breaking of hydrogen bonds is then included implicitly in the larger number of disordered states resulting from the changes in degeneracies, without the introduction of new energy parameters (below).

![Diagram](https://via.placeholder.com/150)

**Figure 4.1:** Illustration of the transformation of configurations from ordered to disordered by the addition of chaotropic substances due to the breaking of hydrogen bonds, which leads to an increasing number of disordered states of both shell and bulk water. The effect is much stronger in the bulk than in the shell.
Because the coarse-grained model treats each water site as containing a number of water molecules, the cosolvent molecules, which are generally rather small in comparison with the hydrophobic particles, may be included implicitly at each water site by adapting only the degeneracies of the energy levels of the water. The states of water clusters containing cosolvent are thus assigned the degeneracies

\[ q_{\text{cos},u} = q_{\text{cos}} - \eta_b, \quad q_{\text{ds},u} = q_{\text{ds}} + \eta_s, \]
\[ q_{\text{ob},u} = q_{\text{ob}} - \eta_b, \quad q_{\text{db},u} = q_{\text{db}} + \eta_b, \]  

where \( u \) denotes urea, a commonly used example of a small chaotropic cosolvent which we adopt for illustration. The cosolvent is taken to affect only the number of hydrogen bonds formed, and not their strength, so in the bimodal approximation the energies of the states remain unchanged.

The effect of a chaotropic cosolvent is much stronger in the bulk than in the shell. While in ordered, bulk water both tangentially and radially oriented hydrogen bonds may be broken, in shell water sites there exist fewer radially oriented bonds, and the ordered bulk configurations with radially oriented water have already become disordered configurations on substitution of the central water by a non-polar solute particle. The number of configurations available to be transformed from ordered to disordered by the addition of cosolvent is therefore much higher in the bulk than in the shell (\( \eta_b \gg \eta_s \)). Indeed, the number of configurations of ordered bulk states may be reduced almost to that of ordered shell water in the presence of a high concentration of strongly chaotropic cosolvent, because the primary difference between the two is the absence of radial hydrogen bonds in the shell. We have thus chosen to use \( \eta_b = 9.0 \) and \( \eta_s = 0.1 \) for the calculations to follow; based on the above considerations these values are expected to be suitably representative of a water/solute/cosolvent system for small cosolvents such as urea. We note briefly that the fractional value of \( \eta_s \) arises from the normalization of \( q_{\text{cos}} \) to unity, and does not imply a fractional degeneracy.

The enthalpically unfavorable reduction of the number of intact hydrogen bonds in bulk water caused by the cosolvent in solution gives rise to a preferential replacement of shell water by the cosolvent, known as preferential binding. The disruption of the local order around a solute particle causes an increase in hydration entropy in addition to the increase in hydration enthalpy (fewer intact hydrogen bonds). Because the entropy gain dominates the process at all temperatures of interest, there is a net free-energy reduction and thus an increase in solubility in the presence of a chaotropic cosolvent [44, 88].

Our first approach is based on an adaptation of the MLG model by increasing the number of disordered states of both shell and bulk water due to the presence of chaotropic substances (Sec. 4.2.1). In Sec. 4.2.2 we refine this study, which indeed provides a qualitative understanding of the chaotropic effect on hydrophobic interactions, and prove that the phenomenon can be reproduced semi-quantitatively within
the MLG framework by a suitably detailed approach which describes accurately
the ternary mixture of solvent (water), simplified cosolvent and solute (hydrophobic
particles).

In the ternary system of \( N \) sites on a cubic lattice, every site is occupied by either
pure water \((n_i = 1)\), hydrophobic particles \((n_i = 0)\) or cosolvent \((n_i = -1)\). The
energy of the system is given by the Potts-like Hamiltonian

\[
H[\{n_i\}, \{\sigma_i\}] = \sum_{i=1}^{N} n_i (n_i + 1) \left( \frac{1}{2} \right) \left( (E_{ab} \delta_{i,\sigma_o} + E_{db} \delta_{i,\sigma_d}) \lambda_i \right) + (E_{oa} \delta_{i,\sigma_o} + E_{da} \delta_{i,\sigma_d})(1 - \lambda_i) \right] + \frac{n_i(n_i - 1)}{2} \left( (E_{ob} \tilde{\delta}_{i,\sigma_{o,b,u}} + E_{db} \tilde{\delta}_{i,\sigma_{d,b,u}}) \lambda_i \right) + (E_{oa} \tilde{\delta}_{i,\sigma_{o,b,u}} + E_{da} \tilde{\delta}_{i,\sigma_{d,b,u}})(1 - \lambda_i)]
\]

(4.2)

where \( \lambda_i \) is defined as the product of the nearest neighbors, \( \lambda_i = \prod_{(j,i)} n_j^2 \), which
takes the value 1 if site \( i \) is completely surrounded by water and cosolvent or 0
otherwise. The first term of the sum, which is multiplied by \( \frac{1}{2} n_i(n_i + 1) \), is the energy
associated with pure water sites, and the second, containing \( \frac{1}{2} n_i(n_i - 1) \), is the energy
of cosolvent sites.

On site \( i \), pure water can be in one of the \( q \) different states, which are divided
among the four energy levels represented in Fig. 2.1. Thus \( \delta_{i,\sigma_o} \) is 1 if site \( i \) is
occupied by water in one of the \( q_{o,b} \) ordered shell states and 0 otherwise, and \( \delta_{i,\sigma_d} \) is 1
if it is occupied by pure water in one of the \( q_{d,b} \) disordered shell states and 0 otherwise.
The same applies for the states, and for the bulk and shell states of water sites which
include cosolvent.

The partition function of this system is given by

\[
Z = \sum_{\{n_i\}, \{\sigma_i\}} e^{-\beta H[\{n_i\}, \{\sigma_i\}]}, \tag{4.3}
\]

where every term of the sum represents the statistical weight of the corresponding
configuration. The number of states of each energy level must be taken into account
when summing over the configurations of the states of each site \{\( \sigma_i \)\}. The partition
function is then

\[
Z = \prod_{\{n_i\}} \left( q_{o,b} e^{-\beta E_{o,b}} + q_{d,b} e^{-\beta E_{d,b}} \right)^{n_i(n_i + 1)} \left( q_{o,b} e^{-\beta E_{o,b}} + q_{d,b} e^{-\beta E_{d,b}} \right)^{\frac{n_i(n_i - 1)}{2} \lambda_i} \tag{4.4}
\]

An effective Hamiltonian for the cosolvent system may be obtained by integrating
over the degrees of freedom of the water particles. To simplify the notation we define

\[
\mathcal{S} = -\frac{1}{\beta} \ln \left[ q_{o,b} e^{-\beta E_{o,b}} + q_{d,b} e^{-\beta E_{d,b}} \right], \tag{4.5}
\]
\[ B = -\frac{1}{\beta} \ln [q_{ob} e^{-\beta E_{ob}} + q_{db} e^{-\beta E_{db}}], \] (4.6)

\[ S_u = -\frac{1}{\beta} \ln [q_{os,u} e^{-\beta E_{os}} + q_{ds,u} e^{-\beta E_{ds}}], \] (4.7)

\[ B_u = -\frac{1}{\beta} \ln [q_{ob,u} e^{-\beta E_{ob}} + q_{db,u} e^{-\beta E_{db}}], \] (4.8)

which leads to the canonical partition function,

\[ Z_N = \sum_{\{n_i\}} e^{-\beta H_{\text{eff}}[\{n_i\}]}. \] (4.9)

The effective Hamiltonians \( H_{\text{eff}}[\{n_i\}] \) are identified with the free energies of the solvent for given particle configurations. If the number of particles may vary, a chemical potential is associated with the energy of particle insertion or removal. \( \mu \) represents the chemical potential for the addition of water to the solution and \( \Delta \mu \) the chemical potential for the addition of a cosolvent molecule to a water site. The grand canonical partition function becomes

\[ \Xi = \sum_N e^{\mu N_w + \beta (\mu + \Delta \mu) N_u} Z_N = \sum_{\{n_i\}} e^{-\beta H_{\text{eff}}[\{n_i\}]}, \] (4.10)

where

\[ H_{\text{eff}}[\{n_i\}] = \frac{1}{2} \sum_{i=1}^N [(S + \mu)n_i(n_i + 1) + (B - S)n_i(n_i + 1) \lambda_i + (S_u + \mu + \Delta \mu)n_i(n_i - 1) + (B_u - S_u)n_i(n_i - 1) \lambda_i], \] (4.11)

\( N_w \) denotes the number of pure water sites, and \( N_u \) is the number of water sites containing a cosolvent molecule. The total system size is \( N = N_w + N_u + N_p \), where \( N_p \) represents the number of particles. Effective interactions among two or more hydrophobic particles may be obtained using this effective Hamiltonian by substituting the density of water and cosolvent by the particle density \( m_i = 1 - \frac{1}{2} n_i(n_i + 1) - \frac{1}{2} n_i(n_i - 1) = 1 - n_i^2 \). This yields the effective Hamiltonian for single particles

\[ H_{\text{eff},p}[\{m_i\}] = K + \frac{1}{2} \sum_{i=1}^N [- (S + S_u + 2 \mu + \Delta \mu)m_i + (S - S_u + \Delta \mu)(1 - m_i)^{\frac{1}{2}} + (B - B_u - (S - S_u))(1 - m_i)^{\frac{1}{2}} + (B + B_u - (S + S_u))(1 - m_i)^{z + \frac{1}{2}}], \] (4.12)

where \( K = N_p \frac{1}{2} (S + S_u + 2 \mu + \Delta \mu) \) and \( z \) is the number of nearest neighbors of a site. It may be verified by complete replacement of the cosolvent by water (\( S = S_u \), \( B = B_u \), \( \Delta \mu = 0 \)) that Eq. (4.12) is equal to the effective Hamiltonian for the binary water-solute system (Chapt. 3). The effective Hamiltonian reduces to the form

\[ H_{\text{eff},p,\text{bin}}[\{m_i\}] = K + \sum_{i=1}^N [- \mu_{\text{eff}} m_i + J_{\text{eff}} (1 - m_i)^{z + 1}], \] (4.13)
where \( K = N(S + \mu) \), the effective chemical potential is \( \mu_{\text{eff}} = (S + \mu) \) and the effective interaction \( J_{\text{eff}} = (B - S) \). As \( B < B_u \) and \( S < S_u \), \( \mu_{\text{eff}} \) increases with rising cosolvent concentration, resulting in a smaller resistance of the system to insertion of particles into the solution. Furthermore, the relations \( \frac{1}{2}(B + B_u - (S + S_u)) > (B - S) \) and \( (B_u - B) > (S_u - S) \) imply a stronger repulsive interaction between the particles in the presence of the cosolvent. This leads to an overall tendency towards increased solubility of the non-polar solute and to a suppression of the coexistence region. At extremely high cosolvent concentrations, the coexistence region disappears completely.

### 4.2 Mean-Field Approximation

In order to make some preliminary qualitative predictions, spatial fluctuations of the order parameter are assumed to be insignificant, and a mean-field study is performed of the thermodynamic properties of the ternary system. We begin with an elementary mean-field approximation in which the concentrations of chaotropic agents in the bulk and in the first solvation shell are assumed to be equal. Allowing the system to have two different concentrations may be expected to provide a better description of the exact mechanism of the chaotropic effect, and will be subject of the next subsection.

#### 4.2.1 Single Mean Cosolvent Concentration

In the simplest approach, we consider a single concentration, \( c \), of chaotropic cosolvent added to the water, which is uniform across the entire system. This calculation requires only the adaptation of the model for a pure-water/solute mixture described in Sec. 2.2, where each site is occupied either by water \((n_i = 1)\) or by a solute particle \((n_i = 0)\). To obtain an uniform cosolvent concentration, each water site is considered to contain the same amount of cosolvent. Thus a fraction \( c \) of water molecules in each cluster represented by a site is in a state which is modified by the presence of the cosolvent. This leads to an additional factor in the partition function described in Eq. 2.3,

\[
Z^u = \sum_{\{n_i\}} \prod_i \left( q_{os} e^{-\beta E_{os}} + q_{ds} e^{-\beta E_{ds}} \right)^{n_i (1 - \lambda_i)} \left( q_{os,u} e^{-\beta E_{os}} + q_{ds,u} e^{-\beta E_{ds}} \right)^{n_i (1 - \lambda_i) c} \times \left( q_{ob} e^{-\beta E_{ob}} + q_{db} e^{-\beta E_{db}} \right)^{n_i \lambda_i (1 - c)} \left( q_{ob,u} e^{-\beta E_{ob}} + q_{db,u} e^{-\beta E_{db}} \right)^{n_i \lambda_i c}.
\]

Here the concentration of chaotropic cosolvent in the bulk is assumed to be identical to the concentration in the first solvation shell of a hydrophobic particle. On defining \( B_u = -\frac{1}{\beta} \ln[q_{ob,u} e^{-\beta E_{ob}} + q_{db,u} e^{-\beta E_{db}}] \) and \( S_u = -\frac{1}{\beta} \ln[q_{os,u} e^{-\beta E_{os}} + q_{ds,u} e^{-\beta E_{ds}}] \), the
effective Hamiltonian function becomes

\[ H^u_{\text{eff}}[\{n_i\}] = \sum_{i=1}^{N} [\mu^u_{\text{eff}} n_i + J^u_{\text{eff}} n_i \lambda_i], \tag{4.15} \]

where the effective interaction \( J^u_{\text{eff}} \) and the effective chemical potential \( \mu^u_{\text{eff}} \) are given by

\[ J^u_{\text{eff}} = B - S + (B_u - B + S - S_u)c, \tag{4.16} \]
\[ \mu^u_{\text{eff}} = S + \mu + (S_u - S)c. \tag{4.17} \]

Because we are concerned with the hydrophobic interaction between solute particles, we again express the Hamiltonian in terms of particles \((m_i = 1 - n_i)\) obtaining

\[ H^u_{\text{eff},p}[\{m_i\}] = K^u - \sum_{i=1}^{N} [\mu^u_{\text{eff}} m_i + J^u_{\text{eff}} (m_i - 1) \lambda_i], \tag{4.18} \]

where \( K^u = N \mu^u_{\text{eff}} \) and \( \lambda_i = \prod_{(i,j)} (1 - m_j). \) The effective interaction between the hydrophobic particles thus depends on the concentration of chaotropic agent in the solution. Because \( S_u > S \) and \( B_u > B, \mu^u_{\text{eff}} \) is larger than \( \mu_{\text{eff}} \) and therefore it is easier to bring the non-polar solute into solution in the presence of chaotropic substances. In addition, the relation \((B_u - B) > (S_u - S)\) results in an increase in \( J^u_{\text{eff}} \) when raising the concentration \( c. \) Hence, the chaotropic particles support the repulsive force between non-polar solute molecules. The coexistence region of the phase diagram consequently shrinks with increasing \( c, \) and for extremely high concentrations may even disappear entirely.

To make qualitative predictions, we use a mean-field approximation for the mean occupancy of a site \( \langle n_i \rangle = \rho. \) The grand canonical free energy per site is given by

\[ f = J^u_{\text{eff}} \rho^{z+1} + \mu^u_{\text{eff}} \rho + \beta^{-1}[\rho \ln \rho + (1 - \rho) \ln (1 - \rho)]. \tag{4.19} \]

If the cosolvent concentration \( c \) vanishes, the results reduce to those of the binary solute-solvent system of Sec. 2.4.

Fig. 4.2 shows the effect on the coexistence curve of different urea concentrations \( c, \) using \( \eta_b = 9 \) and \( \eta_s = 0.1. \) The expected increase in solubility is confirmed, and the LCST and UCST approach each other with increasing cosolvent concentration. At a critical concentration, \( c^*_e = 0.935 \) for the parameters chosen, the closed-loop curve shrinks to a single point \((T^*_c, \mu^*_c = 2.84, \text{ and } \rho^*_c = 1/(z + 1)), \) which represents a double critical point \([52]. \) For concentrations higher than \( c^*_e, \) the aggregation phase disappears completely.
Figure 4.2: $\rho$-$T$ phase diagram for a ternary system consisting of water, hydrophobic particles, and chaotropic cosolvent, obtained by mean-field calculation, showing coexistence curves for different cosolvent concentrations. The $\bullet$ represents a multi-critical point at the critical cosolvent concentration $c_c^* = 0.935$ where the coexistence curve has shrunk to a single point.

4.2.2 Distinct Mean Bulk and Shell Concentrations

In this section we refine the mean-field treatment by allowing for the possibility that the concentrations of chaotropic agents in the bulk and in the first solvation shell may differ which may be expected to provide better insight into the exact mechanism of the chaotropic effect than the results of the previous subsection. In experiment chaotropic molecules are found preferentially in the solvation shell of non-polar solute particles. For this purpose we perform a mean-field calculation for the model introduced in Sec. 4.1, where each site is occupied by either pure water ($n_i = 1$), hydrophobic particles ($n_i = 0$) or a cosolvent/water mixture ($n_i = -1$). In this representation the cosolvent density varies according to the type of site, and preferential binding may emerge.

The local density at a site is approximated by the mean concentration in the system expressed using the densities of hydrophobic particles $\rho_p$, pure water $\rho_w$, and cosolvent $\rho_u$, where $\rho_p + \rho_w + \rho_u = 1$. These densities are thus considered constant throughout the system. The mean values of the variable $n_i$ of site $i$ is then given by $\langle n_i \rangle = (+1) \rho_w +$
By analogous calculation one finds $\langle n_i^2 \rangle = (+1)^2 \rho_w + (-1)^2 \rho_u + 0^2 \rho_p = \rho_w + \rho_u$. The grand canonical mean-field free energy per site $f = \frac{F}{N}$ is then found to be

$$f = (S + \mu)\rho_u + (B - S)\rho_w(\rho_u + \rho_w) + (S_u + \mu + \Delta\mu)\rho_u + (B_u - S_u)\rho_u(\rho_u + \rho_w)^2 + \beta^{-1}(\rho_w \ln \rho_w + \rho_u \ln \rho_u + \rho_p \ln \rho_p).$$

(4.20)

Every site has $z$ nearest-neighbor sites, which can be occupied by a hydrophobic particle, by pure water or by water with a cosolvent molecule. The coordination number $z$ can be interpreted as the average number of hydrophobic particles surrounding any one solute molecule, the quantity relevant for the effective hydrophobic interaction between solute particles. After replacing $\rho_p$ by $1 - \rho_w - \rho_u$, the equilibrium densities at given temperature and chemical potentials may be determined by minimizing $f$ with respect to $\rho_w$ and $\rho_u$ under the constraint $\rho_w + \rho_u \leq 1$.

For the purpose of determining whether the non-polar particles prefer to remain in solution or to precipitate at a certain temperature, we compare the free energy of a ternary system for a given density of dissolved particles with that of a binary system consisting only of water and cosolvent. The latter represents a complete separation of solvent and solute into two phases (the number of solvent sites is equal in both systems) whose interface is neglected (all solvent molecules are considered to be in the bulk), and its equilibrium densities are calculated by minimizing the free energy in Eq. (4.20). For comparison, the free energy of the ternary system is calculated by using the same amounts of solvent and cosolvent, but including a fixed number of solute particles. Here the system is forced to form one homogeneous phase, although a phase separation may in fact be favored.

### 4.2.3 Results

It can be seen in Fig. 4.3 that the free energy per particle [Eq. (3.1)] of the homogeneous ternary mixture (with particle density $\rho_p = 0.1$) is higher than that of the completely separated system between a lower and an upper critical temperature (respectively $T_L$ and $T_U$). The system separates into two phases between $T_L$ and $T_U$, whereas the hydrophobic particles are dissolved below $T_L$ and above $T_U$. For comparison, the difference in free energy per particle of a system without cosolvent (but with the same value of solute density) is also shown: the temperature range in which a coexistence of two phases is found is reduced in the presence of the cosolvent. These qualitative results demonstrate that an idealized chaotropic cosolvent increases the solubility of non-polar particles.

In the above calculation the particle and cosolvent densities are fixed artificially, and therefore the ternary system is not in thermodynamic equilibrium. In the following, the free energy in Eq. (4.20) is minimized with respect to $\rho_w$ and $\rho_u$, and the
Chaotropic Effect Mean-Field Approximation

into two phases, corresponding to nearly pure water and to particle aggregates formed.

As expected, closed-loop transition curves are found, showing corresponding UCSTs and LCSTs. An increase in cosolvent density leads to a diminishing of the coexistence region inside which the system separates into two phases, corresponding to nearly pure water and to particle aggregates formed as a result of their mutually attractive interactions. On the outside of each curve, specified by fixed \( \rho_u \), the hydrophobic interaction becomes repulsive and the particles are in solution. At the critical cosolvent concentration, which in the refined mean-field treatment is \( \rho_u^* = 0.81 \) for the parameters of Secs. 2.1 and 4.1, the coexistence curve shrinks to a double critical point at \( T^* = 1.35 \) and particle density \( \rho_p^* = 0.14 \). For even higher cosolvent concentrations the coexistence region disappears completely.

