Molecular Dynamics Simulations of Protein Adsorption at Interfaces

Giovanni Bruno Brandani

Doctor of Philosophy
The University of Edinburgh
May 2016

Abstract

Proteins can often adsorb irreversibly at fluid/fluid interfaces; the understanding of the adsorption mechanism has relevance across a variety of industrial (e.g. the creation of stable emulsions) and biological (e.g. biofilm formation) processes. I performed molecular dynamics simulations of two surfactant proteins as they interact with air/water and oil/water interfaces, describing the origin of the surface activity, the adsorption dynamics and the conformational changes that these proteins undergo at the interface.

BslA is an amphiphilic protein that forms a highly hydrophobic coat around *B. subtilis* biofilms, shielding the bacterial community from an external aqueous solution. By investigating the behaviour of BslA variants at oil/water interfaces via coarse-grained molecular dynamics, I show that BslA represents a biological example of an ellipsoidal Janus nanoparticle, whose surface interactions are controlled by a local conformational change. All-atom molecular dynamics simulations then reveal the details of the conformational change of the protein upon adsorption, and the self-assembly into a two-dimensional interfacial crystal.

Ranaspumin-2 is one of the main components of the tungara frog foam nest. Contrary to most surfactant proteins, its structure lacks any sign of amphiphilicity. All-atom simulations show that the adsorption proceeds via a two-step mechanism where firstly the protein binds to the interface through its flexible N-terminal tail and then it undergoes a large conformational change in which the hydrophobic core becomes exposed to the oil phase. I then developed a simple structure-based coarse-grained model that highlights the same adsorption mechanism observed in all-atom simulations, and I used it to compare the dynamics of adsorption and the underlying free energy landscape of several mutants. These results agree with and are used to rationalise the observations from Langmuir trough and pendant drop experiments.

Colloids can often be considered simpler versions of proteins that lack conformational changes. I performed coarse-grained simulations of the compression of interfacial monolayers formed by rod-like particles. These simulations show a rich behaviour characterised by the flipping of adsorbed rods, nematic ordering and bilayer formation. I report the series of transitions that take place as the rod aspect ratio is increased from 3 to 15.

Lay Summary of Thesis

We all know that water is not enough to efficiently remove oil (or fat) from the surface of a dirty dish. This is because water and oil do not like to be in contact with each other and they do not mix. In order to get the oil to mix with the water we need soap, which is made of surfactants, small particles that adsorb at (stick to the edge of) the interface between the two liquids. Surfactants prevent the water from directly coming into contact with the oil, and they allow us to mix these two liquids properly.

The need to create particles that adsorb at interfaces originated long before humans needed to wash their dishes. Micro-organisms and animals need them for many of their biological functions, and to this aim they produce surfactant proteins, a class of biomolecules that can adsorb at the interface between water and oil. For instance, the tropical tungara frog uses the protein Ranaspumin-2 to prepare its foam nest (where the oil is substituted with air, which also does not mix with water). Another example is from the bacterium Bacillus subtilis, which produces the surfactant protein BslA to build a kind of coat that protects the whole bacterial colony from the outside.

Currently, we have a good understanding of how proteins work when they are in water, and of how standard surfactants interact with interfaces, but we do not know so much about the molecular mechanisms that govern the adsorption of proteins. In this thesis, I will show how the proteins Ranaspumin-2 and BslA adsorb at an interface between water and oil. To do this, we used advanced computer simulations running on supercomputers. Specifically, I will show that Ranaspumin-2 has a surprising adsorption mechanism similar to the movement of a clamshell; whereas BslA represents a unique example of a protein that folds itself once it reaches the interface, and then it interacts with other proteins to self-assemble into a strong interfacial crystal.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of this work have been published in [1–3].

(*Giovanni Bruno Brandani, May 2016*)

Acknowledgements

Firstly, I thank all the people that supported me during my PhD, as mentors, collaborators, friends and family members: Alberto Brandani, Cait MacPhee, Dario Dell'Arciprete, Davide Marenduzzo, Enrica Govoni, Giulia Foffano, Ioan Magdau, Jane Patterson, Keith Bromley, Kristel Torokoff, Li Tao, Maria Brandani, Marieke Schor, Masako Murakami, Ryan Morris and Ulrich Zachariae. I also thank the Principal's Career Development scholarship for funding and the HECBioSim for the allocation of computer time on ARCHER.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The importance of protein adsorption at interfaces

Proteins are complex polymers made by a combination of charged, polar and hydrophobic residues. The cell environment is mainly made of water, and this is the solution where most proteins fold into their native structure. However, when a protein finds an interface between water and a hydrophobic fluid (such as oil or air), it tends to adsorb to it [4]. By doing so, the protein can reduce the unfavourable interactions between water and the apolar phase, and it can partition the different types of residues into the most favourable environment [4]. In nature, there are a large number of proteins specifically designed to adsorb at interfaces. For instance, hydrophobins are produced by filamentous fungi to aid the formation of areal structures and help hyphae to attach to hydrophobic surfaces [5], whereas Latherin enhances the evaporation of sweat on the pelt of horses by reducing the surface tension between water and air [6]. The bacterial hydrophobin BslA [7] and the protein Ranaspumin-2 [8] are other examples of surfactant proteins with a precise biological function, and they will be discussed in chapters 3 and 4 of this thesis.