The \( \rho-T \) phase diagram in Fig. 4.4 illustrates the coexistence region of the solution for different concentrations of cosolvent. As expected, closed-loop transition curves are found, showing corresponding UCSTs and LCSTs. An increase in cosolvent density leads to a diminishing of the coexistence region inside which the system separates into two phases, corresponding to nearly pure water and to particle aggregates formed as a result of their mutually attractive interactions. On the outside of each curve, specified by fixed \( \rho_u \), the hydrophobic interaction becomes repulsive and the particles are in solution. At the critical cosolvent concentration, which in the refined mean-field treatment is \( \rho_u^* = 0.81 \) for the parameters of Secs. 2.1 and 4.1, the coexistence curve shrinks to a double critical point at \( T^* = 1.35 \) and particle density \( \rho_p^* = 0.14 \). For even higher cosolvent concentrations the coexistence region disappears completely.
Figure 4.4: Closed-loop coexistence curves for ternary systems of water, cosolvent and hydrophobic particles with different cosolvent densities $\rho_u$, obtained by mean-field calculation. On the outside of each curve the solution is homogeneous, and on the inside it separates into two phases. At a critical cosolvent density of $\rho_u^* = 0.81$, the LCST and UCST coincide (at $T^* = 1.35$ and for particle density $\rho_p^* = 0.14$) and the region of phase separation is reduced to a single point. Above this density the solution is homogeneous over the whole temperature range. The initial particle density at the LCST and UCST remains essentially the same for all cosolvent concentrations.

The $\mu$-$T$ phase diagram of the solution is presented in Fig. 4.5 for a given cosolvent concentration. A finite transition line appears which terminates at the LCST and UCST. At these temperatures the coexistence region consists of a single point (see Fig. 4.4), and the two local minima, along with both inflection points of $f$, coincide. Thus the second and third derivatives of $f$ vanish, and the density at which the coexistence region vanishes can be determined exactly. This was calculated in Chapt. 3 for the case without cosolvent.

Fig. 4.5 reveals that the transition line shrinks as the cosolvent concentration increases: $T_U^*$ decreases while $T_L^*$ increases until they coincide at the critical point. Fig. 4.4 indicates that the particle density at these critical solution temperatures remains essentially constant, and may thus be taken to be independent of the cosolvent concentration. The $\mu$-$T$ phase diagram shows a clear shift of the transition line towards larger values of $\mu$ as the cosolvent concentration increases, reflecting the fact that it becomes more difficult to add water to the solution at constant volume in
As the cosolvent density \( \rho_u \) increases, the LCST and UCST approach until they meet at \( T_L^* = T_U^* = 1.35 \) (and \( \mu_L = \mu_U = 7.05 \)), whence \( \rho_p = 0.14 \) for \( \rho_u = 0.81 \). The shift of the transition lines towards larger \( \mu \) values with growing cosolvent density indicates that it becomes more difficult to add water to the solution. This is expected because the particle density remains unchanged at the LCST and UCST (cf. Fig. 4.4), so with increasing cosolvent density at constant volume the water density must decrease.

the presence of the cosolvent. To obtain the same particle density as for pure water, the chemical potential (which represents the resistance of the solution to addition of water) increases as the cosolvent concentration becomes higher. In other words, because the resistance to particle addition is connected to both \( \mu \) and \( \Delta \mu \), and \( \Delta \mu \) decreases when the cosolvent concentration increases, \( \mu \) must increase to maintain a fixed critical particle density.

### 4.3 Molecular-Level Simulations

The mean-field approximation neglects local density fluctuations. Consequently, this approach is unable to detect preferential exclusion of cosolvent particles from the solvation shell of a hydrophobic particle. For this we perform Monte Carlo simulations, which represent an appropriate method for the efficient calculation of
Chaotropic Effect Molecular-Level Simulations

Figure 4.6: Cross-sections of a 3D system of hydrophobic particles (black) in a solution of chaotropic cosolvent (red) in water (suppressed). The cosolvent density is $\rho_c = 0.13$ and the particle density $\rho_p = 0.21$. At temperature $T = 0.5 < T_L$ the system is a homogeneous mixture (left), while at temperature $T_L < T = 1.0 < T_U$ a separation into two phases is found (right). Preferential binding of the cosolvent sites to the solute particles occurs.

Figure 4.6: Cross-sections of a 3D system of hydrophobic particles (black) in a solution of chaotropic cosolvent (red) in water (suppressed). The cosolvent density is $\rho_c = 0.13$ and the particle density $\rho_p = 0.21$. At temperature $T = 0.5 < T_L$ the system is a homogeneous mixture (left), while at temperature $T_L < T = 1.0 < T_U$ a separation into two phases is found (right). Preferential binding of the cosolvent sites to the solute particles occurs.

We begin by showing instantaneous configurations taken from one step of the Monte Carlo simulation, in the equilibrium regime, for a solution of hydrophobic particles in a solution of chaotropic cosolvent in water at different temperatures (Fig. 4.6). At the (coarse-grained) molecular level it is clear that the system forms a homogeneous phase below the LCST, while for temperatures above the LCST a clear phase separation into a pure solvent phase and an aggregation phase is found.

Fig. 4.7 presents the $\rho$-$T$ phase diagram for different chemical potentials $\Delta \mu$, corresponding to different cosolvent concentrations. The qualitative features are as in Fig. 4.3: a closed-loop coexistence curve is found, inside which the system is separated into two phases of different particle densities. Here the effective hydrophobic interaction is attractive and the particles aggregate in order to minimize their surface exposed to the solvent. Outside the curve, however, the non-polar particles are in thermal averages in many-particle systems with statistical fluctuations. A system of $N = 27000$ sites on a cubic lattice is taken, with random initial particle distributions, and periodic boundary conditions to eliminate boundary effects. We have implemented a Metropolis algorithm for sampling of the configuration space, using the effective Hamiltonian in Eq. (4.11) for determination of the statistical weights of different configurations $\{n_i\}$. A more detailed description of the Monte Carlo simulations is provided in Sec. 2.5.
Figure 4.7: Closed-loop miscibility curves of an aqueous solution of hydrophobic particles for different cosolvent concentrations $\rho_u$, obtained by Monte Carlo simulations. Outside the curve the particles are soluble and the system is homogeneous, while inside it two phases are found, namely aggregates of hydrophobic particles and a nearly pure solvent-cosolvent mixture. As the cosolvent concentration grows, the solubility of the particles increases, leading to a reduced coexistence curve.

solution and a homogeneous mixture is found. Because the hydrophobic interaction is repulsive in this region, no aggregation occurs. Consequently, heating a system from a temperature $T_0$, at which the hydrophobic particles are in solution, results in a phase transition at $T_L$ where aggregation occurs, and further heating leads to a disaggregation phase transition at $T_U$.

The $\mu$-$T$ phase diagram of the ternary system is shown in Fig. 4.8. The aggregation phase transition occurs over lines of finite length, which provide a LCST and an UCST. With increasing cosolvent concentration the phase transition is shifted in $\mu$, confirming the increase in resistance of the system towards the addition of water. As displayed in Figs. 4.7 and 4.8, the separation of the LCST and the UCST decreases with increasing cosolvent concentration.

Three-component phase diagrams are best illustrated by a triangular plotting technique for mixtures of any proportions at a given temperature. The percentage of each component is represented by the vertical distance to the opposite side. Fig. 4.9
Figure 4.8: $\mu$-$T$ phase diagram of hydrophobic solute in an aqueous cosolvent mixture for different cosolvent concentrations $\rho_u$, obtained by Monte Carlo simulation. The finite transition lines terminate at an UCST and a LCST, which approach each other with increasing cosolvent concentration. The values of $\Delta\mu$ used for the calculations are determined by the requirement that the cosolvent density remain constant.

shows the ternary phase diagram for temperature $T = 1.0$. At this temperature a clear transition occurs between a two-phase region and one in which the three compounds are completely miscible and no aggregation occurs. By contrast, below $T_L^*$ and above $T_U^*$ the components are miscible in all proportions. The dotted line represents a phase-separation line: a solution whose total composition $a_0$ (with particle density $\rho_{p,0}$) falls on this line is separated into two phases of compositions $a_1$ (particle density $\rho_{p,1}$) and $a_2$ (particle density $\rho_{p,2}$), respectively. The volume occupied by phase $i$ is given by $V_i = \frac{\rho_{p,0} - \rho_{p,i}}{\rho_{p,i} - \rho_{p,j}} V_0$, where $j$ denotes the other phase and $V_0$ is the total volume.

The effect of preferential binding may be illustrated by considering the (relative) concentrations of cosolvent in shell and bulk sites. As shown in Fig. 4.10, the cosolvent concentration is higher in the shell of non-polar solute particles than in the bulk. This implies a preferential binding in the ternary system at all temperatures, which increases the solubility of the solute. The tendency is clearly stronger at low than at high temperatures, where entropy effects are predominant.

Fig. 4.11 presents the relative cosolvent shell concentration compared to the
Figure 4.9: Ternary phase diagram of hydrophobic particles, water and chaotropic cosolvent at temperature $T = 1.0$, obtained by Monte Carlo simulations. Any mixture with composition $\rho_p$, $\rho_w$, and $\rho_u$ (with $\rho_p + \rho_w + \rho_u = 1$) is represented by a point whose vertical distances to the respective sides correspond to the compound densities. A clear transition is found from a homogeneous region to a two-phase region. The dotted line ($\cdots$) represents all compositions $a_0$ which separate into two phases with compositions $a_1$ and $a_2$. The phase diagram is determined by varying the chemical potentials $\mu$ and $\Delta\mu$ to span the entire range of densities.

Overall cosolvent concentration at temperature $T = 1.0$ for different total cosolvent concentrations. At low chemical potential (i.e. at high water and thus low particle density for given cosolvent density) the preferential binding is significant, whereas beyond the phase transition to a high-density solution the effect is only marginal. If the particle concentration is significant the total number of shell sites becomes substantial, and in fact most solvent and cosolvent sites are shell sites. Hence, the shell cosolvent concentration may be high only if the total number of cosolvent sites is large. This is confirmed by the fact that at high particle density the preferential binding is only marginally higher for high than for low cosolvent concentrations. However, at low particle densities (small $\mu$) the shell density of the cosolvent is clearly higher than its total density. The magnitude of the effect can be attributed to the measurement of the relative increase in concentration. If the total number of cosolvent sites is small, the same number of shell sites preferentially occupied by cosolvents gives rise to a larger relative increase in density in the shell as compared to the overall density. An
Figure 4.10: Tendency to preferential binding as a function of temperature for two different cosolvent ($\rho_u$) and particle ($\rho_p$) densities, represented by the cosolvent concentration in the shell of hydrophobic particles relative to the total cosolvent concentration in the solvent (excluding the volume occupied by hydrophobic particles), obtained within the Monte Carlo framework. This ratio is always larger than one, indicating that preferential binding is favored over the whole temperature range.

extremely small number of cosolvent sites (on the order of 100 for the lowest densities used in the simulations) also results in rather noisy curves for low cosolvent densities due to large fluctuation effects.

### 4.4 Pair Approximation

The mean-field approximation of Sec. 4.2 provided confirmation of hydrophobic aggregation, and of the fact that chaotropic cosolvents increase the solubility of non-polar solutes, leading to disaggregation. However, no conclusions can be drawn concerning the molecular distribution in the system, because local densities are replaced by their averages and spatial fluctuations are thus neglected. The cluster-variation method offers an accurate approximation technique for the detailed study of lattice systems on the molecular level [76, 77, 78]. Phase diagrams for a variety of systems have been determined by this method, which is based on a variational
Figure 4.11: Preferential binding for different cosolvent densities $\rho_u$ at temperature $T = 1.0$, represented by the cosolvent concentration in the shell of hydrophobic particles compared to the total cosolvent concentration, computed by Monte Carlo simulations. The effect is most pronounced at low cosolvent concentrations and at low particle concentrations (at small $\mu$). The former can be attributed to the fact that at low cosolvent concentrations a larger relative increase is possible, while the latter is due to the fact that the majority of solvent and cosolvent sites are shell sites at high particle density, thus rendering the total cosolvent concentration nearly equal to its shell concentration.

approach. If the basic clusters under consideration consist of two neighboring lattice sites (pair approximation), it is equivalent to the Bethe approximation [79]. We have implemented the pair approximation, as described in Sec. 2.6 for the solute-solvent-cosolvent system, to obtain additional information about the local concentration of cosolvent and about intersite correlations.

### 4.4.1 Model in Pair Approximation

Nine possible pair configurations exist, namely $ww, wp, wu, uw, up, uu, pw, pp$ and $pu$, where $w$ stands for a pure water site, $p$ for a particle and $u$ for a water site including a cosolvent molecule. The sum of the pair densities is unity, $\rho_{ww} + \rho_{wp} +$
\[ \rho_{wu} + \rho_{uw} + \rho_{up} + \rho_{pu} + \rho_{pw} + \rho_{wp} + \rho_{pu} = 1. \] The effective Hamiltonian

\[ H^\text{eff}[\{n_i\}] = \sum_{i=1}^{N} ((S+\mu)n^w_i + (S_u+\mu+\Delta \mu)n^u_i + (B-S)n^w_i \lambda_i + (B_u-S_u)n^u_i \lambda_i) \] (4.21)

is expressed in terms of single-site densities, which can be represented using the pair densities. The water density \( \rho_w \) may then be written as \( \rho_w = \rho_{wu} + \frac{1}{2}(\rho_{wp} + \rho_{wu}) \), which leads to \( \rho_w = \rho_{wu} + \rho_{wp} + \rho_{wu} \), making use of the symmetries \( \rho_{wu} = \rho_{uw} \), \( \rho_{wp} = \rho_{pw} \), and \( \rho_{up} = \rho_{pu} \). By analogy, the particle and the cosolvent densities become \( \rho_p = \rho_{wp} + \rho_{pw} + \rho_{pu} \) and \( \rho_u = \rho_{wu} + \rho_{uw} + \rho_{up} \), respectively, and the total density \( \rho_w + \rho_p + \rho_u = 1 \) as expected.

In Eq. (4.21) the mean occupancies of water and cosolvent sites can be approximated by \( \langle n^w_i \rangle = \rho_w = \sum_\sigma \rho_{w\sigma} \) and \( \langle n^u_i \rangle = \rho_u = \sum_\sigma \rho_{u\sigma} \), where \( \sigma \in \{w, p, u\} \). The term \( n^w_i \lambda_i \) represents the product for a water site \( i \), and has \( z = 6 \) nearest neighbors which contribute with weight 1 if they all are water or cosolvent sites or 0 otherwise.

The term may therefore be expressed as the sum of all possible configurations of water and cosolvent sites for these seven sites, provided that the central position contains pure water. The term \( n^u_i \lambda_i \) is obtained by analogy. Taking into account all possible permutations by including the factors \( C^z_k = \frac{1}{k!(z-k)!} \), the effective Hamiltonian density becomes

\[ h = (S + \mu) \sum_\sigma \rho_{w\sigma} + (S_u + \mu + \Delta \mu) \sum_\sigma \rho_{u\sigma} + (B_u - S_u) \frac{1}{(\sum_\sigma \rho_{w\sigma})^{z-1}} \sum_{k=0}^{z} C^z_k \rho_{wu}^{z-k} \rho_{wu}^k \]

\[ + (B - S) \frac{1}{(\sum_\sigma \rho_{u\sigma})^{z-1}} \sum_{k=0}^{z} C^z_k \rho_{wu}^{z-k} \rho_{wu}^k . \] (4.22)

The factors \( \rho_{wu}^{z-k} \) and \( \rho_{wu}^k \) in the last two terms represent a normalization of the product required because the site density of the central site appears in all \( z \) pair densities, instead of only once.

In the cluster variation method, the entropy density is obtained from the sum of the entropy densities of all clusters (pairs) and subclusters (sites) \( \alpha \), leading to \( s = -\sum_\alpha a_\alpha \rho_\alpha \ln(\rho_\alpha) \). The coefficients \( a_\alpha \) are given by the geometrical constraint \( a_{site} + z a_{pair} = 1 \), and \( a_{pair} = 1 \) results in \( a_{site} = 1 - z \). The entropy density is then

\[ s = -\frac{z}{2} k_B \sum_{\sigma \in \{w,p,u\}} \sum_{\sigma' \in \{w,p,u\}} \rho_{\sigma,\sigma'} \ln \rho_{\sigma,\sigma'} + k_B (z-1) \sum_{\sigma \in \{w,p,u\}} \rho_\sigma \ln \rho_\sigma, \] (4.23)

and the total free-energy density per site of the three-component system is given in the pair approximation by

\[ f = h - Ts, \] (4.24)
where $h$ is the enthalpy density, $f$ is thus a function of the pair and site densities. To calculate the configuration of minimal free energy, and thus the equilibrium state of the system, the first derivatives of $f$ with respect to all pair densities must vanish,

$$
\frac{\partial f}{\partial \rho_{\sigma,\sigma'}} = \frac{\partial h}{\partial \rho_{\sigma,\sigma'}} - T \frac{\partial s}{\partial \rho_{\sigma,\sigma'}} = 0 \quad \forall \sigma \text{ and } \sigma' \in \{w, p, u\}. \tag{4.25}
$$

These derivatives are best expressed by introducing a function $m_{\sigma,\sigma'}$, which is one if $\sigma = \sigma'$ and two otherwise. The derivative of the entropy density then becomes

$$
\frac{\partial s}{\partial \rho_{\sigma,\sigma'}} = -\frac{z}{2}k_B m_{\sigma,\sigma'} \ln(\rho_{\sigma,\sigma'})
$$

$$
+ k_B(z - 1)\frac{m_{\sigma,\sigma'}}{2} \left[ \ln \left( \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\sigma,\hat{\sigma}} \right) + \ln \left( \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\sigma',\hat{\sigma}} \right) \right] \tag{4.26}
$$

while for the enthalpy density

$$
\frac{\partial h}{\partial \rho_{\sigma,\sigma'}} = (S - \mu)(\delta_{\sigma,w} + \delta_{\sigma',w} - \delta_{\sigma,w}\delta_{\sigma',w}) + (S - \mu - \Delta \mu)(\delta_{\sigma,u} + \delta_{\sigma',u} - \delta_{\sigma',u}\delta_{\sigma',u})
$$

$$
+ (B - S) \frac{1 - z}{\left( \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\omega,\hat{\sigma}} \right)^z} \left( \delta_{\sigma,w} + \delta_{\sigma',w} - \delta_{\sigma,w}\delta_{\sigma',w} \right) \sum_{k=0}^{z} C_k^z \rho_{\omega,w}^{z-k} \rho_{\omega,u}^k
$$

$$
+ (B - S) \frac{1}{\left( \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\omega,\hat{\sigma}} \right)^{z-1}} \left[ \sum_{k=0}^{z-1} C_k^z (z - k) \rho_{\omega,u}^{z-k} \rho_{\omega,u}^k \delta_{\sigma,w}\delta_{\sigma',w} \right]
$$

$$
+ (B_u - S_u) \frac{1 - z}{\left( \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\omega,\hat{\sigma}} \right)^z} (\delta_{\sigma,u} + \delta_{\sigma',u} - \delta_{\sigma,u}\delta_{\sigma',u}) \sum_{k=0}^{z} C_k^z \rho_{\omega,u}^{z-k} \rho_{\omega,u}^k
$$

$$
+ \left[ \sum_{k=1}^{z} C_k^z \rho_{\omega,u}^{z-k} \rho_{\omega,u}^k \delta_{\sigma,w}\delta_{\sigma',w} + \sum_{k=0}^{z-1} C_k^z (z - k) \rho_{\omega,u}^{z-k-1} \rho_{\omega,u}^k \delta_{\sigma,u}\delta_{\sigma',u} \right]
$$

$$
\times (B_u - S_u) \frac{1}{\left( \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\omega,\hat{\sigma}} \right)^{z-1}} \tag{4.27}
$$

The minimization of $f$ involves the simultaneous solution of a set of equations for the different pair densities. These are high-order algebraic equations whose solution may be obtained by the natural iteration method, a technique developed explicitly for solving equations in the cluster variation method [76]. In this procedure the pair densities are expressed in terms of the densities of the subclusters (sites) by solving Eq. (4.25), to yield

$$
\rho_{\sigma,\sigma'} = e^{-\frac{z}{2}m_{\sigma,\sigma'} \rho_{\sigma,\sigma'}} \left[ \rho_{\sigma,\sigma'} \right]^{z-1}, \tag{4.28}
$$

where the site densities are respectively $\rho_{\sigma} = \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\sigma,\hat{\sigma}}$ and $\rho_{\sigma'} = \sum_{\hat{\sigma} \in \{w, p, u\}} \rho_{\sigma',\hat{\sigma}}$. The procedure then consists in the iterative solution of Eqs. (4.28) for all $\sigma$ and $\sigma'$, including the normalization according to $\sum_{\sigma,\sigma' \in \{w, p, u\}} \rho_{\sigma,\sigma'} = 1$, until convergence is attained.
Chaotropic Effect Pair Approximation

Figure 4.12: Closed-loop miscibility curves of hydrophobic particles in a water-cosolvent solution for various cosolvent concentrations $\rho_u$, obtained within the pair approximation. These curves confirm the results of the mean-field calculation and of Monte Carlo simulations. At the critical cosolvent concentration $\rho_u^* = 0.76$ the LCST and the UCST coincide ($T_L = T_U = T^* = 1.35$).

### 4.4.2 Results

Fig. 4.12 presents the closed-loop miscibility curve of the ternary solution for different cosolvent concentrations. Clearly the addition of a chaotropic cosolvent increases the solubility of the solute, and acts to suppress hydrophobic aggregation. At concentrations higher than a critical density $\rho_u^* = 0.76$, the three components are miscible in all proportions over the whole temperature range. The finite transition lines in the $\mu$-$T$ phase diagram (Fig. 4.13) are in good, quantitative agreement with the results obtained from the mean-field calculations of Sec. 4.2 (Fig. 4.5), as well as with the Monte Carlo simulations of Sec. 4.3. Furthermore, the coexistence curves are similar to our previous results, with a critical temperature $T^* = 1.35$, where the transition line is reduced to a single point, in perfect agreement with the mean-field calculation.

In Fig. 4.14 the cosolvent concentration in the shell is compared to the total cosolvent concentration (in the total volume occupied by both water and cosolvent) at chemical potential $\mu = 3.0$. The shell concentration is always higher than the total...
Figure 4.13: Pair approximation to the $\mu$-$T$ phase diagram of an aqueous solution of non-polar solute with different cosolvent concentrations. The diagram is in good agreement with those obtained by mean-field calculation and Monte Carlo simulation. With increasing cosolvent concentration, the LCST and UCST approach until they coincide at $\rho_u^*$. Above this concentration, the phase transition line disappears, and the three components are completely miscible at all temperatures.

Concentration, and we note in particular that above $T_L$ the concentration in the shell can become nearly twice as high as that in the bulk. The fact that this preferential binding is so pronounced for low cosolvent concentrations is not surprising, because in this case a larger proportion of the cosolvent may actually contribute to the effect than is possible for high concentrations (trivially, if the global cosolvent concentration already exceeds 50% it is impossible to find this concentration doubled in the shell). Even for extremely high cosolvent concentrations, where the coexistence phase disappears completely, the concentration in the shell is still over 10% higher than in the bulk at temperature $T = 1.0$. As the temperature increases above $T_L$, the preferential binding decreases due to increasing entropy effects which favor mixing of the components.
4.5 Discussion

In practice, chaotropic substances are used frequently to destabilize folded proteins, micelles and aggregates of hydrophobic particles. Although the existence of the chaotropic effect is well known, its underlying physical mechanism remains controversial. In this study, the effect on aqueous solutions of hydrophobic particles is investigated by adapting the MLG-based model for water to include an idealized chaotropic cosolvent. The objective was to substantiate the idea that chaotropic substances affect primarily the structure of the water, and thus only indirectly the properties of the hydrophobic particles in the solution. By implementing different approaches to the problem, we have found that the adapted model consistently and successfully reproduces both the destabilizing effect on hydrophobic aggregates and the preferential binding of the cosolvents to the solute, without the involvement of any
Figure 4.15: Preferential binding for various cosolvent concentrations at temperature $T = 1.0$, calculated in the pair approximation. At low values of the chemical potential $\mu$, the preferential binding effect is more distinct than at high $\mu$ (\(\mu\) is the chemical potential associated with the insertion of water). At fixed temperature a lower particle density is associated with a smaller value of $\mu$. The values of $\Delta \mu$ are determined by the requirement that the cosolvent density remains constant.

direct interaction of cosolvent with non-polar particles.

Like many measurements of the closed-loop coexistence curve, including those for poly(ethylene glycol) in water [17], our results are obtained by calculations and simulations at constant volume. Mean-field calculations result in two clear phase transitions at a lower and an upper critical point. At temperatures below the lower critical temperature $T_L$, the three-component system forms one homogeneous phase where the forces between the hydrophobic particles are repulsive due to the cage-like arrangement of water molecules around the solute particles. The solution exhibits an aggregation phase transition at $T_L$, leading to a separation into a phase of high particle density and a phase of almost pure water.

Chaotropic cosolvents decrease the number of intact hydrogen bonds between water molecules. If their only impact on the system were the purely geometrical fact that they create a cavity in the solvent, the preferential binding to the solute could not be explained because there would be no essential change in the hydrophobic
surface exposed to water (cosolvents like urea are also primarily hydrophobic). An appropriate model for the interactions of cosolvent with water must therefore include their influence in breaking some of the hydrogen bonds, thus increasing the enthalpy and decreasing the number of ordered states relative to disordered ones, increasing the entropy. We have introduced these properties by reducing the number of ordered states with respect to the number of disordered states, and have shown that this increases the chemical potential associated with the insertion of hydrophobic particles into the solution, promoting their solubility. In addition, the hydrophobic interaction becomes more repulsive and therefore the coexistence region is reduced. For sufficiently high cosolvent concentrations this can even lead to an entirely homogeneous solution over the whole temperature range.

Monte Carlo simulations show that the cosolvent concentration in the shell is larger than in the bulk. This preferential binding reduces the number of possible intact hydrogen bonds in the solvation shell of a solute particle. Hence the number of disordered shell states increases, and at high temperatures a disaggregation of the particle clusters becomes entropically favorable before it would occur in the absence of cosolvent.

These results are confirmed by calculations within the pair approximation. Here again we observe that the cosolvent preferentially occupies the solvation shell of hydrophobic particles, which minimizes the free energy of the system. Both Monte Carlo simulations and the pair-approximation calculations demonstrate a substantially more pronounced preferential binding effect at lower cosolvent concentrations, reflecting the fact that the calculations consider the relative increase in concentration in the shell compared to the bulk, as opposed to the absolute increase. We emphasize that the cosolvents are pushed from the bulk water into the surrounding shell of the solute, rather than being bound to the latter, leading to a weaker net interaction between the solute molecules and the solvent.

The $\mu$-$T$ phase diagrams (Figs. 4.5, 4.8, 4.13) illustrate that the line demarcating the transition region moves to lower temperatures as the cosolvent concentration increases, and a shift of the line is found towards larger values of $\mu$ (the chemical potential associated with insertion of water sites). This reproduces the fact that the resistance of the system to insertion of water increases in the presence of chaototropic cosolvents. Because the particle density at the critical solution temperatures remains essentially independent of the cosolvent concentration, the resistance to water addition increases with growing cosolvent concentration.

Qualitatively, all three approaches yield good agreement for the coexistence curves, and for the transition lines of the $\mu$-$T$ phase diagram. The quantitative difference in position of the UCST is a consequence of the fact that mean-field calculations neglect large fluctuations, and thus often overestimate transition temperatures. The
agreement of the lower critical temperatures indicates the dominant role of strong local interactions, and a suppression of long-range effects, at low temperatures.

Destabilization of aggregates of hydrophobic particles is achieved experimentally only for extremely high concentrations of chaotropic substances. Thus an 8-molar urea solution, which consists of an equivalent volume fraction of urea and water, is required to double the solubility of the highly hydrophobic amino-acid phenylalanin at room temperature, and in Ref. [92] it was found that solubilities of non-polar gases in aqueous solutions of poly(propylene glycol) are raised by factors ranging from 2 to over 100 as the PPG concentration is increased beyond 50%. While within the current framework it is difficult to express the cosolvent concentration explicitly because the model contains more than one water molecule per site, the results are qualitatively correct. A cosolvent concentration of at least 50% is required to reduce significantly the extent of the solubility region.

The overall results demonstrate that the chaotropic effect on aqueous solutions of hydrophobic particles can be explained primarily by a reduction of the number of intact hydrogen bonds between water molecules and an increase of the number of broken ones in the vicinity of the cosolvent. We note that the cosolvent considered is an idealized, small molecule, without polar groups which could interact with water molecules in different ways. However, our results indicate strongly that the hydrophobic interaction and aggregation in the presence of chaotropic effects are indeed well described by such a model when the number of states of water molecules in the solution is modified accordingly.
Chapter 5

Kosmotropic Effect

Kosmotropic substances, such as sucrose and betaine, are known to decrease the solubility of hydrophobic particles and to stabilize their aggregates in aqueous solutions. These cosolvents are highly soluble, polar, uncharged in physiological conditions, and strongly enhance water structure due to their ability to form hydrogen bonds [21]. Their consequent preferential exclusion from the hydration shell of hydrophobic solute particles and proteins (Sec. 1.3) leads to their stabilizing function, and their importance for the osmotic balance in cells has generated a growing interest in the physical origin of the kosmotropic effect, which to date remains unclear [27, 46]. The number of water molecules in the direct vicinity of a solute particle thus increases, which leads to a stronger net repulsion between solute and solvent. In the presence of kosmotropic cosolvents, structural arrangement of the water-cosolvent mixture is enthalpically favorable compared to a cage-like organization around non-polar solute particles. Solute molecules are thus pushed together to minimize their total exposed surface, which results in an enhancement of hydrophobic aggregation. The same process leads to a stabilization of native protein configurations, in spite of the fact that kosmotropic substances have no net charge and do not interact directly with the proteins [21, 93].

In this study, we adapt the two-state Muller-Lee-Graziano model for water to describe the ternary system of water, kosmotropic cosolvents and hydrophobic particles in order to analyze the effect of kosmotropic substances on hydrophobic interactions. A mean-field approximation confirms the stabilizing effect, and Monte Carlo simulations demonstrate preferential exclusion of the kosmotropic cosolvents from the solvation shell of hydrophobic particles. The ability to enhance water-structure formation is shown to explain qualitatively the stabilizing effect of kosmotropic substances on aggregates of hydrophobic particles.
5.1 Model including Kosmotropic Cosolvent

Summarizing the description in the previous chapters, the driving force in the process of solvation and aggregation is the effective hydrophobic interaction between polar water and the non-polar solute particles [10]. This interaction arises from a rearrangement of water around the solute particle, a process which decreases the enthalpy due to reinforced hydrogen bonds between water molecules in comparison to bulk water. These physical features are described by the model of Muller, Lee, and Graziano (MLG), where the energy levels of water molecules (or groups of water molecules) and their respective degeneracies are expressed in terms of local water structure. Water is divided into two different types, shell and bulk water. The difference between solvent molecules and non-polar solute particles lies in their ability to form hydrogen bonds. The continuous range of interaction energies may be simplified to two discrete states of predominantly intact or broken hydrogen bonds [65]. We have extended the model in Chapt. 4 to yield a successful description of ternary systems containing water, hydrophobic particles, and chaotropic cosolvents, which show preferential binding to non-polar solute particles, and which have a destabilizing effect on hydrophobic aggregates, native proteins, and micelles. In this chapter we extend the model to describe ternary systems containing water, hydrophobic particles, and kosmotropic cosolvents.

Kosmotropic cosolvents such as sucrose increase the number of possible intact hydrogen bonds, and thus increase the number of possible structural arrangements of bulk water [44]. It is therefore energetically favorable to maximize the concentration of kosmotropic cosolvent in bulk water with respect to shell water, and hence preferential exclusion is expected. The increased number of water molecules in the solvation shell, or direct surroundings, of non-polar solute particles increases the repulsion between solute and solvent, resulting in a reduction of solubility and an enhanced attraction between solute particles due to the decreased interface with water as they approach each other. Hydrophobic solute particles are generally larger than water, and thus a water site in the model consists of a group of molecules. Furthermore, cosolvent particles are included directly in water sites by changing the number of states of these sites. A site containing water molecules and a cosolvent particle is referred to as a cosolvent site. We incorporate this increased number of intact hydrogen bonds by raising the number of possible ordered states compared to the number of disordered states, \( q_{os,k} = q_{os} + \eta_s \), \( q_{ds,k} = q_{ds} - \eta_s \), and \( q_{ob,k} = q_{ob} + \eta_b \), \( q_{db,k} = q_{db} - \eta_b \), where \( k \) denotes the states of water in contact with kosmotropic cosolvent, and the total number \( q \) of states is kept constant. The energy levels of the states remain unchanged because the cosolvent is taken to affect only the number of intact hydrogen bonds, but not their strength (Fig. 5.1).

The ternary system of \( N \) sites containing either pure water \( (n_i = 1) \), hydrophobic
Figure 5.1: Illustration of the transformation of configurations from disordered to ordered by the addition of kosmotropic substances due to the enhancement of hydrogen-bond formation, which leads to an increasing number of ordered states of both shell and bulk water. The effect is much stronger in the bulk than in the shell.

particles \(n_i = 0\) or cosolvent \(n_i = -1\), is described by the Hamiltonian

\[
H[\{n_i\}, \{\sigma_i\}] = \sum_{i=1}^{N} \frac{n_i(n_i+1)}{2} [(E_{ob}\delta_{i,\sigma_{ob}} + E_{db}\delta_{i,\sigma_{db}})\lambda_i + (E_{os}\delta_{i,\sigma_{os}} + E_{ds}\delta_{i,\sigma_{ds}})(1-\lambda_i)] \\
+ \sum_{i=1}^{N} \frac{n_i(n_i-1)}{2} [(E_{ob}\delta_{i,\sigma_{ob,h}} + E_{db}\delta_{i,\sigma_{db,h}})\lambda_i + (E_{os}\delta_{i,\sigma_{os,h}} + E_{ds}\delta_{i,\sigma_{ds,h}})(1-\lambda_i)], \tag{5.1}
\]

where \(\lambda_i\) is defined as the product of the nearest neighbors, \(\lambda_i = \prod_{j \neq i} n_j^2\), and takes the value 1 if site \(i\) is completely surrounded by water and cosolvent or 0 otherwise.

The first sum defines the energy of pure water sites, and the second the energy of cosolvent sites. Because a water site \(i\) may be in one of \(q\) different states, \(\delta_{i,\sigma_{ob}}\) is 1 if site \(i\) is occupied by water in one of the \(q_{os}\) ordered shell states and 0 otherwise, and \(\delta_{i,\sigma_{ds}}\) is 1 if it is occupied by pure water in one of the \(q_{ds}\) disordered shell states and 0 otherwise. Analogous considerations apply for the bulk states and for the states of cosolvent sites. A more detailed description of the model is presented in Chaps. 2 and 4.

The canonical partition function of the ternary system of \(N\) sites is the sum over the state configurations \(\{\sigma_i\}\) and yields

\[
Z_N = \sum_{\{n_i\}} \prod_i Z_s^{\frac{n_i(n_i+1)}{2}} \left(1-\lambda_i\right)^{\frac{n_i(n_i+1)}{2}} \lambda_i \prod_{j,k} Z_{s,k}^{\frac{n_j(n_j-1)}{2}} \left(1-\lambda_j\right)^{\frac{n_j(n_j-1)}{2}} \lambda_j, \tag{5.2}
\]

where \(Z_\sigma = q_{\sigma,o}e^{-\beta E_{\sigma,o}} + q_{\sigma,d}e^{-\beta E_{\sigma,d}}\) for the shell (\(\sigma = s\)) and bulk (\(\sigma = b\)) states of
pure water sites and cosolvent sites ($\sigma = s, k$ and $\sigma = b, k$).

When the number of particles may vary, a chemical potential is associated with the energy of particle addition or removal. The grand canonical partition function of the system for variable particle number becomes

$$\Xi = \sum_{N} e^{\beta \mu N_{w} + \beta (\mu + \Delta \mu) N_{k}} Z_{N} = \sum_{\{n_{i}\}} e^{-\beta H_{eq}^{sc}([n_{i}])}, \quad (5.3)$$

where $\mu$ represents the chemical potential associated with the insertion of water and $\Delta \mu$ the chemical potential for the insertion of a cosolvent molecule at a water site. $N_{w}$ denotes the number of pure water sites, $N_{k}$ the number of cosolvent sites, and $N_{p}$ the number of solute particle sites, the total number of sites being $N = N_{w} + N_{k} + N_{p}$.

It was shown in Chapt. 4 that the hydrophobic repulsion between water and non-polar solute particles decreases in the presence of destabilizing (chaotropic) cosolvents. By analogy, in the presence of kosmotropic cosolvents this effective repulsion increases leading to increasing hydrophobic attraction between non-polar solute particles, as may be observed on rewriting the partition function for particle sites (Eq. (4.12)).

### 5.2 Methods

To obtain preliminary qualitative predictions, we assume spatial fluctuations of the order parameter to be insignificant in the ternary system and perform a mean-field analysis of the thermodynamic properties. The local density for each site is approximated by the mean density of the system using the densities of hydrophobic particles $\rho_{p}$, pure water $\rho_{w}$, and cosolvent $\rho_{k}$, where $\rho_{p} + \rho_{w} + \rho_{k} = 1$. The mean occupancy of site $i$ is then $\langle n_{i} \rangle = \sum_{\sigma} n_{i,\sigma} \rho_{\sigma} = \rho_{w} - \rho_{k}$, and analogously $\langle n_{i}^{2} \rangle = \sum_{\sigma} n_{i,\sigma}^{2} \rho_{\sigma} = \rho_{w} + \rho_{k}$. One then minimizes the grand canonical mean-field free energy per site

$$f = (\mu - \beta^{-1} \ln Z_{s}) \rho_{w} + (\mu + \Delta \mu - \beta^{-1} \ln Z_{s,k}) \rho_{k} + \beta^{-1} (\ln Z_{s} - \ln Z_{b}) \rho_{b} (\rho_{k} + \rho_{w})^{z} + \beta^{-1} (\ln Z_{s,k} - \ln Z_{b,k}) \rho_{k} (\rho_{k} + \rho_{w})^{z} + \beta^{-1} (\rho_{w} \ln \rho_{w} + \rho_{k} \ln \rho_{k} + \rho_{p} \ln \rho_{p}), \quad (5.4)$$

where $z$ is the number of nearest neighbors of each site. This coordination number can be interpreted as the average number of non-polar solute particles in contact with any other, and thus represents a quantity relevant for characterizing the effective hydrophobic interaction between solute particles. Because mean-field approximation neglects local density fluctuations, we further perform Monte Carlo simulations described in Sec. 2.5.
Figure 5.2: Closed-loop coexistence curves for a ternary system of water, cosolvent and hydrophobic particles for different cosolvent densities $\rho_k$, obtained by mean-field calculation. On the outside of each curve the solution is homogeneous, and on the inside the solution separates into two phases. The dotted arrow represents the heating process of a system with particle density $\rho_0$ and cosolvent density $\rho_k,2 = 0.61$ from temperature $T_0$ in the homogeneous region to $T_1$ in the heterogeneous region. At $T_1$ the system is separated into two phases of different densities $\rho_1$ (nearly pure water) and $\rho_2$ (hydrophobic aggregates).

For the calculations to follow we have used the parameter values for the energies and degeneracies as discussed in Chapt. 2, which have been successful in describing different solutions (Chapts. 3, 4). The values $\eta_b = 9.0$ and $\eta_a = 0.1$ are taken to be suitably representative of a ternary system containing kosmotropic cosolvents (Chapt. 4).

### 5.3 Results and Discussion

Fig. 5.2 shows the closed-loop coexistence curve in the $\rho_p-T$ phase diagram for different cosolvent concentrations. As expected, the system is a homogeneous particle-solvent-cosolvent mixture below a LCST and above an UCST for all concentrations, whereas between these temperatures and for intermediate particle concentrations a
Kosmotropic Effect Results and Discussion

Figure 5.3: Closed-loop coexistence curve measured in the experimentally accessible temperature range for N,N-Diethylmethylamine in water in the presence and in the absence of the kosmotropic cosolvent sodium chloride, reproduced from Ref. [94]. The LCST is reduced from 51°C in the salt-free case to −0.6°C in the saturated cosolvent solution.