Even outside their natural environment, proteins are widely used as surfactants in industrial applications [4, 9]. Proteins represent ideal candidates for the creation and stabilisation of emulsions [4], which are the basis of many food [4] and pharmaceutical [10] products. There are many factors contributing to this success. Firstly, proteins are biocompatible and naturally occurring; for instance, milk proteins are often used in food products [4]. Secondly, attractive protein-protein interactions can make the resulting emulsions extremely resistant to shear stress [4]; one noteworthy example is the self-assembly behaviour of hydrophobins, which form highly ordered films [11].

In order to understand the function of surfactant proteins in biology and improve their use for industrial applications, it is important to understand the principles that govern their adsorption at interfaces. So far, most of our understanding of protein adsorption comes from experiments. Using either a Langmuir-Blodgett trough [12] or pendant drop tensiometry [9], researchers can study how adsorption affects the surface tension of the air/water or oil/water interface. The change in surface tension from the reference value in the absence of proteins is defined as surface pressure. The data are then interpreted by comparison against theoretical models. For instance, the model by Ward and Tordai [13] is used to test whether adsorption is diffusion-limited or there is an energetic barrier [9], and the surface pressure isotherm by Frumkin [14] can provide an estimate of the strength of protein-protein interactions at the interface [4]. The level of unfolding of the adsorbed proteins can be inferred from the thickness of the film [8] or the area per molecule [12] by comparison against the respective values of the native structure, and changes in the secondary structure content can be qualitatively characterised from circular dichroism experiments [15].

Experiments highlighted some general features about protein adsorption. In the initial stages, when surface coverage is low, adsorption is usually diffusion-limited and irreversible, whereas at later times the proteins that are already at the interface introduce an energetic barrier to further adsorption, which becomes reversible [12]. At low surface coverage, adsorption usually causes the unfolding of the protein, whereas at high surface coverage the native conformation is preserved [4]. Unfolding is stronger at oil/water than at air/water interfaces, because oil solvates hydrophobic residues [12]. In general, adsorption at interfaces causes a decrease of α-helices and an increase of β-sheets [15]. However, interfacial conformational changes are highly dependent on the specific protein considered [4]. For instance, β-Casein adopts completely disordered conformations characterised by ‘trains’ (all residues lie on the interface) and ‘loops’ (protrusions out of the interface) [12]; Lysozyme partially unfolds but some secondary structure elements are present even at low surface coverage [12]; the highly surface active hydrophobin EAS does not unfold at the interface, but

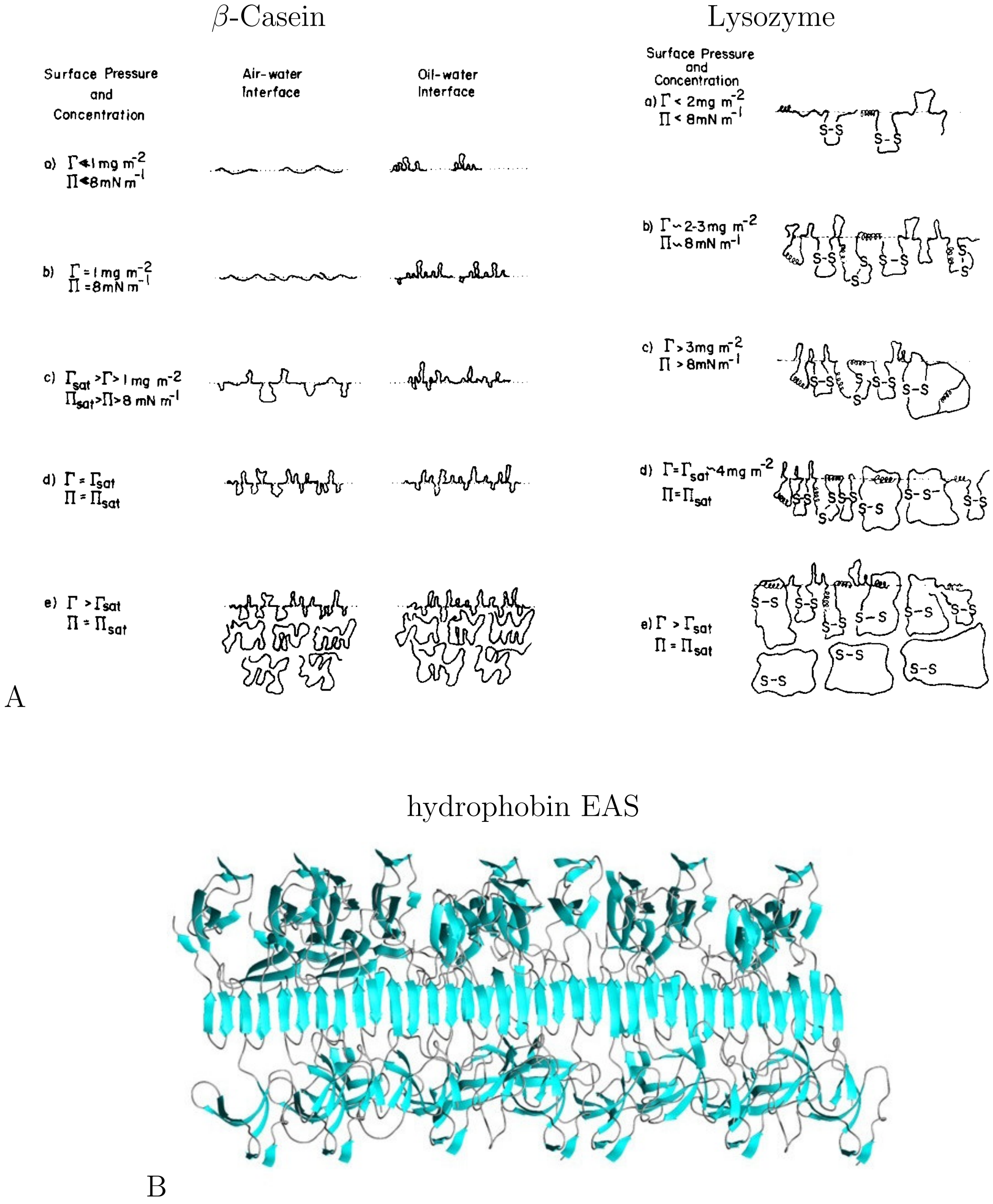


Figure 1.1 *(A) Qualitative representation of β-Casein and Lysozyme interfacial conformational changes from Ref. [12] as a function of the density* Γ *and the surface pressure* Π*. β-Casein is highly disordered, forming 'trains' at the air/water interface at low density and 'loops' at high density. At the oil/water interface loops are enhanced because the oil phase solvates hydrophobic residues. At low density, Lysozyme is only partially unfolded at interfaces, and, as the density increases, the protein maintains its native structure upon adsorption. (B) Model of hydrophobin EAS self-assembly from Ref. [16]. The protein undergoes a conformational change that allows it to associate into a rodlet stabilised by a amyloid fibril.*

it undergoes a conformational change that allows it self-assembly into stiff linear chains [16] (or rodlets, see Fig. 1.1).

1.2 Motivation for the study

The main limit of the characterisation of protein adsorption in experiments is that it is often only qualitative, in particular regarding conformational changes and self-assembly [9, 12]. Moreover, the complex nature of proteins makes the prediction of their interfacial behaviour from the native conformation particularly challenging. In this thesis I will show that molecular dynamics (MD) simulations represent a powerful tool to complement experimental observations by providing molecular details that would not be otherwise accessible.

MD simulations have already contributed to unravel the microscopic behaviour of a number of proteins at interfaces [17–20] (see background chapter). In what follows I will apply them to the study of two important surfactant proteins: BslA [7] and Ranaspumin-2 [8] (Rsn-2). Firstly, understanding the surface activity of these two proteins is important for the control of their function in-vivo and for their industrial applications as biocompatible surfactants [21, 22]. BslA forms a protective raincoat around *B. subtilis* biofilms [23] and it has been used to produce a slower melting ice-cream [21]. Rsn-2 is the fundamental ingredient that stabilises the foam nest of the tropical tungara frog [8] and it has been used to engineer a foam for artificial photosynthesis [22]. Secondly, the aim is to test a wide variety of approaches for MD simulations, and to show how they can be useful for the study of the microscopic forces that govern protein adsorption, conformational changes and self-assembly at interfaces. In this way, I hope to create transferable strategies that will be useful for the wider community of researchers interested in surface active proteins.

1.3 Thesis outline

In chapter 2 I will review the relevant scientific background. I will start from the physics behind the adsorption of colloidal particles at interfaces and the structure of proteins. Then, I will discuss the theory of MD simulations, focusing on the main computational methods used during my PhD. I will conclude with a review of the past applications of MD simulations to the study of protein adsorption at interfaces.

Chapter 3 concerns the surface activity of the protein BslA. Firstly, I will show that the energetics of adsorption of BslA can be well captured by a simpler continuum model of a Janus ellipsoidal colloid. Then, I will show that all-atom simulations highlight a conformational change upon adsorption, where an unstructured region of the protein in solution becomes highly ordered and hydrophobic at the interface. Finally, I will propose a model for the BslA interfacial self-assembly in agreement with the current experimental observations, and I will report the binding free energies of BslA dimers obtained from MD simulations.

The surface activity of the protein Ranaspumin-2 is discussed in chapter 4. Here I will show that all-atom and coarse-grained MD simulations consistently highlight the same two-step adsorption mechanism. In the first step, the disordered and hydrophobic N-terminal tail captures the interface, and in the second one the protein unfolds as a clamshell by exposing the hydrophobic residues in the core towards the apolar phase. Interestingly, the observed unfolding is not accompanied by a change in secondary structure content. The employed coarse-grained model has been developed and tested for this system, but it is transferable to the study of the unfolding of other proteins at interfaces.

In the final results chapter I will discuss the behaviour of colloids at interfaces, whose adsorption is driven by the same forces that control protein adsorption, but are simpler in that they lack conformational changes. I will consider the compression of an interfacial monolayer formed by rod-like particles of various aspect ratios. MD simulations show that the ordering of the rods and the mechanism employed to release the compression stress highly depends on the aspect ratio.

In the conclusions I will discuss the key findings presented in the previous chapters, compare the surface activities of BslA and Ranaspumin-2, and comment on what this work can teach about the use of MD simulations for the study of surfactant proteins.