Phase separation is found into a pure solvent phase and an aggregate phase with fixed solute density. The temperature and density ranges of the aggregation region grow with increasing cosolvent concentration. Further, the particle density of the aggregate phase in the two-phase region increases when adding cosolvent, demonstrating the stabilizing effect of kosmotropic cosolvents and the strengthening of hydrophobic interactions between solute particles due to a growing water-solute repulsion. This is illustrated by the process of heating a system at constant density (Fig. 5.2). The dotted arrow represents a solution with particle density $\rho_0$, which is heated from temperature $T_0$ in the homogeneous region to a temperature between the LCST and the UCST. In the heterogeneous region ($T_1$), the solution separates into two phases of densities $\rho_1$ (almost pure water) and $\rho_2$ (hydrophobic aggregates) under the constraint $\rho_0(V_1 + V_2) = \rho_1V_1 + \rho_2V_2$, where $V_i$ is the volume occupied by phase $i$. An increase in cosolvent density results in a higher particle density $\rho_2$ of the hydrophobic aggregates and a lower density $\rho_1$, showing a clear strengthening of the hydrophobic interaction between solute particles.

The phase diagram obtained by mean-field calculation is in qualitative agreement
Figure 5.4: Closed-loop miscibility curves for an aqueous solution of hydrophobic particles with different cosolvent concentrations $\rho_k$, obtained by Monte Carlo simulations. Outside the curve the particles are soluble and the system is homogeneous, while inside it two phases are found, namely aggregates of hydrophobic particles and a nearly pure solvent-cosolvent mixture. As the cosolvent concentration grows, the solubility of the particles decreases, leading to an expanded region of coexistence.

with the experimentally determined coexistence curves for N,N-Diethylmethylamine in pure water, and in a mixture of water and the kosmotropic cosolvent sodium chloride, as shown in Fig. 5.2. The presence of sodium chloride causes a dramatic decrease of the LCST from $51^\circ C$ to $-0.6^\circ C$ for a solution saturated with sodium chloride [94]. Below $-0.6^\circ C$ only one homogeneous liquid phase is observed.

The hydrophobic aggregation in a certain temperature range is further confirmed by Monte Carlo simulations, presented in Fig. 5.3. Below the LCST and above the UCST, one homogeneous mixture is found where the hydrophobic particles are dissolved. Between these temperatures the solute particles form aggregates of a given density. As in Chaps. 3 and 4 the LCST and the critical aggregate densities are in good agreement with the mean-field calculation, whereas the UCST determined by Monte Carlo simulations lies at a temperature lower than the mean-field result. We remind the reader that this is to be expected because mean-field calculations neglect fluctuation effects, generally overestimating both transition temperatures, and that the good agreement at low temperatures indicates local effects in this region and that
Figure 5.5: \( \mu-T \) phase diagram of hydrophobic solute in an aqueous cosolvent mixture for different cosolvent concentrations \( \rho_k \), obtained by Monte Carlo simulations. The endpoints of the finite transition lines correspond to the UCST and LCST.

The \( \mu-T \) phase diagram of the ternary system obtained by Monte Carlo simulations is presented in Fig. 5.5. Lines of finite length terminated by the UCST and LCST demarcate the aggregation phase transition. An increase in cosolvent density leads to a larger separation of UCST and LCST, and the transition line is shifted to higher \( \mu \) values, confirming the increased resistance of the system to addition of water at constant volume in the presence of cosolvent. This is a consequence of the fact that the resistance to addition of hydrophobic particles depends on both \( \mu \) and \( \Delta \mu \). With increasing cosolvent density, \( \Delta \mu \) decreases and hence \( \mu \) must increase for the critical particle density to remain constant.

The Monte Carlo simulations confirm the expectation of a lower cosolvent concentration in shell water than in bulk water. Fig. 5.6 shows the relative cosolvent shell concentration compared to the overall cosolvent concentration in the solution, where preferential exclusion of the kosmotropic cosolvent from the solvation shell of the hydrophobic solute particles is observed. A sharp drop occurs at the phase transition temperature, which is due to a sudden change in solute particle density. At very low particle density, a clear exclusion of the cosolvent from the solvation shell.
Kosmotropic Effect Results and Discussion

Figure 5.6: Density of kosmotropic cosolvent in the solvation shell of hydrophobic particles, obtained from Monte Carlo simulations at chemical potential $\mu = 2.5$, exhibiting preferential exclusion. The effect is more pronounced for low cosolvent densities because it is the relative decrease in cosolvent density in the shell compared to the bulk which is shown. At high particle density, the preferential exclusion effect is small because most solvent and cosolvent sites are shell sites, leading to a shell density which is almost equal to the total cosolvent density.

appears. However, at high particle density (for temperatures below the aggregation phase transition) the effect is only marginal, because here the number of particle sites becomes substantial, most solvent and cosolvent sites are shell sites, and thus the total cosolvent density is almost identical to the shell cosolvent density. The strong fluctuations at temperatures above the phase transition may be attributed to the very low particle density in the system. Small, thermal fluctuations in the number of shell sites occupied by cosolvent molecules then result in large fluctuations in the cosolvent shell concentration. This effect increases with temperature due to stronger entropy effects. The system shows a substantially stronger preferential exclusion effect for low than for high cosolvent densities, which is largely a consequence of considering the relative, as opposed to absolute, decrease in cosolvent concentration in the shell. Preferential exclusion is simply less pronounced for high cosolvent densities because a smaller proportion of the total number of cosolvent sites can contribute to the effect. At high temperatures, entropy effects become dominant and the preferential exclusion of cosolvent particles shows a slight decrease.
Preferential exclusion of kosmotropic cosolvents from the immediate vicinity of hydrophobic particles in aqueous solutions is measured in experiments only for very high cosolvent concentrations. Concentrations of compatible osmolytes, such as sucrose and betaine in the cytoplasm, typically reach values well in excess of 0.5M, and the stability of the protein lactate dehydrogenase against thermal denaturation increases by about 90% in a 1M sucrose solution at room temperature [27]. In the current model it is difficult to determine the cosolvent concentration because one site contains a group of water molecules, but the results are nevertheless qualitatively correct. We have found that cosolvent concentrations of more than 10% are required to show a clear stabilization effect on hydrophobic aggregates (Fig. 5.3).

In summary, we have studied the physical mechanism underlying the stabilizing effect of kosmotropic cosolvents on hydrophobic aggregates in aqueous solutions. By altering the state degeneracies we include the ability of kosmotropic substances to enhance the structure of liquid water, and have found stabilization of aggregates, expansion of the coexistence regime, and preferential exclusion of the cosolvent from the hydration shell of hydrophobic solute particles. We have shown that this preferential exclusion arises as a consequence of the altered shell and bulk degeneracies which result from the energetically favorable enhancement of bulk water structure due to strong hydrogen-bond formation between solvent and cosolvent molecules. As a result, the cosolvent density increases in the bulk and decreases in the shell, causing a relative increase in shell water density. This high shell water density can be considered to increase the hydrophobic effect, thus stabilizing solute aggregates.

The overall results demonstrate that the stabilizing effect may be explained to a large extent purely by the propensity of kosmotropic cosolvents to enhance water structure. The properties of aqueous solutions of hydrophobic particles including cosolvents are described successfully by our extension of the MLG model, which focuses on the formation of two distinct types of solvent and on changes in the structural arrangements of liquid water.
Chapter 6

Micelle Formation

In an aqueous solution, the hydrophobic parts of amphiphilic molecules tend to separate themselves from water molecules by forming aggregates, such as micelles and microemulsion droplets. The simplest possible structure occurs if the polar and hydrophobic parts of the amphiphilic molecule are well separated into head and tail regions. For this type of molecule, to which we will refer as surfactant by adopting the notation of Ref. [20], micelles consist of a hydrophobic core which contains all the tails. Amphiphilic species whose polar and hydrophobic regions are distributed over the entire molecule, rather than being clearly separated, may aggregate to form assemblies which minimize the hydrophobic area per molecule exposed to the aqueous phase. We will refer to this general category of mixed HP molecules we consider as ‘amphiphiles’. Poly(N-isopropylacrylamide), a polymer belonging to this class of molecule, exhibits a phase transition at a lower critical solution temperature (LCST) from a homogeneous solution, where the polymers are completely soluble, to a system of two separated phases [51]. At an upper critical solution temperature (UCST) the organic phase disaggregates, and above this temperature the amphiphilic molecules are again soluble due to entropic effects [52]. Substitution of polar by hydrophobic monomers in amphiphiles of given length leads to alterations of the critical solution temperatures which depend on the size of hydrophobic surface regions [52].

Micelle formation, and the aggregation of amphiphilic molecules in general, may be treated as a phase separation occurring at a critical micellar concentration (CMC) which describes the density of amphiphiles where the system enters the two-phase region [53, 54]. Above the CMC, amphiphilic molecules in aqueous solutions self-associate, forming small aggregates to decrease the net contact between the hydrophobic parts of the molecules and the surrounding solvent. A small fraction of the surfactant molecules remain free in the solution, with a concentration close to the CMC value. As the concentration increases, the onset of a semi-dilute regime is found, where the system may be considered as a solution of relatively few water molecules dissolved in an amphiphilic medium [55].

The CMC decreases as the length of the hydrophobic chain increases [4], indicating
Micelle Formation Hydrophobic-Polar Model for Amphiphiles

A stabilization of aggregates as a consequence of the stronger net repulsion between the hydrophobic tail of an amphiphilic molecule and the surrounding water. For the same reason, the LCST is thought to decrease as the tail grows, as suggested in Ref. [52]. One means of probing the nature of the effective hydrophobic interactions would be by systematic alteration of the polarity of the amphiphilic polymers in solution. Replacement of hydrophobic monomers by polar ones within the polymer chain may then be expected to cause changes in the phase diagram. Many theoretical and experimental studies have been carried out to describe the various aggregation phases of amphiphilic molecules in aqueous solutions [4, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63]. However, no comprehensive investigations have yet been performed concerning the influence of the distribution of polar groups in amphiphiles on the aggregation phase diagram, or concerning the mechanism underlying the process of self-aggregation.

The aim of this study is to substantiate preliminary experimental results indicating a decrease in LCST as the degree of hydrophobicity increases, and to analyze the dependence on density and hydrophobicity of the aggregation phase diagram. We will conclude that the principal properties of amphiphiles in aqueous solutions are solvent-induced, in that they are explained by alterations in the formation of hydrogen bonds in liquid water in the vicinity of the hydrophobic regions of solute particles. We begin by introducing a simple hydrophobic-polar (HP) description of amphiphilic solute particles (Sec. 6.1) on a cubic lattice, and then extend the model to describe surfactant particles of varying tail length (Sec. 6.2). We investigate the changes in the phase diagram associated with an increasing proportion of polar groups in the amphiphilic molecules and with changes in their distribution, and present similar analysis for the formation of micelles of surfactants in aqueous solutions.

6.1 Hydrophobic-Polar Model for Amphiphiles

Summarizing the concepts in Chapt. 1 and 2, the solvation of amphiphiles in aqueous solutions and their self-association into micellar aggregates are due primarily to the hydrophobic interaction between polar water and the non-polar parts of the solute molecules. Water molecules have the ability to form strong hydrogen bonds both among themselves and with the polar groups of the amphiphilic solute molecules (if they have the ability to form hydrogen bonds), which can lead to extended hydrogen-bonded networks. Although the insertion of a hydrophobic molecule leads to a destruction of local hydrogen-bond structure, and hence to considerable entropy gain and enthalpy loss, at low temperatures water molecules are found to rearrange in a cage-like structure around the solute molecule formed by even slightly stronger hydrogen bonds. The net reduction of energy during the insertion process results in dissolution of the solute particle. At higher temperatures, however, hydrophobic solute particles are found to aggregate, minimizing their total surface exposed to
Micelle Formation

Hydrophobic-Polar Model for Amphiphiles

water. This process is driven by the concomitant minimization of local restructuring of water, which allows additional entropic contributions from the solvent to lower the free energy. At still higher temperatures entropy effects due to the many possible arrangements of water molecules are dominant favoring once again a homogeneous mixture. The effective hydrophobic interaction between hydrophobic solute molecules is thus generally accepted to be primarily solvent-induced, i.e. to be a consequence of changes in the ordering of water molecules, rather than being explained by water-solute interactions.

These physical considerations may be described by the concept of the solvent forming two physically distinct types of state, shell (hydration) water and bulk water, and is contained in the MLG model. The distinction between solvent particles and non-polar solute molecules lies in their ability to form hydrogen bonds. Because solute particles are relatively large compared with single water molecules, we use an adapted version of the MLG model in which each site contains either a solute particle or a group of water molecules. The continuous range of interaction energies within this group may be simplified to two discrete states (bimodal approximation), namely those with predominantly intact or broken hydrogen bonds [65]. Water sites in the coarse-grained model may then be characterized by two states, where an “ordered” site represents a water cluster with mostly intact hydrogen bonds, while a “disordered” site contains relatively fewer intact hydrogen bonds. We have shown that this extended model contains implicitly the many-body interactions between hydrophobic particles, and that it exhibits hydrophobic aggregation between the upper (UCST) and lower (LCST) critical solution temperatures (Chapt. 3).

To include not only purely hydrophobic solute particles but also amphiphilic molecules with varying conformations of polar and non-polar regions, as in the experiments in Ref. [51], we represent the particles as cubes on which each face may be either polar or hydrophobic. A neighboring water site, which is homogeneously polar, interacts only with the side of the particle which is pointing in its direction. If this side represents a polar group, the neighboring water site is in a bulk state, whereas if the side is hydrophobic, the water site is considered a shell site. Thus polar sides and water are considered as having the same effect on neighboring water. However, only water sites but not polar faces contribute to the free energy, which is represented by the sum over water sites (Eq. 6.1). In this coarse-grained model, one side of a site may represent more than one chemical group, and thus corresponds to a net characterization of the surface area of the solute molecule under consideration.

Hydrophobic solute particles are generally larger than water, and thus a water site in the model consists of a group of molecules. On a lattice where each site has $z$ nearest neighbors, the energy of a system of $N$ sites, occupied either by amphiphilphiles

87
(n_i = 0) or by water (n_i = 1), is given by the Potts-like Hamiltonian

\[ H\{\{n_i\}, \{\sigma_i\}\} = \sum_{i=1}^{N} n_i [E_{os}\delta_{i,\sigma_{os}} + E_{ds}\delta_{i,\sigma_{ds}}] \left( \frac{1}{z} \sum_{(ji)} \left( 1 - \lambda_{i,j(k)} \right) \right) \]

\[ + \left( E_{ob}\delta_{i,\sigma_{os}} + E_{db}\delta_{i,\sigma_{ds}} \right) \left( \frac{1}{z} \sum_{(ji)} \lambda_{i,j(k)} \right), \tag{6.1} \]

where \( \lambda_{i,j(k)} \) is a side variable depending on the nearest neighbor \( j \) of site \( i \), whose orientation is specified by the variable \( k \) of the neighboring site \( j \). The variable \( \lambda_{i,j(k)} \) takes the value 1 if the neighboring face of site \( j \) a polar face or a water face and 0 otherwise. The orientation variable \( k \) varies from 1 to \( n_{c,j} \), where \( n_{c,j} \) takes properly into consideration all possible rotations of site \( j \), and \( n_{c,j} = 1 \) for water and solute particles with \( N_H = 6 \) or \( N_H = 0 \), \( n_{c,j} = 6 \) for \( N_H = 5 \) or \( N_H = 1 \), \( n_{c,j} = 3 \) for \( N_H = 4 \) (\( N_H = 2 \)) if the two polar (hydrophobic) sides are opposite from each other, \( n_{c,j} = 12 \) for \( N_H = 4 \) (\( N_H = 2 \)) if they are adjacent, \( n_{c,j} = 12 \) for \( N_H = 3 \) if the three polar sides are all adjacent to each other, and \( n_{c,j} = 8 \) for \( N_H = 3 \) if two of them are opposite from each other.

Because a water side may be in one of \( q \) different states, \( \delta_{i,\sigma_{os}} \) is 1 if it is one of the \( q_{os} \) ordered shell states and 0 otherwise, and \( \delta_{i,\sigma_{ds}} \) is 1 if it is water in one of the \( q_{ds} \) disordered shell states and 0 otherwise. Analogous considerations apply for the bulk states.

We note that for completely hydrophobic solute particles (\( N_P = 0 \)) the Hamiltonian in Eq. 6.1 is not equivalent to the Hamiltonian in Eq. 2.2, because the sites are treated differently. In Eq. 6.1 each face of a water site \( i \) contributes to the free energy, and a sum over neighboring faces is performed instead of the sum over sites in Eq. 2.2. This description is consistent with the calculations for surfactant molecules in Sec. 6.2.

To determine the canonical partition function, the sum over the state configurations \( \{\sigma_i\} \) is performed. By taking into account the possible rotations of amphiphilic particles through the variable \( k \) the canonical partition function of the system of \( N \) sites yields

\[ Z_N = \sum_{\{n_i\}} \prod_{i=1}^{N} \prod_{k=1}^{n_{c,i}} \frac{1}{Z_a} \sum_{(ji)} \lambda_{i,j(k)} \frac{1}{Z_s} \sum_{(ji)} \left( 1 - \lambda_{i,j(k)} \right), \tag{6.2} \]

where \( Z_a = q_{\sigma,0}e^{-\beta E_{\sigma,0}} + q_{\sigma,1}e^{-\beta E_{\sigma,1}} \) and \( Z_s = q_{\sigma,0}e^{-\beta E_{\sigma,0}} + q_{\sigma,1}e^{-\beta E_{\sigma,1}} \) for the shell and bulk states of pure water sides, and \( N_w \) denotes the number of pure water sites. The partition function in Eq. 6.2 takes into consideration the constraints for local configurations which occur due to the orientational degree of freedom in the system. When the number of particles may vary, a chemical potential is associated with the energy of particle addition or removal. The grand canonical partition function of the system for variable particle number is

\[ \Xi = \sum_{\{n_i\}} e^{\beta \mu N_w} Z_N = \sum_{\{n_i\}} e^{-\beta H_{en}[\{n_i\}]} \tag{6.3} \]
where $\mu$ represents the chemical potential associated with the insertion of water. Although the explicit terms of the model describe solely the states of water molecules in solution, it contains implicitly both two- and even many-particle interactions between hydrophobic solute molecules.

The coexistence regions are characterized by measuring the UCST and the LCST for various numbers $N_P$ of polar sides per particle. On a cubic lattice, the total number of sides of a solute particle is $z = 6 = N_P + N_H$, where $N_H$ is the number of hydrophobic faces of the solute cube. To investigate variations in the effective hydrophobic interactions due to changes in polarity of the solute particles, we increase systematically the number of polar sides per particle and determine the coexistence region in each case.

To measure the extent of aggregate formation in the system, we determine the number of contacts between two hydrophobic sides, $H-H$, between two polar sides $P-P$, and between a polar and a hydrophobic side, $P-H$. The number of contacts in a randomly distributed system with the same particle density $\rho_p$ is also calculated for comparison. In a random solution of solute particles with $N_P$ polar faces, whose positions and orientations are completely independent, in the thermodynamic limit these probabilities are

$$
\begin{align*}
  p(H-H) &= p(H)^2 \\
  p(P-P) &= p(P)^2 \\
  p(P-H) &= 2p(H)p(P),
\end{align*}
$$

where $p(H)$ is the probability that a cube face is hydrophobic, $p(P)$ is the probability that it is polar, and by symmetry $p(H-P) = p(P-H)$. In a random system, the probabilities of occurrence of the different faces are independent of the neighboring sites, and are given by

$$
\begin{align*}
  p(H) &= \frac{N_H}{z}\rho_p, \\
  p(P) &= (1 - \rho_p) + \frac{N_P}{z}\rho_p.
\end{align*}
$$

If the density of $H-H$ contacts, $n_{H-H}$, and the density of $P-P$ contacts, $n_{P-P}$, are larger than their probabilities of random occurrence, and the density of $P-H$ contacts $n_{P-H}$ is correspondingly smaller than its random expectation value, the system has formed aggregates which reduce the number of hydrophobic sides exposed to water.

### 6.1.1 Method

Our interest is focused on the orientation and location of amphiphilic molecules in solution, which may be captured by molecular-level simulations. We thus restrict
our considerations in this chapter to Monte Carlo studies, because the processes involved depend strongly on local, spatial effects which are neglected in mean-field calculations. Monte Carlo simulations constitute an appropriate method for efficient calculation of thermal averages in many-particle systems with statistical fluctuations, and we use them here to detect the aggregation of amphiphilic solute particles as a function of temperature and solute concentration. We work with a system of $N = 27000$ sites on a cubic lattice with random initial particle distributions and periodic boundary conditions to eliminate boundary effects, using a Metropolis algorithm for sampling of the configuration space. The number (100 000) of relaxation steps for the system to achieve thermal equilibrium is similar to that found in our previous studies, as is the number (1 000 000) of measurements over which averages should be taken to estimate thermodynamic quantities. Every $20^{th}$ step is an attempt to add or remove a particle and the others are attempts to rotate a particle.