Chapter 2

Background

The nature of the topic, the adsorption of proteins at interfaces, makes this background section rather heterogeneous. Firstly, I will introduce the general concepts that will be useful for the understanding of the considered systems, starting from a discussion on the structure of proteins, followed by the theoretical background on colloidal adsorption at interfaces. Then, I will review the theory of molecular dynamics simulations, focusing on the force-fields used to represent the dynamics of proteins, the algorithms for the integration of the equations of motion, and the enhanced sampling techniques that will be employed throughout the thesis.

2.1 Background on proteins and interfaces

2.1.1 The structure of proteins

Proteins are polypeptide chains, heteropolymers formed by amino acids covalently bound together [24]. We refer to the unique sequence of amino-acids that defines a protein as primary structure. All proteins share the same series of nitrogen, carbon and oxygen atoms linked together, forming the backbone, but the differences between amino acid side chains confer a unique character to each protein. In solution, each protein folds into a unique native conformation [25], even though conformational changes [26] and disorder [27] (see later on intrinsically disordered proteins) are also common. Protein folds display only a few types of structural motifs, or secondary structure elements, each of them characterised by a specific pattern of backbone hydrogen bonds and dihedral angles. The most common elements are α-helices and β-sheets; in the former, the structure is stabilised by hydrogen bonds between the backbone C=O group of residue i and the backbone N-H group of residue i+4, whereas in the latter, two or more strands in an extended conformation are held together by inter-strand backbone hydrogen bonds. The tertiary structure represents the organisation of the different secondary structure elements of the protein into a compact domain.

One of the key principles that determines how the protein secondary structure elements will form the tertiary structure is the hydrophobic dipole [28] (see section 3.2.3 for the precise definition). Amino-acids can be classified according to the hydrophobicity of their side-chains; a large hydrophobic dipole of a secondary structure element corresponds to a highly asymmetric distribution of the residues and it confers an amphiphilic character to the element. Within a protein, two secondary structure elements tend to interact in a way to cancel the total hydrophobic dipole [28], so that the hydrophobic sides of the two regions are closed on top of each other, whereas the polar sides are exposed to the solvent. In chapter 4 will see that this principle also governs the adsorption mechanism of the protein Ranaspumin-2 [8] at interfaces.

One of the biggest questions of biology is how each protein folds into its native conformation in a reasonable amount of time [25]. This problem is also known as Levinthal's paradox [29], a though experiment that shows that if a protein would have to find its native conformation by performing a random search through all the possibilities it would take much longer than the typical experimental folding time-scales, which is normally within the order of milliseconds. The solution to this paradox is that the motion of a protein is not random, but directed towards the native state, moving in what appears to be a funnel-shaped energy landscape [25]. This is also why computational models based solely on the topology of the native state are very popular at representing the folding pathway of proteins [30]. This idea will also be useful for modelling protein unfolding at interfaces (see section 4.3).

However, such structural characterisation does not apply to intrinsically disordered proteins (IDP) [27]. Until recently, the function of a protein was considered tightly linked to its structure, but it has been now recognised that there is a large number of proteins that perform important functions despite being disordered in solution [27]. Intrinsically disordered regions often undergo a disorder-to-order transition upon binding to a target site, and a single region can even adopt different conformations depending on the binding partner [31]. DNA-binding proteins are often rich in intrinsically disordered regions, which are for instance employed by transcription factors to efficiently search the target sequence along the DNA [32], or by histones to regulate the condensation of DNA into chromatin [33]. Throughout this thesis, I will argue that disordered regions are also important for the adsorption mechanism of the surfactant proteins BslA and Rsn-2.

2.1.2 Colloidal adsorption at interfaces

The art of coarse-graining allows scientists to use simple models to provide insights into complicated problems. Because of their globular structure, folded proteins can often be coarse-grained into much simpler colloidal particles, and be studied using the wide range of experimental and theoretical techniques that are typically employed in soft matter physics [34]. For instance, the crystallisation of both proteins and colloids is only controlled by the second virial coefficient of the system [34]. This coefficient represents the deviation from the ideal gas behaviour given by the pairwise interaction between two particles. Given the complexity and specificity of protein-protein interactions, this is a fundamental result, that enables to easily predict the conditions under which protein crystallisation will occur. As I will argue in section 3.2, the driving forces to the adsorption of proteins at interfaces are the same as those of simpler colloidal particles. In what follows, I will explain what is an interface and why colloids like to attach to it, even though they may lack the typical amphiphilic structure of small surfactants, characterised by a polar head and an apolar tail [35].

Water and oil are the simplest example of a pair of fluids that do not mix at ambient temperature. The surface tension γ (or interfacial tension) is defined as the excess energy per unit area due to the creation of an interface between two immiscible fluids, i.e. the energy cost of an interface of area A between two fluids is given by $E = \gamma A$, which means that two immiscible fluids will always tend to minimise the interfacial area between them (and therefore phase separate). But what is the origin of the surface tension? The simplest way to demonstrate the existence of an interfacial energy is by considering a model of a (symmetric) mixture of two fluid components, where the free energy of the system is given by the following Ginzburg-Landau functional [36]:

$$F(\phi) = \int dV \left(\frac{a}{2}\phi^2 + \frac{b}{4}\phi^4 + \frac{k}{2}(\nabla\phi)^2 \right) \tag{2.1}$$

where ϕ is a position-dependent function proportional to the difference between the local concentrations of the two components, b and k are positive parameters whereas a is positive only above a certain critical temperature T_c, and negative below. The given free energy can be used to model the demixing of many different fluid mixtures near the critical temperature T_c in a general way. The values of the parameters a, b and k will depend on the microscopic details of the considered system and they can be extracted from experiments or computer simulations. If we solve Eq. 2.1 by imposing the conservation of the components, we find that above T_c the solution of ϕ with the lowest energy is the homogeneous mixture with $\phi_0 = 0$ everywhere. On the other hand, below T_c the state with the lowest energy is obtained for a demixed system separated by an interface (that is taken along $\hat{z}$ at position z_0):

$$\phi_I(z) = \sqrt{-a/b}\tanh\left(\frac{z - z_0}{\sqrt{-k/2a}} \right) \tag{2.2}$$

where $\phi_b = \pm\sqrt{-a/b}$ correspond to the bulk concentrations of the two fluids far away from the interface ($z = \pm\infty$). The energy cost for the creation of the interface is then obtained by subtracting the free energy of the demixed system in Eq. 2.2 by the one of the bulk state with $\phi = \phi_b$ everywhere, and the surface tension is simply this excess energy per unit area:

$$\gamma = \int dz \, (F(\phi_I(z)) - F(\phi_b)) = \left(\frac{-8ka^3}{9b^2} \right)^{1/3} \tag{2.3}$$

Note that at the critical temperature T_c the surface tension is zero (since a=0), which means that there is no energy cost to create the interface, and the two fluids start to mix together.

In general, two fluid components can demix when the attraction between components of the same type is higher than the attraction between the two different components, despite the decrease in entropy of mixing. For instance, polar water molecules form a network of strong hydrogen bonds in bulk [37]; however, these bonds cannot be formed between water and apolar oil molecules. This means that the system has to pay a free energy penalty for the inclusion of every oil molecule into water. When this energy cost is high enough (or temperature is low enough), phase separation becomes favourable, since the total cost for a homogeneous mixture scales with the oil volume, whereas that of a demixed system scales with the interfacial area [37].

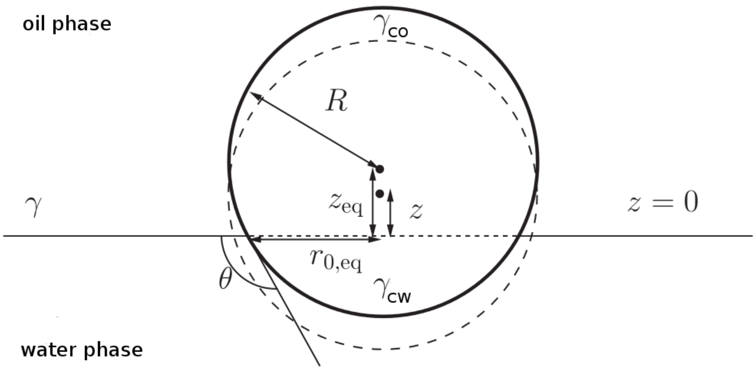


Figure 2.1 *Adsorption of a spherical colloid at an oil/water interface. θ is the contact angle between the surface of the colloid and the interface. Image reproduced from Ref. [38].*

The interfacial energy described above also exists for solid/liquid interfaces, which will have their own surface tensions. Let us now consider a spherical colloid of radius R adsorbed at an oil/water interface as in Fig. 2.1. The free energy of the system then depends on the surface tensions γ, γ_{CW} and γ_{CO} respectively between water and oil, colloid and water, and colloid and oil [38]:

$$E = -\gamma A_I + \gamma_{CW} A_{CW} + \gamma_{CO} A_{CO} \tag{2.4}$$

where A_{CW} and A_{CO} are respectively the areas between colloid and water, and colloid and oil, and A_I is the area of the oil/water interface that has been covered by the particle. By minimising the free energy in Eq. 2.4, we can find the equilibrium contact angle (see Fig. 2.1) between the surface of the particle and the interface [38]:

$$\theta = \arccos\left(\frac{\gamma_{CO} - \gamma_{CW}}{\gamma}\right) \quad , \tag{2.5}$$

and the total free energy of adsorption relative to the most favourable fluid phase [38]:

$$\Delta G_{\text{ads}} = \pi \gamma R^2 (1 - |\cos\theta|)^2 \tag{2.6}$$

This shows that the maximum free energy of adsorption is achieved for neutral colloids with a contact angle of 90°. For $\theta > 90°$ the colloid is considered hydrophobic, whereas for $\theta < 90°$ the colloid is considered hydrophilic. At oil/water interfaces, colloids can adsorb irreversibly if they are larger than a few nm (a size comparable to that of proteins). This is why they can replace standard surfactants and be employed to make stable emulsions [39] or bijels [40].

The condition of the equilibrium contact angle has to be satisfied at all points along the contact line between three phases (colloid, water and oil) regardless of the shape of the colloid. Hence, for most particle shapes the contact line cannot be flat, and the adsorption of the colloid causes a deformation of the interface between the two fluids [41]. In turn, this deformation generates anisotropic capillary interactions that drive the self-assembly of the particles at the interface [42–44].