The closed-loop coexistence curves in the $\rho_p$-$T$ phase diagram are obtained from the transitions determined by increasing the temperature at fixed chemical potential $\mu$ (grand canonical sampling), which results in a sudden density jump at the transition temperature, and from the corresponding solute particle densities.

### 6.1.2 Results

We begin by attempting to capture the qualitative features of the behavior of amphiphilic molecules with varying conformations of polar and non-polar regions. Using the simple model described in Sec. 6.1, where solute particles are represented as cubes whose face may be either polar or hydrophobic, the hydrophobic sides are substituted systematically by polar ones.

Fig. 6.1 shows the $\rho_p$-$T$ phase diagram for different numbers of polar sides per particle. In the coexistence regime the amphiphilic particles aggregate, minimizing the contact of their hydrophobic regions with water and polar solute segments. Outside this region the amphiphiles are soluble at all densities below the LCST and above the UCST, and the system forms a homogeneous solution. For temperatures between these values, the amphiphiles are soluble only at very low densities, while at the CMC $\rho_1$ the particles aggregate and the system separates into two phases: nearly pure water (of particle density $\rho_1$) and an organic phase of density $\rho_2$. With increasing polarity, the solubility of the solute particles is enhanced and the CMC increases.

As expected, the coexistence region is reduced as the number of polar sides increases. If all sides of the solute particles are hydrophobic ($N_P = 0$ and $N_H = 6$), the system represents a solution of purely hydrophobic particles and the coexistence curve is in agreement with that of Chapt. 3. Substitution of one hydrophobic side per particle by a polar side ($N_P = 1$ and $N_H = 5$) leads to a decrease of the UCST.
Figure 6.1: $\rho_p$-$T$ phase diagram for micelle formation in the 3D, HP model for different numbers of polar sides per solute particle. The coexistence region is reduced and aggregation suppressed, as the number of hydrophobic sides substituted by polar sides increases. At a given temperature the system is homogeneous for solute particle concentrations below the CMC $\rho_1$, while above this value it separates into two phases of densities $\rho_1$ and $\rho_2$. 

and a slight increase of the LCST, and to an overall suppression of the temperature and density range of the coexistence region. In this case, which may be taken to represent simplified surfactant molecules (see Fig. 6.9), the effect is rather moderate, indicating that the substitution of one hydrophobic side by a polar one increases the solubility of the solute particles only marginally. However, substitution of a second hydrophobic side per particle ($N_{P} = 2$ and $N_{H} = 4$) reduces the coexistence region dramatically (Fig. 6.1). The solubility of solute particles with two polar sides is thus much higher than that of those with only one polar side. When substituting three or more hydrophobic sides by polar ones, the solute particles become soluble at all temperatures and no aggregation phase transition is observed.

Figures 6.2 and 6.3 illustrate the nature of the aggregated phase using “snapshots” of two-dimensional (2D) systems at $T = 1.0$, obtained in the coexistence region for solute particles with different numbers of polar sides. The snapshots are taken after allowing the system to relax for $1000000$ steps, where every $20^{th}$ step is an attempt to exchange two sites and the others are attempts to rotate a particle. Systems containing
Figure 6.2: Snapshots of 2D systems in the coexistence phase, obtained by Monte Carlo simulations at $T = 1.0$. Left: completely hydrophobic solute particles in water ($N_P = 0$); right: mainly hydrophobic solute particles in water ($N_P = 1$). White circles (right) represent the polar sides of the solute particles, which are shown as black squares. The particles form compact micelles which shield the hydrophobic sides from water.

Figure 6.3: Snapshots of 2D systems of partially hydrophobic solute particles ($N_P = 2$) in water in the coexistence phase, obtained by Monte Carlo simulations at $T = 1.0$. White circles represent the polar sides of the particles, which are shown as black squares.

primarily hydrophobic particles ($N_P = 0$ and $N_P = 1$) form mostly compact clusters, which minimize the number of hydrophobic surfaces exposed to the solvent. For surfactant-like solute particles with $N_P = 1$ [Fig. 6.4(a)], the formation of perfect micelles with a hydrophobic core and a polar surface is prevented by the nature of the square lattice, which causes frustration on the edges of the micelles. An edge particle, which is in contact only with two other particles, is forced to expose one of
its hydrophobic sides to water. The model allows a surfactant particle to occur in the core of micelles, where its polar side is in direct contact with the polar side of another surfactant particle, because no distinction is made between a group of water molecules and the polar face of a particle. Incorporating this distinction into the description could be expected to generate more realistic micellar structures, albeit within the confines of the cubic geometry.

Amphiphilic solute particles whose surface is half polar \((N_P = N_H = 2\) in 2D) show differing behavior depending on the polarity pattern of the sides. If the polar sides are adjacent on the square, small micelles consisting of four solute particles can be formed, which is energetically the most favorable configuration because no hydrophobic sides are exposed to water (Fig. 6.3). Short, diagonal lines of molecules, which may be considered to represent condensed bilayers, can also be formed, although their ends are hydrophobic, and this configuration is therefore less favorable than are “circular” micelles of four solute particles. These configurations are expected from the construction of the sites, which is shown in Fig. 6.4, to appear as the ground states. Solute sites with two adjacent polar sides may be considered to represent sections of circular micelles [Fig. 6.4(b)], with the formation of small micelles as a consequence. In contrast, if the polar sides are opposite each other, the only possibility to avoid hydrophobic sides being in contact with water is to form lines of particles, although the hydrophobic ends remain exposed to water. These squares may be taken as schematic representations of cross-sections of a bilayer [Fig. 6.4(c)].

Figure 6.4: Illustration of schematic analogs obtained using square particles with different arrangements of polar sides in 2D. Hydrophobic sides of the square solute sites are shown in black, polar sides in red. a) A cube with one polar side, \(N_P = 1\), may be taken to represent a single surfactant molecule. b) Solute sites which contain two adjacent polar sides, \(N_P = 2\), may be considered as segments of a small micelle. c) A solute site with two opposite polar sides, \(N_P = 2\), may be taken to represent a part of the cross-section of a bilayer.

The contact densities \(n_{H-H}\), \(n_{P-P}\), and \(n_{P-H}\), are shown in Fig. 6.5 for a system
of solute particles with one polar side ($N_P = 1$) at different temperatures. The results are normalized to the probability of these contacts in a randomly distributed system with the same particle density, $\rho_p = 0.24$. At low temperatures, the solute particles are clearly soluble, because $n_{H-H} < p(H-H)$, $n_{P-P} < p(P-P)$, and $n_{P-H} > p(P-H)$. This reflects the fact that water forms strongly hydrogen-bonded, cage-like arrangements around hydrophobic particles when entropy effects are minor, resulting in a net energy gain due to dissolution. The number of $H-H$ contacts is thus smaller than in a random system while the number of $P-H$ contacts is larger. At temperatures higher than the lower critical temperature for density $\rho_p$, the solute particles aggregate and micelles are formed. In this regime, the hydrophobic sides form many mutual contacts to avoid the proximity of water sites, and of the polar sides of other solute particles, whence $n_{H-H} > p(H-H)$, $n_{P-P} > p(P-P)$, and the density of polar-hydrophobic contacts is suppressed, $n_{P-H} < p(P-H)$. At high temperatures where entropy effects dominate, $n_{H-H}$ and $n_{P-P}$ decrease due to increasing thermal fluctuations. At an infinite temperature the system approaches a random configuration, and the contact densities converge to their random values.

Figure 6.5: Contact densities between polar and hydrophobic sides as a function of temperature for a 3D system of solute particles with one polar side, $N_P = 1$. The contact densities between sides are normalized to the values expected in a random system for $\rho_p = 0.24$. 
6.2 Surfactant Model

Micelles are generally formed by amphiphilic molecules which may be divided into two distinct regions, the polar head and the hydrophobic tail, and usually referred to as surfactants. The length of the hydrophobic tail, which is usually composed of one or more hydrocarbon chains, is generally rather greater than the size of the polar head. Experimental observations [20] suggest that surfactants with tails shorter than 10 carbon atoms are highly soluble in aqueous solutions, while those whose tails exceed approximately 20 carbon atoms are almost completely insoluble.

For simple geometrical reasons, the total repulsion between such a long chain and the water molecules surrounding its sides is much stronger than that between the small tip of the chain and the neighboring water molecules, as illustrated in Fig. 6.6. We adapt the model described in Sec. 6.1 to include this aspect by assigning different energy levels to shell water clusters in contact with the sides of a solute particle compared to those in contact with the tip. Because the number of shell water molecules interacting with one side of the hydrophobic tail with length \( l \) is approximately \( l \) times that of those interacting with the tip, the energy associated with a site representing all of these water molecules is taken to be \( l \) times that for a site representing the water molecules interacting with the tip.

While surfactant molecules are naturally also amphiphilic, we will refer henceforth to molecules modeled with a tail length \( l \) (Fig. 6.6) as “surfactants,” and to the molecules represented in the simple cubic model of Sec. 6.1 as “amphiphiles.”

Figure 6.6: Schematic representation of a surfactant in water (left). White circles represent water molecules in the shell site of a tail side (S), red circles those for a polar side (P), and blue circles those for a tail tip (H). In the anisotropic HP model, a surfactant is represented by a cube whose sides interact with neighboring water sites with a strength which differs according to their positions on the surfactant (right).
A cubic solute particle in the surfactant model is shown in Fig. 6.6. It consists of a polar head (P) and a hydrophobic tail, which in turn is divided into the moderately hydrophobic tip (H) situated opposite the polar head and the long sides of the tail (S) which are represented by strongly hydrophobic sides. In the coarse-grained model, both the tip and each long side of the tail are represented by a face of the cubic particle, but the sides interact much more strongly with a neighboring water site than does the tip simply because the neighboring water site represents a larger group of water molecules.

The Hamiltonian of a system of \( N \) sites on a cubic lattice which are either occupied by water (\( n_i = 1 \)) or by a surfactant molecule (\( n_i = 0 \)), is then

\[
H[\{n_i\}, \{\sigma_i\}] = \sum_{i=1}^{N} n_i \left[ (E_{os,H} \delta_{i,\sigma_{os,H}} + E_{ds,H} \delta_{i,\sigma_{ds,H}}) \frac{1}{Z} \sum_{\langle ji \rangle} (1 - \lambda_{i,j(k)}) \right] \\
+ (E_{os,S} \delta_{i,\sigma_{os,S}} + E_{ds,S} \delta_{i,\sigma_{ds,S}}) \sum_{\langle ji \rangle} \frac{1}{Z} \lambda_{i,j(k)} (\lambda_{i,j(k)} - 1) \\
+ (E_{ob} \delta_{i,\sigma_{ob}} + E_{db} \delta_{i,\sigma_{db}}) \sum_{\langle ji \rangle} \frac{1}{2Z} \lambda_{i,j(k)} (1 + \lambda_{i,j(k)}) ,
\]

(6.6)

where \( \lambda_{i,j(k)} \) is a side variable depending on the nearest neighbor \( j \) of site \( i \), whose orientation is specified by the variable \( k \) of the neighboring site \( j \). The variable \( \lambda_{i,j(k)} \) takes the value 1 if the neighboring face is a polar face or a water face, 0 if it represents a slightly hydrophobic tail tip H, and -1 if it represents a strongly hydrophobic tail side S. The orientation variable \( k \) varies from 1 to \( n_{c,j} \), where \( n_{c,j} \) takes properly into consideration all possible rotations of site \( j \), and \( n_{c,j} = 1 \) for water and \( n_{c,j} = 6 \) for surfactant particles. Because a polar side \( i \) may be in one of \( q \) different states, \( \delta_{i,\sigma_{os,i}} \) is 1 if site \( i \) is occupied by a polar face in one of the \( q_{os} \) ordered shell states of an H face and 0 otherwise, and \( \delta_{i,\sigma_{ds,j}} \) is 1 if it is occupied by pure water in one of the \( q_{ds} \) disordered shell states of an H face and 0 otherwise. Analogous considerations apply for S faces and for the bulk states.

As in Sec. 6.1 only water sites contribute to the free energy, and a polar face is considered having the same effect on a neighboring water site as water. We define

\[
Z_{s,H} = q_{os,H} e^{-\beta E_{os,H}} + q_{ds,H} e^{-\beta E_{ds,H}} \\
Z_{s,S} = q_{os,S} e^{-\beta E_{os,S}} + q_{ds,S} e^{-\beta E_{ds,S}} \\
Z_b = q_{ob} e^{-\beta E_{ob}} + q_{db} e^{-\beta E_{db}},
\]

(6.7)

and

\[
S_H = -\frac{1}{\beta} \ln Z_{s,H} \\
S_S = -\frac{1}{\beta} \ln Z_{s,S}
\]

96
For a system of \( N \) sites, the canonical partition function may be expressed as

\[
Z_N = \sum_{\{n_i\}} \prod_{i=1}^{N} \prod_{k=1}^{n_{c,j}} e^{-\beta [S_H n_i \left( \sum_{j(i)} (1 - \lambda_{i,j(k)}^2) + B n_i \sum_{j(i)} \frac{1}{2z} \lambda_{i,j(k)} (1 + \lambda_{i,j(k)}) \right)]} 
\times e^{-\beta [S_S n_i \sum_{j(i)} \frac{1}{2z} \lambda_{i,j(k)} (\lambda_{i,j(k)} - 1)]},
\]

where \( n_{c,j} \) is the number of possible orientational configurations, and \( n_{c,j} = 1 \) for water sites and \( n_{c,j} = 6 \) for surfactants.

In a system where the surfactant density is not fixed, a chemical potential \( \mu \) associated with the insertion of particles is included, and the grand canonical partition function may be expressed as

\[
\Xi = \sum_{\{n_i\}} e^{\beta \mu N_w} Z_N = \sum_{\{n_i\}} e^{-\beta H_{\text{eff}}^{\text{gc}}[\{n_i\}]},
\]

where the effective, grand canonical Hamiltonian is given by

\[
H_{\text{eff}}^{\text{gc}}[\{n_i\}] = \sum_{i=1}^{N} \sum_{k=1}^{n_{c,j}} \left[ n_i (S_H \sum_{\langle j \rangle} (1 - \lambda_{i,j(k)}^2) - \mu) + B n_i \sum_{\langle j \rangle} \frac{1}{2z} \lambda_{i,j(k)} (1 + \lambda_{i,j(k)}) \right] + S_S n_i \sum_{\langle j \rangle} \frac{1}{2z} \lambda_{i,j(k)} (\lambda_{i,j(k)} - 1).
\]

To measure the formation of aggregates in the surfactant system, the number of contacts between different faces is again determined and compared with the number of contacts in a randomly distributed system of the same particle density \( \rho_p \) in the thermodynamic limit. In a random solution of solute particles, whose positions and orientations are completely independent, these probabilities are given in the thermodynamic limit by

\[
p(H-H) = p(H)^2,
p(P-P) = p(P)^2,
p(S-S) = p(S)^2,
p(P-S) = 2p(S)p(P),
p(H-S) = 2p(S)p(H),
p(H-H) = 2p(H)p(P),
\]

where the symmetry of the system requires that \( p(H-P) = p(P-H), p(S-P) = p(P-S), \) and \( p(H-S) = p(S-H) \). Here \( p(H) \) is the probability that the adjacent side of the nearest neighbor is slightly hydrophobic, \( p(P) \) is the probability that it is polar, and \( p(S) \) is the probability that it is strongly hydrophobic. These probabilities are given in
Micelle Formation  

Surfactant Model

general by

\[
p(H) = \frac{N_H}{z} \rho_p, \\
p(P) = (1 - \rho_p) + \frac{N_P}{z} \rho_p, \\
p(S) = \frac{N_S}{z} \rho_p.
\]  

(6.13)

although henceforth we will consider only the values \(N_H = N_P = 1, N_S = 4\) (Fig. 6.6). Aggregate formation has occurred in this system if the number of \(H-H\) contacts, \(N_{H-H}\), of \(P-P\) contacts, \(N_{P-P}\), and of \(S-S\) contacts, \(N_{S-S}\), are larger than their random expectation values, and the number of \(P-H\) contacts, \(N_{P-H}\), and of \(P-S\) contacts, \(N_{P-S}\), are correspondingly smaller.

These probabilities may also be obtained by calculating the average density of each face in a mean-field approach if each site has \(z\) nearest neighbors. The mean density of the parameter \(n_i\) may be expressed in terms of the particle density \(\rho_p\) according to

\[
\langle n_i \rangle = 1 \cdot (\rho_w + \frac{1}{6} \rho_p) + 0 \cdot \frac{1}{z} \rho_p + (-1) \cdot \frac{z - 2}{z} \rho_p = 1 - \frac{3}{2} \rho_p,
\]  

(6.14)

where the water density is \(\rho_w = 1 - \rho_p\). By analogy,

\[
\langle n_i^2 \rangle = 1 - \frac{1}{z} \rho_p.
\]  

(6.15)

The mean densities of the different faces \(P, H,\) and \(S\) may then be expressed as

\[
\rho_P = \langle \frac{n_i(n_i + 1)}{2} \rangle = 1 - \frac{z - 1}{z} \rho_p, \\
\rho_S = \langle \frac{n_i(n_i - 1)}{2} \rangle = \frac{z - 2}{z} \rho_p, \\
\rho_H = \langle (1 - n_i^2) \rangle = \frac{1}{z} \rho_p.
\]  

(6.16)

Fig. 1.5 shows a typical surfactant molecule with a tail containing 12 carbon atoms. The space-filling representation illustrates the relative size of the polar head compared with the non-polar tail. In this case, the tail is approximately three times longer than the head resulting in a number of neighboring water molecules which is three times larger. This fact is also illustrated schematically in Fig. 6.6, where the number of water molecules interacting with the tail tip \(H\) of a surfactant, and the number of those in contact with the polar head, are multiplied by a relatively small factor to obtain the number of water molecules interacting with the tail side \(S\). We define therefore the length \(l\) of the tail as an effective length, associated with the effective interaction between the side of the tail and the neighboring water molecules relative to the effective interaction between water sites and the tail tip or the head. The approximation describes also surfactants containing more than one hydrophobic tail.
as exemplified in Fig. 6.10(b).

The absolute energy levels and the degeneracies of water sites facing S sides are higher because this site represents the number of water molecules contained in a site in the shell of H multiplied by the factor $l$. The energy of a S shell site is thus obtained from that for a H shell site, for both ordered and disordered states, using

$$E_{os,S} = l E_{os,H}, \quad (6.17)$$

$$E_{ds,S} = l E_{ds,H}, \quad (6.18)$$

where $E_{os,H}$ ($E_{ds,H}$) is the energy of an ordered (disordered) water site in the shell of the tip (H) and $E_{os,S}$ ($E_{ds,S}$) that of an ordered (disordered) shell water site of the tail side (S). Under the assumption that neighboring water sites are rather independent, the total number of configurations for two sites may be approximated by the product of their numbers of configurations. Thus, the number of configurations of a S shell site is related to the number of configurations of a corresponding H shell site by

$$q_{os,S} = q_{os,H}^l, \quad (6.19)$$

$$q_{ds,S} = q_{ds,H}^l. \quad (6.20)$$

The parameter values of the energy levels, and the degeneracies of bulk water and H shell sites, are chosen as described in Sec. 2.1.

### 6.2.1 Methods

Analysis of the properties of a surfactant-water system at the molecular level is possible by similar Monte Carlo simulations using the model of Sec. 6.2. However, the procedure described above must be redefined in one respect. In the Metropolis algorithm, the relative transition probability for configuration $\{n_i^l\}$ from a previous one $\{n_i\}$ depends on the difference in free energy of the two configurations according to

$$r = e^{-\beta[H_{st}^{gc}(\{n_i^l\}) - H_{st}^{gc}(\{n_i\})]}.$$  \hspace{1cm} (6.21)

This free-energy difference must be calculated for two states with the same number of molecules. One step of the simulation procedure consists either in rotation of a solute particle or in a site exchange between two randomly chosen sites. The only contributions to the difference in free energy are then the energy change of the sites concerned, and of their nearest-neighbor sides. If a side changes from polar to S or vice versa, the bulk and shell states of neighboring water sites must contain the same number of water molecules to be comparable. In this case, we attribute the bulk energies

$$E_{os,b} = l E_{ob}, \quad (6.22)$$

$$E_{ds,b} = l E_{db}. \quad (6.23)$$
and their respective degeneracies

\[ q_{ob,S} = q_{ob} \]  \hspace{1cm} (6.24) \\
\[ q_{ob,S} = q_{ob}^l \]  \hspace{1cm} (6.25)

to the relevant water sites when calculating the transition probability \( r \). The index \( S \) refers to the comparison of this bulk site with a \( S \) shell site.

### 6.2.2 Results

We have performed Monte Carlo simulations for surfactant molecules in water on a cubic lattice for varying lengths \( l \) of the hydrophobic tail to investigate the stability of their aggregation as a function of hydrophobicity \((l)\) and density.