Interesting interfacial behaviours can also be observed for the adsorption of Janus particles, where the surface chemistry is not uniform but divided into polar (hydrophilic) and apolar (hydrophobic) sides [45]. The amphiphilic character of Janus particles confers them higher interfacial activity compared to neutrally wetting colloids [46]. Moreover, for non-spherical Janus particles, the competition between the minimisation of the different interfacial energies (there will be 5 surface tensions) gives rise to multiple stable orientations of the particle at the interface [47]. We will see that this scenario is analogous to the one found in the adsorption of the protein BslA [7] (section 3.2).

2.2 Background on molecular dynamics simulations

While purely analytical tools are often sufficient for the understanding of colloidal adsorption at interfaces [48], the same is not true for protein adsorption, because of their non-uniform shape and surface chemistry, and their tendency to undergo conformational changes upon adsorption [12]. Furthermore, it is experimentally very challenging to observe the dynamics of protein adsorption at the molecular level. Hence, molecular dynamics (MD) simulations are necessary to provide a complete understanding of the surface activity of proteins, and to interpret correctly the experimental observations. More generally, MD simulations are useful whenever we wish to have a dynamic picture of static structural data (for instance a protein crystal structure). In what follows, I will review the background on MD simulations, with particular emphasis on the models and methods employed for the studies described in the results chapters.

2.2.1 Force-fields: all-atom and coarse-grained models

In the context of MD simulations, we refer to the force field as the set of parameters and functional forms of the inter-particle interactions that define the Hamiltonian of the system [49]. To simulate the dynamics, the equations of motion are then integrated in a chosen thermodynamic ensemble, which is usually NVT (constant number of particles, volume and temperature) or NPT (constant number of particles, pressure and temperature). In general, the force field potential is the sum of several m-body interactions (m=2,3,..), and for proteins it can be typically written as a sum of bonds, angles, dihedral angles, Van der Waals and electrostatic interaction terms [50]:

$$\begin{aligned} U_{\text{tot}} = & \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} K_n \left[1 + \cos(n\phi - \gamma)\right] + \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon r_{ij}^2} \right] \end{aligned} \tag{2.7}$$

where b is the bond length, θ the bond angle, ϕ the dihedral angle, and r_{ij} the distance between particles i and j. The coefficients K_n, K_θ, K_n, A_{ij}, B_{ij}, q_i, q_j, b_0, θ_0, γ and ϵ depend on the types of the particles involved in the interaction and they are carefully parametrised to reproduce the correct behaviour of the system.

All-atom force fields represent each individual atom in the system; they are computationally expensive, but they are very accurate and transferable. MD simulations with all-atom force fields have been successfully used to describe complex processes such as protein folding [51], conformational changes [52] and oligomerization [53]. However, they are still not able to perfectly represent intrinsically disordered proteins [54], and various approaches such as changes in the protein-water interactions [54] and experimental restraints [55] have been investigated to solve this problem. I will employ all-atom force fields (AMBER ff99SB-ILDN [50] and ff99SB*-ILDN [50, 56, 57]) to study the adsorption of Rsn-2 (section 4.2), and the conformational change and self-assembly of BslA upon adsorption (section 3.3).

In order to reduce the computational cost of simulations, coarse-grained force

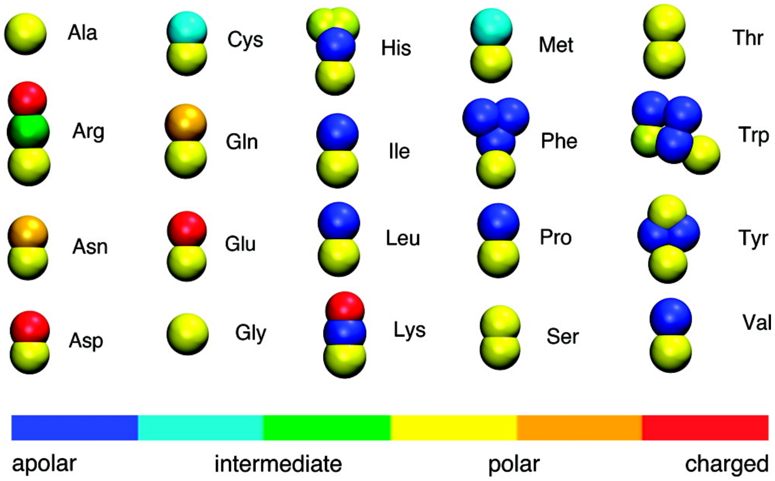


Figure 2.2 *MARTINI coarse-grained model of amino-acids [58]. On average, 4 heavy atoms are mapped onto a single coarse-grained bead, classified according to its hydrophobicity and charge.*

fields map groups of atoms into single beads according to certain rules that depend on the precise model [30, 58]. This approach allows simulations of larger systems for longer time-scales, at the expense of resolution and accuracy. In physics-based models the inter-particle potential is based on the chemical properties of the components. For instance, in the popular MARTINI model the amino acids side chains are coarse-grained according to their polarity using a 4 to 1 mapping of the heavy atoms [58] (see Fig. 2.2). This model has been applied to the study of large systems where all-atom simulations would be computationally too expensive, such as the aggregation of membrane proteins [59]. I will use the MARTINI force-field to study the orientational transition of BslA at interfaces and to compute its free energy of adsorption (section 3.2). In structure-based models the interaction potential is not based on the chemistry of the components, but is specifically designed to reproduce certain experimental observations. A common approach to model protein folding and conformational changes is to define a favourable interaction between every pair of residues that form a native contact in the NMR or crystal structure of the protein [30].