Fig. 6.7 shows the \( \rho_p-T \) phase diagram of surfactant particles in water for different values of \( l \). The coexistence regime is significantly enhanced with increasing tail-length \( l \), which can be attributed to the stronger effective repulsive interaction between the longer hydrophobic tail of the surfactant molecule and the surrounding water. Outside this region, the surfactants are soluble at all densities below the LCST and above the UCST, and the system forms a homogeneous solution. For temperatures between these values, the surfactant molecules are soluble only at very low densities, while at the CMC, \( \rho_1 \), the particles aggregate and the system separates into two phases: nearly pure water (of surfactant density \( \rho_1 \)) and an organic phase of density \( \rho_2 \). With increasing tail-length, as the solubility of the solute particles is reduced, the CMC decreases, and the UCST increases while the LCST is reduced.

As described above, the ratio \( l \) between the size of the polar head and the length of the tail is the important quantity to characterize the degree of amphiphilicity of a surfactant molecule. Comparison with Fig. 1.5 indicates that for a typical polar head, \( l = 1 \) corresponds to a tail containing approximately four carbon atoms. Fig. 6.7 shows that for surfactant molecules with a tail containing approximately 8 carbon atoms \((l = 2)\) there is already an enhancement of the coexistence region. The aggregation of solute particles with a tail composed of approximately 12 carbon atoms \((l = 3)\), which represent sodium lauryl sulfate, is reinforced very significantly. Within our simplified model, at temperature \( T = 1 \) surfactants with \( l \geq 4 \) are basically insoluble and form a completely separated phase and thus perfect micelles for all densities, in rather good qualitative agreement with expectations based on experiments in Ref. [20].

We may analyze the aggregation in more detail by studying the density of side contacts as a function of temperature. The density of contacts between the different sides in a system of surfactant particles of tail length \( l = 3 \) in water is shown for different temperatures in Fig. 6.8. The results are normalized to the corresponding
Figure 6.7: $\rho_p$-$T$ phase diagram illustrating micelle formation in the extended 3D, HP model for surfactant molecules of varying tail length. $l$ represents the relative length of the hydrophobic tail compared with the size of the head (Fig. 6.10). The coexistence region is enhanced as the length of the hydrophobic tails increases, and aggregation is promoted.

The probability of these contacts in a randomly distributed system with the same particle density, $\rho_p = 0.25$. At low temperatures the solute particles are clearly soluble, because $n_{H-H} < p(H-H)$, $n_{P-P} < p(P-P)$, $n_{S-S} < p(S-S)$, and $n_{P-H} > p(P-H)$, $n_{P-S} > p(P-S)$, $n_{H-S} > p(H-S)$. This again reflects the formation of cage-like structures around the hydrophobic tails when entropy effects are minimal, favoring dissolution and consequent suppression of $n_{S-S}$ and $n_{H-H}$. At temperatures higher than the lower critical temperature for density $\rho_p$ (Fig. 6.7), the surfactant particles aggregate to minimize their total exposed surface, whence $n_{H-H} > p(H-H)$, $n_{P-P} > p(P-P)$, and $n_{S-S} > p(S-S)$, while $n_{P-H} < p(P-H)$ and $n_{P-S} < p(P-S)$. At high temperatures the contact densities converge to their random values, indicating complete mixing.
Figure 6.8: Contact densities between the different faces as a function of temperature for a 3D solution of surfactant molecules with effective tail length $l = 3$. The contact densities between sides are normalized to the values expected in a random system for $\rho_p = 0.25$.

6.3 Discussion

Amphiphilic molecules in aqueous solutions can form different types of micelles depending on their concentration and on the distribution of polar regions at their surfaces. Our initial approach to capture the qualitative properties of micelle formation in a hydrophobic-polar model involved systematic substitution of the hydrophobic sides of cubic solute particles by polar ones. We determined the $\rho_p-T$ phase diagram for different surface patterns and found closed-loop coexistence curves (Fig. 6.1), in accord with experiments using hydrophilically modified copolymers of poly(N-isopropylacrylamide) [52]. With increasing polarity, the coexistence region is reduced as the solubility of the model amphiphiles increases. This tendency was confirmed by the same experiment, where the LCST of purely hydrophobic poly(N-isopropylacrylamide) P3 was observed to increase from 37°C at atmospheric pressure to 42-44°C when approximately 13% of the monomers were substituted by polar species (CP2 and CP3). In our model we observe the same quantitative increase of 2-3% in absolute temperature from $T_{LCST} = 0.545$ for purely hydrophobic solute particles to $T_{LCST} = 0.56$ for amphiphiles with one polar side, which represents 16% of the particle surface. Thus the most simple cubic model appears to yield good
agreement with available data at this level of comparison.

For particles with two polar sides, the coexistence region is reduced dramatically. This is not surprising, considering the fact that two polar sides represent one third of the total particle surface, and the attractive interactions with water are rather strong. In amphiphilic molecules, the polar regions are in general relatively small compared with the hydrophobic surface area. The solubility of the molecules thus increases considerably on substitution with polar monomers, leading to a decrease in UCST and an increase in LCST and CMC. In the cubic model no significant difference in the size of the coexistence region is observed for different distributions of the two sides on the solute cubes. Any further substitution of hydrophobic cube faces by polar ones results in molecules which are at least half polar: the solubility of such particles is always high, and thus no aggregation is found.

Figure 6.9: Representation of different surface patterns on cubic solute particles in the HP model, and their schematic correspondence to different micelle types. Polar surfaces and micelle segments are marked in red, hydrophobic surfaces in white. a) A cube with one polar side \( (N_P = 1) \) represents a surfactant molecule. b) A cube with two opposite, polar sides \( (N_P = 2) \) corresponds to a section of a bilayer. c) A cube with two adjacent, polar sides \( (N_P = 2) \) represents a section of a cylindrical micelle. d) A section (one eighth) of a spherical micelle is represented by a cube with three adjacent, polar sides \( (N_P = 3) \). e) A cube with \( N_P = 3 \) and two opposite polar sides corresponds to a section of a cylindrical micelle.

Possible interpretations of the various surface patterns of cubes representing solute
particles in the HP model are shown in Fig. 6.9. The model is applicable for any surface pattern and density of solute, with the premise that one site may contain one amphiphilic molecule or a group of solute molecules. A given surface distribution of polar groups on a polymer may be characterized by a corresponding arrangement of polar sides on the surface of a cubic solute molecule. A single surfactant molecule with a clear distinction between a polar head and a hydrophobic tail may be represented by a cube with one polar face [Fig. 6.9(a)]. For the formation of “spherical” micelles, each site must have three adjacent polar sides [Fig. 6.9(d)]. If two of the three sides are situated opposite each other [Fig. 6.9(d)], the site may be considered to symbolize a section of a cylindrical micelle. A smaller section of a cylindrical micelle would be represented by a cube with two adjacent polar sides [Fig. 6.9(c)]. Finally, a cubic particle with two opposite polar sides corresponds at the same level of approximation to a cross-section of a bilayer [Fig. 6.9(b)].

Although by construction \( N_p = 3 \) should give either small spherical or cylindrical micelles, depending on the distribution pattern at the surface of each site, the solubility is too high to find aggregates. This is due to the fact that surfactants are composed of a long tail and a small polar head. Thus even for sections of these micelles the net polar surface is far less than half of the total surface of such a site, and the model overestimates the polar surface area.

Such micellar structures occurring in 3D simulations are difficult to extract and display. To confirm the formation of different micelle types depending on the surface pattern of the solute particles, snapshots of an analogous 2D system are shown in Figs. 6.2 and 6.3. Here the square solute particles may be interpreted in a manner similar to the 3D case (Fig. 6.4), and the formation of small micelles and layers is found in the coexistence region. During the relaxation process, micelles grow from initial dimers to larger entities. Although the solute particles may rotate at a given position, they can be trapped in a configuration which disables the construction of perfect micelles or extended layers. Because there is no preference for growing a layer in one direction rather than in the other, short line segments are formed which are incompatible with others, resulting in a network of short layers. In the model, no distinction is made between the polar side of a water molecule and that of a solute particle, and a \( P-P \) contact contributes the same energy independent of the molecules to which the sides under consideration belong. Such contacts between the polar sides of solute molecules are found in the interior of a micelle, which also influences the formation of perfect micelles. The non-zero temperature in the coexistence region, and the observation that the upper critical density is much smaller than unity, might further imply the formation of imperfect micelles. In fact the extent to which the upper critical density is significant remains unclear, because shell water sites in the model may be considered as belonging to the micelle phase rather than to the pure water phase, which would explain the low density of the organic phase even for perfect micelles containing no water molecules.
A quantitative measure of micelle formation is provided by the contact densities, \( n_{H-H}, n_{P-P}, \) and \( n_{P-H} \), normalized by the corresponding probabilities of the contacts in a random system (Fig. 6.5). Below the LCST the solute particles are dissolved due to the favorable enthalpy gain associated with a cage-like arrangement of water molecules around the hydrophobic parts of the amphiphiles. This preferential hydration is confirmed by a \( P-H \) contact density which is higher in the solution than for a random system, while the \( P-P \) and \( H-H \) contacts are found to be lower than for a random distribution. Above the LCST this picture is inverted because increasing entropy effects favor a screening of hydrophobic faces from polar ones, leading to micelle formation. The effect on \( n_{H-H}/p(H-H) \) is much more pronounced than \( n_{P-P}/p(P-P) \) because the number of polar sides in the system is much higher than the number of hydrophobic ones. Thus a slight change in the number of \( H-H \) contacts leads to a larger change in \( n_{H-H}/p(H-H) \) than in \( n_{P-P}/p(P-P) \), and in fact the majority of \( P-P \) contacts are intact even in the dissolved phase due to the high number of water molecules. For low particle densities, the effect is even larger because fewer \( H-H \) contacts are possible, and a small absolute change causes a considerable relative change. At high temperatures, the entropy becomes dominant and the contact densities approach their respective random values as complete mixing is obtained.

Figure 6.10: Space-filling model of a typical glycolipid containing two hydrocarbon tails and a relatively large polar head compared with the length of the tails. Carbon atoms are represented in black, hydrogen atoms in white, nitrogen atoms in blue, and oxygen atoms in red.

We have extended our analysis to describe surfactants, which represent a particular type of amphiphilic molecule. Surfactants are distinguished by a special partition of the polar and hydrophobic segments along the molecule: a typical surfactant molecule consists of a polar head and one or two hydrophobic tails. To incorporate these geometrical features in the model we have adapted the energy levels and their
degeneracies according to the tail length. The surfactant molecules are represented by cubic solute particles containing one polar face $P$ corresponding to the head, one hydrophobic face $H$ opposite to it, and of the same length ($l = 1$), corresponding to the tail tip, and four strongly hydrophobic faces $S$ ($l > 1$) connecting head and tail tip, which represent the sides of the tail. An $S$ face is in contact with a water site containing a number of water molecules $l$ times greater that that in contact with a $H$ or $P$ face. The tails of typical surfactant molecules exhibiting surface-active properties contain 10 to 20 carbon atoms, and these tails may be on average some three to five times longer than the dimensions of the polar head. In surfactants composed of two or more tails, this ratio is generally smaller. As an example, the head size of the characteristic glycolipid in Fig. 6.10 is close to one third of the length of the tail composed of 14 carbon atoms. We have defined the effective tail length $l$ as the ratio between the tail length and the head-size of the surfactant, so that a typical surfactant is represented by $l$ values between two and five, which may also be fractional.

As expected, Monte Carlo simulations of surfactants of increasing length illustrate a significant enhancement of the coexistence region (Fig. 6.7). Surfactants with a longer tail have more pronounced characteristics of hydrophobic solute particles than do surfactants where the polar head represents a considerable fraction of the molecule. The solubility of long-tailed surfactants is therefore lower than that of short-tailed ones, causing a decrease in LCST and an increase in UCST with lengthening of the tail. In addition, at a given temperature the CMC decreases as the tail becomes longer. For the parameters used in the model we find that an increase from $l = 1$ to $l = 2$ already enhances the coexistence region substantially, and for that $l \geq 4$ the system is completely separated into pure water and micelles of density $\rho_p = 1$ over a wide temperature range, which is fully consistent with experiment [20].

The contact densities $n_{H-H}$, $n_{P-P}$, $n_{S-S}$, $n_{P-H}$, $n_{P-S}$, and $n_{S-H}$, normalized by the corresponding probabilities of the contacts in a random system, provide a quantitative measure of aggregation (Fig. 6.8). We have found fewer contacts between hydrophobic faces and more hydrophobic-polar contacts than would be expected for a random distribution below the LCST, indicating highly dissolved surfactant molecules. Above the LCST this picture is inverted, thus confirming the aggregation of surfactants and the separation of the solution into two phases. Because of the much stronger effect of an $S$ face than of an $H$ face on its neighboring water site, the normalized contact density of two $S$ faces is highest in the aggregation phase and lowest below the LCST. A higher relative contact density between $S$ and $H$ faces is observed for the same reason as between two $H$ faces, while below the LCST water sites prefer to form the solvation shell of $S$ faces rather than of $H$ faces. As temperature increases towards infinity, the solution approaches a random system due to dominant entropy effects, and thus all the relative contact densities approach unity.

A solution of surfactants in water may produce a lamellar phase of bilayers at
low temperature and rather high densities \[63\]. Such formation of bilayers requires the possibility of smooth curvature and high flexibility \[55\]. In the Monte Carlo simulations, the geometrical constraints presented by the lattice prevent the formation of extended surfactant bilayers. For the same reason it is difficult to find well-formed micelles in the more general case of amphiphiles in water. The model considers explicitly the energy states of water sites, and therefore no distinction is made between S-S, S-H, H-H, and P-P contacts. An S-H contact which is formed during the Monte Carlo simulation is as favorable as an S-S contact, although it is more likely to be broken in a later step. The model allows a surfactant particle to occur in the core of micelles, where its polar side is in direct contact with the polar side of another surfactant particle, because no distinction is made between a group of water molecules and the polar face of a particle. In addition, the orientation of the surfactant is irrelevant for the formation of an S-S contact, which prevents efficient alignment of the heads and may further hinder the formation of extended bilayers. The incorporation of these distinctions in a more sophisticated description of the surfactant solution may be expected to reproduce further detailed properties of real systems.

In summary, we have extended the MLG framework to include the solvation of amphiphilic solute particles in water. Within a cubic HP model we have found the aggregation of solute particles, and the formation of various types of micelle as a function of the distribution of the hydrophobic regions. By successive substitution of hydrophobic sides by polar ones, we have studied the aggregation behavior and the influence of the degree of hydrophobicity on the upper and lower critical solution temperatures. We have refined this model to describe surfactant molecules of varying length, by adapting the interaction of the hydrophobic tail to include a corresponding number of neighboring water molecules, and have demonstrated the enhanced stability of aggregates with increasing tail length (increasing hydrophobicity). We have shown that primary features of micelle formation, which are often attributed solely to the amphiphilic nature of the solute particles under consideration, are reproduced by our extension of the solvent-based MLG model to describe alterations of water structure in the vicinity of the different surface regions of dissolved amphiphiles.
Chapter 7

Hydrophobicity in Protein Folding

One of the most persistent challenges in modern molecular biology is to understand the underlying mechanism involved in the folding of proteins into their unique conformation [64]. The difficulty of the protein-folding problem is the fact that a protein is a long polymer chain (polypeptide) formed by 20 different amino-acids. The main carbon atom of each residue, at which its side chain is attached, is called the $C_{\alpha}$ atom and forms the backbone of the protein. Fig. 7.1 shows the atoms of a short region of a protein. Amino-acids can be combined to form an extremely large number of different proteins, which consist typically of hundreds of monomers. In addition, many proteins which differ considerably in their sequence fold into highly similar three-dimensional structural domains. Currently, there are almost 120,000 determined protein sequences in the protein sequence database SWISS-PROT.\(^1\) However, the nearly 20,000 known structures in the Protein Data Bank\(^2\) resulted in only 701 different folds, which are classified in SCOP\(^3\) [95]. Although the number of determined sequences and structures increases rapidly, the number of “new folds” increases only slowly, which suggests that the total number of structures for natural proteins is extremely small compared with the number of different proteins [96]. Thus there is some simplifying force at work which results in many different sequences folding into only a few structures.

The concept of designability was introduced to understand how nature has selected such a small number of protein structures [97, 98, 99, 100]. The designability of a structure is defined as the number of amino-acid sequences which have this fold as their native configuration. It has been shown that almost all possible structures have low designability, and only a rather small subgroup of structures is distinguished by high designability. These highly designable structures possess further protein-like properties, including thermodynamic and mutational stability, rapid folding, and tertiary symmetries. Thus designability may be an appropriate measure for how likely it is that a particular structure is selected in nature. Because very different

amino-acid sequences may have as their native configuration the same fold, it is difficult to distinguish common characterization patterns associated with their monomers.

One way of characterizing the 20 amino-acids may be according to their hydrophobicity, which is thought to represent the essential property determining whether an amino-acid is more likely to be found in the core or on the surface of a folded protein. Because the hydrophobicity is believed to represent the main force driving the folding of proteins [10], certain models, usually referred to as HP models, classify the amino-acids into hydrophobic (H) and hydrophilic (P for polar), sometimes including intermediates (see for example Refs. [102, 103]). Many attempts based on different approaches have therefore been made to determine the hydrophobicity of the 20 amino-acids [87, 104, 105, 106, 107, 101]. However, the various hydrophobicity scales in the literature often disagree as to the hydrophobicity ranking of the amino-acids. These discrepancies may be attributed to the fact that hydrophobicity is a relative quantity, which depends on the environment and on reference molecules.

Hydrophobic side chains of amino-acids are densely packed inside the protein (called the hydrophobic core), while more polar residues are found on the exposed surface of the protein, and form hydrogen bonds with the surrounding water molecules. In order to bring the hydrophobic side chains into the core, the polypeptide chain must fold also in the interior of the protein. This dense packing is obtained by forming secondary structures (the primary structure is the sequence), including the two major motifs, namely \( \alpha \)-helices and \( \beta \)-strands, which are connected through turns (hairpin loops), as shown in Fig. 7.2. In general, proteins are formed by a number of secondary-structure elements (cf. Fig. 7.3). These motifs are essentially stabilized by hydrogen bonds between N-H and C=O groups of the main chain, the backbone. The core of a protein is thus rigid and stable, and proteins with similar structures have essentially the same interconnections between almost identical
secondary-structure elements, but vary typically in the *turns*. These loop regions between secondary-structure motifs often appear on the surface and exposed to the solvent, and are consequently rich in charged and polar amino-acids. When a protein sequence is known it is therefore easier to identify hairpin loops than the more regular motifs. Because the turns are most exposed, functional groups are attached frequently to these loop regions. The most common, natural folds are distinguished by tertiary symmetry (*i.e.* the arrangement of the secondary-structure elements is symmetric), and by thermodynamic and mutational stability.

**Figure 7.2:** $\alpha$-helix domain (left), small protein formed by $\beta$-strands (middle), and a small protein with both $\alpha$-helices and $\beta$-strands (right).

**Figure 7.3:** Two highly symmetric proteins which are formed by $\alpha$-helices and $\beta$-strands: Triosephosphate Isomerase (1TPF), and Lumazine Synthase (1DI0).

Although on average hydrophobicity has been shown to correlate with surface exposure, it is not clear to what extent that a protein’s fold, and hence its surface
exposure pattern, correlates with the hydrophobic pattern dictated by its amino acid sequence. If the average hydrophobic behavior of amino acids is true in general then there should be a high correlation between the hydrophobicity sequence and the corresponding surface exposure pattern. If the correlation is not high, this has implications for folding models that only incorporate solvation effects. It also has potential implications for protein design that is based purely on hydrophobic-polar patterning, since the hydrophobicity pattern is assumed to dictate the final fold.

In this section we analyze the correlation between the accessible surface areas and hydrophobicities of amino-acids on a structure to structure basis. Based on this analysis we are able to assess which of the currently available hydrophobicity scales yield the best correlations with surface exposure patterns of known protein structures. Surprisingly, we find the correlation between a protein’s hydrophobicity sequence and its surface exposure pattern to be poor. Based on this analysis we gauge the accuracy of certain hydrophobicity scales for amino-acids in the real protein environment. We then define a new scale of “hydrophobicity” based on the actual, measured surface exposure of each amino-acid in the entire database. We extend the database analysis to the main secondary-structure elements, α-helices and β-strands, to confirm the validity of the surface-exposure scale. We conclude by comparing the results with a hydrophobic-polar model for protein structures which is free of any prior assumptions concerning the mechanism of the folding process, and obtain good agreement with the new scale. We show that it may be more appropriate to use a statistically derived scale based on actual solvent exposures for classification over those derived from transfer measurements.

### 7.1 Protein-Structure Classification

The first step towards an understanding of the mechanisms underlying protein folding was taken by Anfinsen in 1973, who established the “thermodynamic principle” for proteins [64]. The hypothesis is that the native conformation of a protein in a given environment is determined completely by the minimum of the Gibbs free energy, and thus by all intermolecular interactions in the system. Consequently, the structure is specified entirely by the amino-acid sequence of the protein in a certain environment. This conclusion was drawn from the observation that the proteins ribonuclease A and staphylococcal nuclease can be denatured reversibly. After removal of a denaturant, such as urea, from the solution they fold spontaneously into their native structure, and cooling a solution of thermically denatured proteins leads to their folding.

The energy landscape of a random amino-acid sequence is extremely rough, and the folding process can be easily trapped in one of the numerous metastable states. Real proteins, however, fold typically within milliseconds and almost always find the
global minimum of the free energy. This indicates that natural proteins differ very significantly from random sequences in that they have adapted during evolution, and each one has a clear global minimum, well separated from all local minima. This short folding time is important, because the unfolded state is more likely to be involved in undesirable chemical reactions which may modify its amino-acid sequence to such an extent that it becomes unable to fold correctly, and thus to perform its original function in the cell. Because the function of a protein is critically dependent on its structure, the latter has evolved to be robust against small mutations and easy to stabilize thermodynamically.

Because rather few protein folds actually occur in nature, and these are composed essentially of only two major secondary-structure elements [108], they may be classified according to their similarities. Structural classification databases reflect the structural and the evolutionary relationships between proteins, and group them into hierarchy trees according to their level of similarity. A comprehensive comparison of the structures of proteins is provided by the FSSP database\(^4\). In the compilation of this database, all residues of the known protein structures are compared in three dimensions, and the results are reported in the form of alignments of equivalent residues. Redundancy is eliminated by removing proteins with mutual sequence identity larger than 25\%, because these result in almost complete structural overlap, and in most cases in similar functions. After suppression of sequence redundancy, some hundreds of representative proteins remain to be considered. Each of these representatives is split into domains, which are defined as parts of proteins observed either as whole proteins or as motifs in various proteins, and are thought to be evolutionary units. All known structures are then compared with each of these domains, and the structural similarity \(\tilde{S}\) is defined as the sum of similarities of equivalent intramolecular distances between all residues \(i\) and \(j\) of the common core of two proteins \(A\) and \(B\),

\[
\tilde{S} = \sum_i \sum_j \left( d - \frac{|d_{ij}^A - d_{ij}^B|}{d_{ij}^*} \right) e^{-\left(\frac{d_{ij}^*/d_{ij}^*}{20.4}\right)^2}, \tag{7.1}
\]

where \(d_{ij}^*\) denotes the mean value of the \(C_\alpha-C_\alpha\) distances \(d_{ij}^A\) and \(d_{ij}^B\). The threshold of similarity is taken as 20\% \((d = 0.2)\), and the exponential factor reduces the weight of residues with large separations. Because the protein is not completely rigid, atoms must be allowed to move, typically over distances on the order of 10 Å. Optimal structural alignments of the equivalent residues in the core maximize \(\tilde{S}\).

Each representative structure has a set of aligned structures. Each structure in turn has a corresponding amino acid sequence. Thus at each position in the structural alignment, there is a corresponding set of amino acids. A given structure’s amino acid sequence also has a list of sequence homologues (whose folds are assumed to be

---

\(^4\)Fold classification based on Structure-Structure alignments of Proteins, http://www2.ebi.ac.uk/dali/fssp/
identical). We obtain these homologues from the HSSP\(^5\) and add them to the list of
sequences which fold into the given structure. Thus for each representative structure
in the FSSP we have a list of aligned structures as well as a potentially much larger set
of amino acid sequences, all of which are assumed to fold into this structure.

### 7.2 Testing Hydrophobicity Scales

To test the suitability of different hydrophobicity scales which are based on the
spatial positions of amino-acids in natural proteins, we have considered the complete
available data concerning protein structures. For this purpose, we have aligned all
known structures in the Protein Data Bank (PDB) with the representative structures
in the FSSP. Each residue of a PDB structure which is aligned with a residue of a rep-
resentative structure occurs at the same geometrical position within the fold as the one
with which it is associated.

A hydrophobicity scale \(s\) assigns a hydrophobicity value \(h_{a,a}^s\) to each amino-acid
(a.a.). \(h_{i,j}^s\) is the hydrophobicity of the \(i\)th aligned residue of sequence \(j\) which is
aligned with a representative structure, based on the hydrophobicity scale \(s\). For the set
of amino acid sequences that fold into a given structure we wish to consider what the
average hydrophobicity sequence for the set. We consider the average sequence since
it gives a good characterization of the hydrophobicity sequence that adopts the given
representative structure. The average hydrophobicity value \(\overline{h_i^s}\) of a position \(i\) within
this representative structure using scale \(s\) is

\[
\overline{h_i^s} = \frac{1}{M} \sum_{j=1}^{M} h_{i,j}^s,
\]

where \(M\) is the number of sequences in which a residue is aligned with residue \(i\).
Calculating this average for all residues of the representative structure with length \(N\)
gives the average hydrophobicity sequence of this structure, \((\overline{h_i^s})_{i=1..N} = \overline{h_1^s} h_2^s \cdots \overline{h_N^s}\).

The hydrophobicity of a residue is critical in determining its surface exposure. If
the hydrophobicity scale \(s\) is appropriate for amino-acids in proteins, a high correlation
should emerge between the average hydrophobicity \(\overline{h_i^s}\) of each residue \(i\) and its degree
of surface exposure.

The surface exposure \(a_i\) of residue \(i\) in structure \(\gamma\) is quantified by the number of
water molecules which can access the surface of the residue simultaneously in this
particular structure. (For each structure we obtain the surface exposure pattern from
the FSSP file.) This quantity is normalized by the maximal number of water molecules
able to access the total surface of the residue, which depends on its size and on the

---

geometrical distribution of the molecules within the residue [105]. Many known structures are found in the alignments of the representative structures. To obtain maximal accuracy in the surface-accessible area of the residues, we take the average over all determined positions of aligned residues,

$$\bar{a}_i^Y = \frac{1}{L} \sum_{j=1}^{L} a_{i,j}^Y,$$  \hspace{1cm} (7.3)

where $L$ is the number of known structures which have a residue aligned with residue $i$. $a_{i,j}$ denotes the surface-accessible area of residue $i$ in sequence $j$ of the alignment. Performing this procedure for each residue $i$ of the representative structure leads to a sequence of surface accessibilities $(\bar{a}_i^Y)_{i=1..N} = \bar{a}_1 \bar{a}_2 \cdots \bar{a}_N$.

The correlation coefficient $c^a$ between the sequences $(\bar{a}_i^Y)_{i=1..N}$ and $(\bar{h}_i^Y)_{i=1..N}$ is given by

$$c^a = \frac{\sum_i^N (\bar{a}_i - \bar{a})(\bar{h}_i^Y - \bar{h}^*)}{\sqrt{\sum_i^N (\bar{a}_i - \bar{a})^2 \sum_i^N (\bar{h}_i^Y - \bar{h}^*)^2}}.$$  \hspace{1cm} (7.4)

The alignments described above have been analyzed in order to determine the average surface accessibilities of each monomer. To each position $i$ of the representative

![Figure 7.4: Correlation coefficients between surface accessibilities for the alignments of each representative structure. The average correlation coefficient is $c = 0.831$.](image-url)
structure these corresponds a normalized surface accessibility \(a_i\). The average surface accessibility (ASA) of amino-acid a.a. in the alignment of a representative structure \(\gamma\) is

\[
\langle a_{\text{a.a.}} \rangle_\gamma = \frac{1}{A_{\text{a.a.}}} \sum_{i=1}^{A_{\text{a.a.}}} a_i,
\]

(7.5)

where \(A_{\text{a.a.}}\) denotes the number of occurrences of amino-acid a.a. in the complete alignment of known structures and their respective sequence homologs. Averaging over the alignments of the \(N_\gamma\) representative structures leads to the total ASA of amino-acid a.a.,

\[
\overline{a_{\text{a.a.}}} = \frac{1}{N_\gamma} \sum_{\gamma=1}^{N_\gamma} \langle a_{\text{a.a.}} \rangle_\gamma.
\]

(7.6)

Surface exposure and secondary-structure characterization of each amino-acid in a folded protein are extracted from the FSSP database. Before looking at the hydrophobicity/surface-area correlation we have assessed the quality of the structural alignments for determining the average surface exposure pattern. For this purpose, we calculated the correlation coefficient between the surface exposure patterns in each alignment (2158 structural alignments). Each point in Fig. 7.4 represents the correlation coefficient between the surface accessibilities [Eq. (7.3)] within each alignment. The natural, minor deviations of aligned structures from their representative folds lead to a relatively large spread in the correlation coefficients. Nevertheless, the average correlation coefficient \(c = 0.831\) represents the best value which may be obtained from the current database. Thus most aligned structures share similar surface exposure patterns.

We have tested several well-known hydrophobicity scales for their accuracy in the protein environment. These scales are based on different approaches, such as measurements of water-vapor transfer free energies and analysis of side-chain distributions [87], semi-theoretical approaches determining transfer free energies for \(\alpha\)-helical amino-acid side chains from water to a non-aqueous environment [104], determination of transfer free energies by measuring solubilities in water and ethanol relative to the reference amino-acid Glycine [105], calculating residue-residue potentials with pairwise contact energies [106], and a refined study of the latter using the Bethe approximation for determination of relative contact energies with respect to the native state [107].

Calculation of the correlation \(c^s\) between the sequences \((\overline{a_i})_{i=1..N}\) and \((\overline{b_i})_{i=1..N}\) for the five chosen hydrophobicity scales \(s\) resulted in extremely low correlations between the average hydrophobicities and the corresponding surface accessibilities, as shown in Fig. 7.5. In comparison, the correlation coefficients between the surface-accessibility sequences in each alignment (Fig. 7.4) exceed significantly the values
obtained from Eq. 7.4, which suggests that a more appropriate hydrophobicity scale is required for describing the tendency of amino-acids to occur at protein surfaces.

None of the average correlation coefficients calculated for any of the tested hydrophobicity scales exceeded 0.5, the highest being $c^* = 0.454$ for the scale in Ref. [107], although the correlation between the surface-accessibility sequences within the alignments was significantly higher. Thus a structure’s hydrophobicity sequence (using the current hydrophobicity scales) seems to correlate poorly with its pattern of surface exposure. This discrepancy may arise from the scales themselves in that they may not accurately capture the hydrophobic behavior of residues within a protein (the hydrophobicity scales used were determined either by comparing single residues in purely aqueous solutions [87], or in solvents such as ethanol [104, 105], instead of in the environment of real proteins, or by minimizing the interactions between residues within proteins on the assumption that the essential intramolecular interactions within proteins are two-body only, without taking into account other interactions [106, 107, 109]), or it may serve to highlight that other forces also contribute significantly to the final fold of a protein.
7.3 Surface-Accessibility Scale

As shown in Sec. 7.2, known hydrophobicity scales provide a rather poor quantification of the tendency of amino-acids to occur at the surface or in the core of a natural protein. Based on values determined by structural analysis of proteins, a surface-accessibility scale may be defined for better classification of amino-acids in proteins, and for this purpose we have investigated the average surface accessibility (ASA) of each amino-acid in protein structures. This scale provides a quantity for describing the actual position of amino-acids in proteins, and represents the “real, statistical hydrophobicity” of amino-acids in their natural protein environment. It may thus be used for experimental comparison of results obtained from theoretical models using hydrophobicity. For this purpose, we have investigated the distribution of ASA for each amino-acid in all known protein structures.

<table>
<thead>
<tr>
<th>Amino-Acid</th>
<th>ASA</th>
<th>Class</th>
<th>Amino-Acid</th>
<th>ASA</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cysteine</td>
<td>0.268</td>
<td>C</td>
<td>Proline</td>
<td>0.502</td>
<td>M</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0.273</td>
<td>C</td>
<td>Arginine</td>
<td>0.539</td>
<td>M</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>0.279</td>
<td>C</td>
<td>Asparagine</td>
<td>0.568</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0.290</td>
<td>C</td>
<td>Serine</td>
<td>0.568</td>
<td>S</td>
</tr>
<tr>
<td>Valine</td>
<td>0.306</td>
<td>C</td>
<td>Glutamine</td>
<td>0.573</td>
<td>S</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0.319</td>
<td>C</td>
<td>Glutamic acid</td>
<td>0.586</td>
<td>S</td>
</tr>
<tr>
<td>Leucine</td>
<td>0.321</td>
<td>C</td>
<td>Glycine</td>
<td>0.588</td>
<td>S</td>
</tr>
<tr>
<td>Methionine</td>
<td>0.364</td>
<td>C</td>
<td>Lysine</td>
<td>0.607</td>
<td>S</td>
</tr>
<tr>
<td>Alanine</td>
<td>0.405</td>
<td>C</td>
<td>Aspartic acid</td>
<td>0.615</td>
<td>S</td>
</tr>
<tr>
<td>Histidine</td>
<td>0.425</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threonine</td>
<td>0.480</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1: Average surface accessibilities of amino-acids obtained by analysis of the complete structure and sequence database [Eq. (7.6)], and our classification based on the surface-accessibility distribution of the amino-acids (see Fig. 7.7).

Tab. 7.1 shows the ASAs of the 20 amino-acids. Although the scale ranges from 0 (completely hidden in the core) to 1 (100% exposed to water), the averages do not take extreme values. 11 amino-acids have rather moderate tendencies to prefer the core of proteins, while 9 are more polar. Tyrosine occurs mostly in the core, and thus shows quite hydrophobic properties in a protein environment. Charged amino-acids including Aspartic acid, Glutamic acid, Lysine, and Arginine tend to occur on the surface, and may thus be associated with hydrophilic properties. Cysteine is the monomer most frequently found in the core, and thus by the ASA criterion represents the most markedly hydrophobic amino-acid. This is in disagreement with Ref. [101], where Cysteine is classified as a polar amino-acid. The ASA scale agrees mostly
with the method of Ref. [107] as regards the distinction between hydrophobic and polar amino-acids. However, the ranking is rather different, and Glycine, which occurs mostly on the surface, is classified in Ref. [107] as neutral or slightly hydrophobic.

The correlations between the ASA values for the 20 amino-acids and their hydrophobicity values determined using the scales under consideration are shown in Fig. 7.6. The three scales based on the transfer free energies of amino-acid side chains from water into either vapor or non-aqueous solvents have the lowest correlation with the ASA scale. An improvement is observed for the scales obtained by determination of the pairwise interaction between amino-acids. Although the correlation between ASA and hydrophobicity determined using these scales is relatively high for individual amino-acids, the situation for real amino-acid surface exposures in natural proteins is unsatisfactory (Fig. 7.5), indicating that the forces involved in protein folding are not limited to pairwise interactions between residues.

Figure 7.6: Correlation between ASA values of the 20 amino-acids (Tab. 7.1) and their hydrophobicity values deduced from the scales of a) Ref. [87], b) Ref. [104], c) Ref. [105], d) Ref. [106], and e) Ref. [107].

The ASA value of each amino-acid is a representative measure for the tendency of an amino-acid to occur preferentially in the core or on the surface of natural proteins. However, deviations from these average values may be considerable in some proteins. To analyze the magnitude of these deviations, we have determined the occurrence of
each amino-acid at different degrees of surface exposure.

Folded proteins are dense, three-dimensional clusters of amino-acids. The core thus represents a considerable portion of the whole protein, whereas only a relatively small number of amino-acids is to some extent exposed to water. To prevent underrating of exposed areas, we define a mutual normalization of the occurrences of the amino-acids at a given surface exposure. We compute the surface distributions of the 20 amino-acids according to

$$P = \frac{p(a.a. \& A)}{p(a.a.)p(A)}$$

(7.7)

where $p(A)$ is the probability that any amino-acid occurs with a given surface exposure $A$ (see Fig. 7.13), $p(a.a.)$ is the probability of occurrence of amino-acid $a.a.$ anywhere in the database (see Fig. 7.15), and $p(a.a. \& A)$ is the probability that amino-acid $a.a.$ occurs with surface exposure $A$.

Figure 7.7: Histograms of degree of surface exposure of the core (top left), surface (top right), and medium (bottom) amino-acids.

Fig. 7.7 shows histograms of the distributions of surface exposures for the 20 amino-acids. The distributions are rather broad. Tests using only a fraction of the
database, and others using only a fraction of the length of the sequences, led to very similar results. Based on these distributions we have distinguished three classes of amino-acids: core amino-acids (\(C\)) with a peak at low surface exposure, surface amino-acids (\(S\)) with a peak at high surface exposure, and intermediate amino-acids (\(M\)) with flat distributions.

These classifications agree well with the classification based on the ASA, as shown in Tab. 7.1. As an example, histidine has a slightly hydrophobic ASA value, which is confirmed by the distribution. We determine ASA sequences by associating the values in Tab. 7.1 to each amino-acid in all protein sequences in the FSSP database. We have calculated the correlation between these ASA sequences and the average surface-exposure sequences (Eq. 7.3) finding the correlation coefficient \(c^{ASA} = 0.663\). Given the correlation coefficients of the accessibility sequences within each alignment (Fig. 7.4), this correlation is excellent. The correlation coefficients of all representative structure alignments are shown in Fig. 7.8, and show a very significant improvement compared to the correlations for known hydrophobicity scales (Fig. 7.5).

Figure 7.8: Correlation coefficient between ASA sequences and average hydrophobicity sequences obtained by using the surface-accessibility scale shown in Tab. 7.1 for all representative structures. The average coefficient is \(c^* = 0.663\).
7.4 Hydrophobic-Polar Model

The protein databases presently cover only a small part of the very large number of existing proteins. Even considering the full PDB and FSSP databases may therefore lead to biased and incomplete results. We investigate the accuracy of the results of the statistical analysis by reproducing the correlations within a specific model for protein structures. We use a 3D, off-lattice, hydrophobic-polar (HP) model, in which the residue hydrophobicities range from 0 (hydrophobic) to 1 (polar) [110]. The structures of polypeptide chains of length $N = 18$ are generated by considering the backbones as self-avoiding walks in free space (“off-lattice”), using a fixed set of three dihedral angle pairs $(\phi, \psi)$ between consecutive residues along the backbone [111]. The angle pairs considered in this study are the two pairs $(-140, 150)$ and $(-65, 125)$ from the $\beta$-strand region of the Ramachandran plot and the pair $(-60, -50)$ from the $\alpha$-helical region [112]. The side chains are represented by hard spheres of radius $r = 1.9\,\text{Å}$ centered on the $C_\beta$ positions of the amino-acids, which are the side-chain carbon atoms bound to the $C_\alpha$ atom in the backbone. The accessible surface $a_i^\gamma$ of residue $i$ in a given structure $\gamma$ is then determined by calculating the exposure of its sphere to a water molecule, which is represented by a sphere of radius $1.4\,\text{Å}$. For each sequence, a sequence $(h_i)_{i=1,N}$ of hydrophobicities is generated as a string of $N$ random, real numbers between 0 and 1. Ref. [110] provides a more detailed description of the model.
Because hydrophobicity is the main force driving the folding process, in this model the energy of a sequence which forms structure $\gamma$ may be expressed as

$$E^\gamma = -\sum_{i=1}^{N} h_{ii} a^\gamma_i.$$  

(7.8)

The native structure of a sequence is that minimizing this energy. The correlations were calculated for the 50 most designable structures, which are thought to be the most protein-like, using a sample of 2000 sequences which fold into these.

![Figure 7.10: Surface-accessibility distributions of different hydrophobicity classes, using the 50 top designable structures of a 3D, off-lattice, HP model.](image)

We have calculated hydrophobicity and surface-exposure sequences in the 3D, off-lattice, HP model. Fig. 7.9 shows the relation between the surface accessibility of each residue and its hydrophobicity. For a given hydrophobicity value, a large variation in surface area is found, explaining the low correlation $c = 0.59$ between hydrophobicity and accessible surface area. However, classifying the residues into ten groups, computing the ASA for each of these hydrophobicity classes, and determining the average surface accessibility of each class leads to a dramatic improvement, with a correlation coefficient $c = 0.99$. Classifying the residues by surface accessibility and then calculating the correlation with the average hydrophobicity of each class also leads to a coefficient of 0.99.

The normalized surface-exposure distribution of each hydrophobicity class is shown in Fig. 7.10. As in the study of natural proteins (Fig. 7.7), the distributions
Figure 7.11: Correlation coefficients between surface-exposure sequence of a structure and average hydrophobicity sequence of all corresponding amino-acid sequences as a function of designability, obtained within the HP model. The correlation improves markedly with increasing designability.

are rather broad, although more pronounced peaks are evident at the lowest (highest) surface-exposure values for the most hydrophobic (polar) residues.