Very often, the optimal strategy is to combine physics-based and structure-based models together. For instance, in MARTINI, the natural fluctuations of the protein are best reproduced when the original force field is combined with an elastic network model based on the native structure [60] (as we will see for the protein BslA in section 3.2). In section 4.3 I will model the Rsn-2 adsorption using a structure-based potential for the protein, and defining a physics-based interaction with the interface (using the water to oil partitioning free energies of the amino acid side chains [61]).

2.2.2 Integrating the equations of motion

In order to simulate the dynamics of the system, we have to integrate the equations of motion, that in the NVE ensemble of constant number of particles, volume and energy, are simply the equations of Newton:

$$\begin{aligned} \frac{d\mathbf{r}_i}{dt} &= \mathbf{v}_i \\ \frac{d\mathbf{v}_i}{dt} &= \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i}\frac{dU}{d\mathbf{r}_i} \end{aligned} \tag{2.8}$$

where $\mathbf{r}_i$, $\mathbf{v}_i$, $\mathbf{F}_i$ and m_i are the position, the velocity, the force and the mass of particle i=1,..,N, and U is the potential energy. A system of first order differential equations evolves from time zero to time t according to:

$$\mathbf{\Gamma}(t) = \exp(iLt)\mathbf{\Gamma}(0) \tag{2.9}$$

where L is the Liouville operator satisfying $iL = \dot{\Gamma} \cdot \nabla_\Gamma$ and Γ is the vector of independent variables. In our case Γ represents positions and velocities of the particles. The evolution operator can be discretised by making use of the Trotter decomposition [62]:

$$\exp(iLt) = \exp(iLn\Delta t) = \prod_{k=1}^{n} \exp(iL\Delta t) \tag{2.10}$$

In the NVE ensemble of Eq. 2.8, the Liouville operator can be split into two additive operators that act on velocities and positions independently:

$$iL = iL_1 + iL_2 = \sum_{j=1}^{N} \mathbf{v}_j \cdot \nabla_{\mathbf{r}_j} + \sum_{j=1}^{N} \frac{1}{m_j}\mathbf{F}_j \cdot \nabla_{\mathbf{v}_j} \tag{2.11}$$

Then, a symmetric, and hence reversible, approximation of the short-time evolution operator is given by:

$$\exp(iL\Delta t) = \exp(iL_2\Delta t/2)\exp(iL_1\Delta t)\exp(iL_2\Delta t/2) + O(\Delta t^3) \tag{2.12}$$

If we expand the exponentials to the linear order in the small timestep Δt, the integration of the equations of motion can be implemented as:

$$\begin{aligned}
\mathbf{v}(t+\Delta t/2) &= \mathbf{v}(t) + \frac{\Delta t}{2m}\mathbf{F}(t) \\
\mathbf{r}(t+\Delta t) &= \mathbf{r}(t) + \Delta t \mathbf{v}(t+\Delta t/2) \\
\mathbf{v}(t+\Delta t) &= \mathbf{v}(t+\Delta t/2) + \frac{\Delta t}{2m}\mathbf{F}(t+\Delta t)
\end{aligned} \tag{2.13}$$

This algorithm is referred to as velocity Verlet [63]. Alternatively, the leapfrog integration [64] follows the Trotter decomposition:

$$\exp(iL\Delta t) = \exp(iL_1\Delta t)\exp(iL_2\Delta t) + O(\Delta t^3) \tag{2.14}$$

This operator is still symmetric as long as the full velocity step is performed between the times $t - \Delta t/2$ and $t + \Delta t/2$:

$$\begin{aligned}
\mathbf{v}(t+\Delta t/2) &= \mathbf{v}(t-\Delta t/2) + \frac{\Delta t}{m}\mathbf{F}(t) \\
\mathbf{r}(t+\Delta t) &= \mathbf{r}(t) + \Delta t \mathbf{v}(t+\Delta t/2)
\end{aligned} \tag{2.15}$$

The two algorithms lead to the same result up to numerical precision in the NVE ensemble, but they differ in the way the kinetic energy is estimated at each time step, which will affect the result in a thermodynamic ensemble with constant temperature, e.g. NVT or NPT.

A wide range of algorithms have been developed to simulate the coupling of the system to an external reservoir that maintains a constant temperature (thermostat) or pressure (barostat). The Berendsen thermostat [65] mimics the weak coupling to a heat reservoir by rescaling the velocities of the system at each timestep according to:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.16}$$

where T is the current temperature of the system and τ is the characteristic time of the exponential relaxation towards the reference temperature T_0. This thermostat does not generate a proper thermodynamic ensemble, but the error scales with $1/N$, and most of the ensemble averages will not be significantly affected for very large systems. The temperature coupling by Bussi et al. [66] is a modified version of the Berendsen thermostat with an additional stochastic term that ensures a correct canonical sampling of configurations. The Nose-Hoover thermostat [67], which also produces a correct canonical ensemble, modifies the

Hamiltonian dynamics of the system by adding a dissipative term that controls the coupling to the external reservoir:

$$\frac{d\mathbf{v}_i}{dt} = \frac{\mathbf{F}_i}{m_i} - \frac{p_\xi}{Q}\frac{d\mathbf{r}_i}{dt} \tag{2.17}$$

$$\frac{dp_\xi}{dt} = T - T_0 \tag{2.18}$$

$Q = \tau^2 T_0/4\pi^2$ is a coupling constant related to the oscillatory relaxation time of the system τ. Barostatting is obtained by rescaling the system box in order to maintain a certain reference pressure $\mathbf{P}_0$. An exact NPT ensemble can be ensured using the Parrinello-Rahaman pressure coupling [68]. In this barostat, the system box matrix satisfies its own equation of motion, which is coupled to the equations of motion for the particles via a dissipative term, in a similar way to the Nose-Hoover thermostat.