By analogy with the analysis of natural proteins performed in Sec. 7.3, we have

Figure 7.12: Correlation between surface-exposure sequence of a structure and average hydrophobicity sequence for the 50 most designable structures.
calculated the average hydrophobicity sequence of all amino-acid sequences which fold into a given structure, and compared this with the surface-exposure sequence of the structure. The correlation coefficients for all structures with designability ranging from 1 to 1000 is shown in Fig. 7.11. The correlation is rather poor for structures with low designability, but improves significantly with increasing designability. It is clear that the average hydrophobicity sequence becomes more accurate as the number of sequences to be averaged increases. Thus the average correlation coefficient for the 50 most designable structures is $c = 0.936$ (Fig. 7.12).

Comparison of Fig. 7.11 with the correlation coefficient obtained in Fig. 7.8 may indicate that the number of known sequences for a given structure in the protein databases is very low, or that these sequences do not represent a uniform sample of all sequences which could fold into that particular structure. Even for structures with high designability, the sequences in the databases would appear from the low correlation coefficient obtained by averaging not to represent a uniform sample of the possible sequences which might fold into a given structure.

### 7.5 Secondary-Structure Analysis

The native configuration of a folded protein is characterized by secondary-structure elements, namely $\alpha$-helices and $\beta$-strands, which are connected by turns (Fig. 7.2). The formation of secondary structure helps to reduce the many degrees of freedom involved in the folding process thus reducing the conformational search space. Segments of secondary-structure elements are rather short, because they are limited by the diameter of the folded protein. It is often difficult to define the exact ends of the segment, a process which is additionally complicated by the fact that the end regions of a secondary-structure element may have irregular conformations. Analyzing the occurrence of the 20 amino-acids in these structural elements may help to understand the basis of protein structure.

As shown in Fig. 7.13, most of the residues in $\alpha$-helices and $\beta$-strands occur in the interior of native protein configurations. However, this effect is much stronger for $\beta$-strands, indicating that $\alpha$-helices appear more frequently than $\beta$-strands on the surface of proteins. However, if an $\alpha$-helix is situated on the surface of a folded protein, only half of the residues are actually exposed to the solvent, because one side of the helix faces the solution and the other side the hydrophobic interior of the protein, as illustrated in Fig. 7.14. Thus, although $\alpha$-helices are more often found at the surface of the protein, they are not composed only of polar amino-acids, but there is a tendency for the monomers to change from hydrophobic to polar with a periodicity of three to four residues, because there are 3.6 residues per full helix twist. $\alpha$-helices which cross membranes are in a hydrophobic environment, and thus are
mostly composed of hydrophobic amino-acids. These facts may partly explain the high fraction of α-helix residues which appear in the core. In contrast to α-helices, β-sheets (composed of parallel or anti-parallel β-strands) are formed of different regions within the protein sequence. β-sheets tend to be much more deeply buried in the core of proteins, which may explain the large number of β-strand residues found in the core.

To analyze in more detail the structure of α-helices and β-sheets, we have studied the frequency of occurrence of each amino-acid in these motifs. The probability of occurrence of an amino-acid in α-helices is compared to the probability of occurrence of this amino-acid in the entire protein database.

Fig. 7.15 shows the frequency of occurrence of each amino-acid in α-helices and β-strands compared to the frequency of occurrence in the entire database. The amino-acids are arranged according to their ASA values in increasing order. Compared to the total database, β-strands tend to be composed of a high portion of amino-acids with low ASA (and rather large side chains such as V, I, and T, or an aromatic ring as in F, Y, and W) and relatively few amino-acids with high ASA. This is fully consistent with the experimental observation that β-strands are found mainly in the core of

![Graph showing frequency of amino-acids in α-helices and β-strands.](image)

Figure 7.13: Probability of finding a residue at a given degree of surface exposure $\mathcal{A}$ ($\mathcal{A} = 0$: core, $\mathcal{A} = 1$: surface) compared to the probability of finding an α-helix residue and a β-strand residue at a given degree of surface exposure $\mathcal{A}$. The total number of residues in proteins is 352,707, in α-helices 129,643, and in β-strands 74,543.
Figure 7.14: Projection of the position of each amino-acid residue of an α-helix onto a plane perpendicular to the helical axes. The amino-acids are hydrophobic (green), polar (red), or charged (blue), and the part of the protein sequence which folds the α-helix is shown at the bottom. This α-helix has clearly a hydrophobic and a polar side.

proteins. The residues which are found more frequently in α-helices than in other parts of the proteins are divided into comparable numbers of amino-acids with low and high ASA. This supports the idea that α-helices are often composed on one side of hydrophobic and on the other side of polar residues.

Fig. 7.16 shows the surface-exposure distributions of the 20 amino-acids in α-helices and in β-strands, in comparison with the distributions for the entire database. For the core (C) amino-acids, the differences are rather small. However, the intermediate (M) amino-acids tend to occur less frequently at the surface in α-helices, with the exception of Arginine (R). The tendency of surface (S) amino-acids to be exposed to the solvent is enhanced in α-helices, with the exception of Glycine (G), which avoids surface exposure in helices more than in other parts of proteins. Glycine and Proline, which tend to avoid both α-helices and β-strands (Fig. 7.15), are known to occur most often in turns.

For β-strands, the residues which are found most frequently are those with low ASA (Fig. 7.15). In addition, both Figs. 7.15 and 7.16 show that the deviations from the observations of the total database are more significant. In general, core (C) amino-acids tend to show a less pronounced peak at low surface exposure when they occur in β-strands which may be explained by the experimental observation that many of these C amino-acids never appear completely exposed when composing a β-strand. The residues which are most frequently found in β-strands, most of these
residues have low ASA. Finally, our classification is also consistent with the fact that the charged amino-acids (D, E, K, and R) are found more often on the surface when building a \( \beta \)-strand, but their occurrence in \( \beta \)-strands is quite rare. The same holds for most other \( S \) and \( M \) amino-acids, while Glycine avoids surface exposure in \( \beta \)-strands, and in fact avoids \( \beta \)-strands in general (Fig. 7.15).

### 7.6 Discussion

Based on a comprehensive study of the entire protein-structure database, we have determined the average surface accessibility (ASA) of each amino-acid, which is summarized in Tab. 7.1. This quantity measures the average degree of surface exposure of amino-acids in the protein environment. The corresponding ASA value may be assigned to each residue, which leads to a considerable improvement in describing the tendency of amino-acids to occur at the surface of native proteins, as shown in Fig. 7.8. It may therefore be interpreted as a statistical hydrophobicity scale for amino-acids in their protein environment, and thus represents a quantity better suited for the comparison of results determined by theoretical studies with experimental values.
Figure 7.16: Histograms of degree of surface exposure of the core amino-acids $C$ (top left), of the intermediate amino-acids $M$ (top right), and of the surface amino-acids $S$ (bottom) in the complete database, only in $\alpha$-helices, and only in $\beta$-strands.

Many theoretical models used to investigate the mechanism of protein folding, including the HP model [102, 103], simplify the complex problem by assuming the possibility of classifying amino-acids into two or more groups according to their hydrophobicity. To compare the resulting observations with experimental data, the latter must also be analyzed by classifying the amino-acids into a corresponding number of hydrophobicity groups. We have shown that a protein’s fold which determines its surface exposure pattern correlates poorly with its corresponding hydrophobicity sequence (as determined from known scales). A considerable improvement was gained using a scale based on ASA values. Thus for theoretical models based solely on solvation, using the database derived ASA values (which represent the amino acids average propensity to be solvated in the context of a folded protein) will give a closer correspondence between models and known structures. Nevertheless, our results showing the poor correlation between sequence and structure using just hydrophobicity show the importance of other forces in the folding problem.
For most amino-acids, the ASA scale indicates rather moderate tendencies to occur on the surface or in the core. As an example, Cysteine has the smallest average surface exposure, but still has an ASA of 0.268, and the ASA values of 10 amino-acids lie between 0.4 and 0.6. To determine how representative these ASA values are, we have studied the surface-exposure distributions of each amino-acid in the protein-structure database. The distributions are predominantly broad, showing more or less pronounced peaks for certain amino-acids, whereas for others they are practically flat (Fig. 7.7). Nevertheless, it is possible to classify the amino-acids according to their ASA by considering the surface-exposure distributions. We have distinguished three types of distributions: those with a peak at the core (C), with a peak at the surface (S), and flat distributions (M). As shown in Tab. 7.1, such a classification according to the distributions is in agreement with the ASA scale.

The statistical results are confirmed by a study of the 50 most designable structures of length 18 obtained within a 3D, off-lattice, HP model. Calculating the correlation between the hydrophobicities and the surface accessibilities of all residues results in a correlation coefficient \( c = 0.59 \). However, the correlation coefficient between the average hydrophobicity and the average surface accessibility of ten groups of residues, classified according to their surface accessibility, becomes \( c = 0.99 \). The same correlation is obtained by classifying the residues according to their hydrophobicity. Averaging over the entire database thus leads to a substantial improvement, and a classification of amino-acids then becomes very reasonable.

Within this framework we have analyzed the correlation between the surface exposure sequence of each HP model structure and its average hydrophobicity sequence as a function of designability. The correlation improves with increasing designability (Fig. 7.11). For the 50 most designable structures, the average correlation coefficient is \( c = 0.936 \), which is considerably higher than that obtained for real proteins even when using the ASA scale, where \( c^{ASA} = 0.663 \). This indicates that the present databases are far from being complete, and that the low correlation is a consequence of a relatively strong bias in the known sequence database, so that the sampling of all possible sequences is consequently not uniform.

The high correlation coefficient obtained from the HP model may be partly attributed to the fact that computational limitations restrict the calculations to short proteins of 18 monomers. However, Fig. 7.8 shows that for real proteins the correlation improves with increasing length of the polypeptide chains under consideration, which suggests that even better results would emerge for longer proteins also in the model.

To obtain further insight, we have studied the main secondary-structure elements, \( \alpha \)-helices and \( \beta \)-strands. Here we also find significant differences in the positional distributions for most amino-acids. Amino-acids with low ASA occur predominantly in \( \beta \)-strands, which are usually found to occur deep in the core. In contrast, \( \alpha \)-helices
are often situated at the surface of proteins. However, a large number of helix-forming residues have low ASA, which can be attributed to the fact that only one side of the helix is exposed to the solvent, while the other side is directed towards the interior of the protein. This is reflected in the observation that half of the residues which are found preferentially in α-helices have low ASA, while the other half have high ASA. Amino-acids with short polar side chains (Serine (S), Asparagine (N), and Aspartic Acid (D)), Proline (P) (cyclic) and Glycine (C) (no side chain) are rarely found in α-helices or β-strands, but rather in turns.

It is clear that hydrophobicity remains a controversial quantity also in protein science. Well-known hydrophobicity scales for amino-acids frequently disagree with each other, and often fail to predict the degree of surface exposure of a residue in the folded protein structure. Statistical studies have shown that some amino-acids are found preferentially either in the core or on the surface of native folds, whereas many are present at all positions with equal probability. Based on these results we have presented the average surface-accessibility scale.

In summary, the ASA scale measures the tendency of each amino-acid to occur at protein surfaces. In that sense, it may be called a “real, statistical hydrophobicity scale” for amino-acids in their natural protein environment, because it is assumed that hydrophobic residues occur preferentially in the protein core, whereas polar ones are found at the protein surface. It is important to notice that the ASA contains all possible forces, because for its determination no assumptions are required concerning the nature of these forces or the mechanism of the folding process. Because of its model-independence, it represents a good measure of the degree to which an amino-acid is driven to the surface of native proteins, and may be used for comparison of results obtained using theoretical HP models with experimental measurements. Although the ASA scale yields a significant improvement in the accuracy of predicting the surface exposure of an amino-acid in protein structures, it is still not perfect, and the knowledge of the amino-acid species alone may not be sufficient to predict its geometrical position within some folded proteins. This observation indicates that further factors, such as intermolecular forces, the exact position of a residue within the protein sequence, and correlations of pairs and triples of amino-acids within sequences, may be required in a full description of protein structures.
In this study we have shown that the qualitative features of the liquid-liquid demixing process of hydrophobic aggregation may be explained successfully within a simple model for aqueous solutions of non-polar particles by including hydrophobic interactions only in terms of changes in water structure. Although the explicit terms of the model Hamiltonian describe solely the states of water molecules in solution, we have demonstrated that it contains implicitly both two- and even many-particle interactions between hydrophobic solute molecules. We have established the phase diagram of the binary system, which shows the characteristic properties of hydrophobic aggregation. These include an upper and a lower critical solution temperature, which define temperature and density ranges in which aggregation occurs. Within the model the high solubility of hydrophobic particles in water at low temperatures is a consequence of the energetically favorable, cage-like rearrangement of water molecules around solute particles. At higher temperatures entropic contributions increase considerably, driving an aggregation of hydrophobic solute particles to minimize the total surface exposed to water. This leads to a reduction in local restructuring of water, and thus to the emergence of a two-phase coexistence region in the phase diagram. We have been able to show the occurrence of an upper critical solution temperature, beyond which the system forms one homogeneous phase, where the solute particles are once again soluble, due to dominant entropy effects. Overall, we have demonstrated that the hydrophobic effect may be described by a model which focuses on the formation of two distinct types of solvent, and on changes in the structural arrangements of liquid water, rather than on any direct interactions of the hydrophobic solute particles.

We have also studied the aggregation of hydrophobic particles in aqueous solutions in the presence of cosolvents. Here we demonstrated that the important features of the destabilizing effect of chaotrope cosolvents on hydrophobic aggregates may be described within the same two-state model, with adaptations to focus on the ability of such substances to alter the structurally ordered arrangements of water molecules around solute particles. The relevant phenomena include a significant enhancement of the solubility of hydrophobic particles in the presence of chaotropic substances, and preferential binding of these substances to solute molecules. We were able to show
that the rather simple mixture model reproduces the implicit effects of chaotropic substances on the many-body interactions between solute molecules.

In a similar fashion, we analyzed the physical mechanism underlying the stabilizing effect of kosmotropic cosolvents on hydrophobic aggregates in aqueous solutions. By altering the state degeneracies one may include the ability of kosmotropic substances to enhance the structure of liquid water, which leads to preferential exclusion of the cosolvent from the hydration shell of hydrophobic solute particles. We have shown that high concentrations of kosmotropic cosolvents stabilize aggregates of hydrophobic particles, leading to an enhancement of the coexistence region. We have demonstrated that this preferential exclusion is a consequence of the energetically favorable enhancement of bulk water structure due to strong hydrogen-bond formation between solvent and cosolvent molecules. As a result, the cosolvent density increases in the bulk and decreases in the shell, causing a relative increase in shell water density. We illustrated the enhancement of the hydrophobic effect due to this high shell water density, and thus the stabilization of solute aggregates.

The overall results for solutions including cosolvent molecules demonstrate that the stabilizing or destabilizing effect may be explained to a large extent purely by the propensity of cosolvents to enhance or reduce water structure. The properties of aqueous solutions of hydrophobic particles including cosolvents are described rather well by our extension of an MLG-type model, which focuses on reproducing the altered structural arrangements of the two distinct types of solvent.

We have extended the MLG model to include the solvation of amphiphilic solute particles in water. By implementing a simplified model which allows the description of different distributions of hydrophobic regions at the molecular surface, we have found aggregation of the solute particles and formation of various types of micelle as a function of the hydrophobicity pattern. We then refined this model to describe surfactant molecules of varying length by adapting their interaction with water molecules. The strong interaction of a long hydrophobic tail with a correspondingly large number of water molecules is incorporated in the model as an increase in the energy and degeneracy of a shell site representing the neighboring solvent particles. We have demonstrated that the essential features of micelle formation, attributed to the amphiphilic nature of the solute particles under consideration, may be reproduced by our extension of the model to describe alterations of water structure in the vicinity of different surface regions of the amphiphiles in solution.

Hydrophobicity remains a controversial quantity also in protein science. We have found that well-known hydrophobicity scales for amino-acids frequently disagree with each other, and often fail to predict the degree of surface exposure of a residue in a folded protein structure. Statistical studies have shown that some amino-acids are found preferentially either in the core or on the surface of native folds, whereas many
are present at all positions with equal probability. Based on these results we have presented the average surface-accessibility scale, which may lead to an improvement in the comparison of experimental data with the results from theoretical HP models. We have verified the validity of the new scale on secondary-structure elements (α-helices and β-strands).

In summary, we have shown that the primary features of the hydrophobic interaction in aqueous solutions may be captured within a model which focuses on the ability of water to form structurally ordered arrangements around solute molecules. Extensions of the model to describe solutions containing cosolvent particles confirms the validity of the framework. Because we were able to demonstrate that the simple mixture model contains implicitly the many-body interactions between the solute molecules, reproduces the implicit effects of chaotropic and kosmotropic substances on these interactions, and describes the formation of micelles for amphiphilic molecules, the study constitutes an important contribution towards advancing the qualitative and quantitative understanding of the hydrophobic effect. These results should be of interest to both experimental and theoretical communities working on aqueous systems, biological reactions, and protein structures.
Résumé en français

Malgré son importance dans notre vie de tous les jours, certaines propriétés de l’eau restent inexppliquées. L’étude des interactions entre l’eau et les particules organiques occupe des groupes de recherche dans le monde entier et est loin d’être finie. Dans mon travail j’ai essayé de comprendre, au niveau moléculaire, ces interactions importantes pour la vie. J’ai utilisé pour cela un modèle simple de l’eau pour décrire des solutions aqueuses de différentes particules.

Récemment, l’eau liquide a été décrite comme une structure formée d’un réseau aléatoire de liaisons hydrogènes. En introduisant une particule hydrophobe dans cette structure à basse température, certaines liaisons hydrogènes sont détruites ce qui est énergétiquement défavorable. Les molécules d’eau s’arrangent alors autour de cette particule en formant une cage qui permet de récupérer des liaisons hydrogènes (entre molécules d’eau) encore plus fortes : les particules sont alors solubles dans l’eau. A des températures plus élevées, l’agitation thermique des molécules devient importante et brise les liaisons hydrogènes. Maintenant, la dissolution des particules devient énergétiquement défavorable, et les particules se séparent de l’eau en formant des agrégats qui minimisent leur surface exposée à l’eau. Pourtant, à très haute température, les effets entropiques deviennent tellement forts que les particules se mélangent de nouveau avec les molécules d’eau. En utilisant un modèle basé sur ces changements de structure formée par des liaisons hydrogènes j’ai pu reproduire les phénomènes principaux liés à l’hydrophobicité. J’ai trouvé une région de coexistence de deux phases entre les températures critiques inférieure et supérieure de solubilité, dans laquelle les particules hydrophobes s’agrégent. En dehors de cette région, les particules sont dissoutes dans l’eau. J’ai démontré que l’interaction hydrophobe est décrite par un modèle qui prend uniquement en compte les changements de structure de l’eau liquide en présence d’une particule hydrophobe, plutôt que les interactions directes entre les particules.

Encouragée par ces résultats prometteurs, j’ai étudié des solutions aqueuses de particules hydrophobes en présence de co-solvants cosmotropiques et chaotropiques. Ce sont des substances qui stabilisent ou déstabilisent les agrégats de particules hydrophobes. La présence de ces substances peut être incluse dans le modèle en décrivant leur effet sur la structure de l’eau. J’ai pu reproduire la concentration élevée de co-solvants chaotropiques dans le voisinage immédiat de la particule, et l’effet
inverse dans le cas de co-solvants cosmotropiques. Ce changement de concentration du co-solvant à proximité de particules hydrophobes est la cause principale de son effet sur la solubilité des particules hydrophobes. J’ai démontré que le modèle adapté prédit correctement les effets implicites des co-solvants sur les interactions de plusieurs corps entre les particules hydrophobes.

En outre, j’ai étendu le modèle à la description de particules amphiphiles comme des lipides. J’ai trouvé la formation de différents types de micelles en fonction de la distribution des régions hydrophobes à la surface des particules.

L’hydrophobicité reste également un sujet controversé en science des protéines. J’ai défini une nouvelle échelle d’hydrophobicité pour les acides aminés qui forment des protéines, basée sur leurs surfaces exposées à l’eau dans des protéines natives. Cette échelle permet une comparaison meilleure entre les expériences et les résultats théoriques.

Ainsi, le modèle développé dans mon travail contribue à mieux comprendre les solutions aqueuses de particules hydrophobes. Je pense que les résultats analytiques et numériques obtenus éclairent en partie les processus physiques qui sont à la base de l’interaction hydrophobe.
Bibliography