2.2.3 Free energy calculations and enhanced sampling techniques

In order to describe MD simulations and to make sense of complex biomolecular processes, it is often useful to project the original high-dimensional space onto a low-dimensional free energy landscape [69, 70]. This is done by defining a set of collective variables $\mathbf{s}(\mathbf{r})$, functions of the microscopic configuration $\mathbf{r} = (\mathbf{r}_1, .., \mathbf{r}_N)$, that can efficiently capture the behaviour of the system, and in particular the transitions between the different metastable states. For instance, these variables can be used to study the folding pathway of a protein or the adsorption at an interface. The free energy of the system as a function of the collective variables $\mathbf{s}$, also known as the potential of mean force [71] (PMF), is defined from the average distribution function $\langle\rho(\mathbf{s})\rangle$ as:

$$\begin{aligned} F(\mathbf{s}) &= -k_B T \ln\langle\rho(\mathbf{s})\rangle \\ &= -k_B T \ln \int d\mathbf{r} e^{-U(\mathbf{r})/k_B T} \delta(\mathbf{s} - \mathbf{s}(\mathbf{r})) \end{aligned} \tag{2.19}$$

where T is the temperature, k_B is the Boltzman constant and $U(\mathbf{r})$ is the potential energy of the system (the equalities are defined up a non-physical constant). In principle, the PMF could be simply calculated from a histogram of the average distribution function $N(\mathbf{s}) \sim \langle\rho(\mathbf{s})\rangle$. In practice, this is often impossible when one wish to employ an all-atom force-field, since we are usually interested in studying rare events where the system transition between two or more metastable states. In unbiased MD simulations the protein will spend a long time near its initial conformation, without being able to explore the relevant phase space within the available computational time. I will now give an overview of some of the most popular strategies to enhance the sampling of the conformational space in MD simulations, which will be also employed throughout the thesis.

In parallel tempering [72] (or replica exchange) multiple replicas of the same system are simulated in parallel at different temperatures. At a certain rate, we attempt to exchange the configurations of neighbouring replicas with temperatures T_1 and T_2. The exchange is accepted with a probability that ensures the condition of detailed balance:

$$P(1 \leftrightarrow 2) = \min\left(1, e^{(1/k_B T_1 - 1/k_B T_2)(U_1 - U_2)}\right) \tag{2.20}$$

where U_1 and U_2 are the current potential energies of the two replicas. This exchange allows the replicas at ambient temperature to escape their local energy minima by jumping at a higher temperature, where they can easily overcome free energy barriers and diffuse through the whole phase space. I will use replica exchange simulations to enhance the conformational change of BslA adsorbed at the interface (section 3.3.5).

Umbrella sampling [73], on the other hand, aims to enhance rare events by introducing a non-physical bias potential $V(\mathbf{s})$. The correct estimate of the potential of mean force can then be recovered from the obtained average distribution function using:

$$F(\mathbf{s}) = -\frac{1}{\beta} N(\mathbf{s}) - V(\mathbf{s}) \tag{2.21}$$

The form of the potential $V(\mathbf{s})$ that optimise the sampling of the phase space is exactly the opposite of the PMF. In practice, a series of umbrella samplings with harmonic potentials centred at different positions along $\mathbf{s}$ can be combined together using the weighted histogram analysis method [74] to produce the full PMF. We will see an application of this approach in section 3.2, where I will use it compute the orientational free energy landscape of BslA at the interface. Alternatively, in well-tempered metadynamics [75], the bias potential is constructed during the simulation as a sum of Gaussians: every time τ, a Gaussian of width σ, height w and centred at the current values of the collective variables $\mathbf{s}$ is added to the bias potential; the height w is set to $w = w_0 e^{-V/k_B \Delta T}$, where V is the current value of the bias potential. w_0 is the initial Gaussian height and ΔT is a parameter with the dimension of a temperature. Under this history-dependent bias, free-energy barriers are effectively reduced by a factor of $\frac{T+\Delta T}{T}$, and at long times, the bias converges to a fraction of the potential of mean force: $F(\mathbf{s}) = -\frac{T+\Delta T}{\Delta T}V(\mathbf{s})$. In section 4.3, the full free energy landscape of Rsn-2 adsorption at interfaces has been obtained from coarse-grained metadynamics simulations.

In order to study a specific transition, the system can also be steered from an initial state with collective variable s_0 to a target state s, using a harmonic potential moving at constant speed v [76]. In the limit $v \to 0$ the external work $W(s)$ performed on the system will be equal to the free energy difference between the initial and the final states. If v is finite, the system dynamics is generally out of equilibrium and the work depends on the initial microscopic configuration of the system. However, the equilibrium PMF can still be reconstructed using the Jarzynski equality [77]:

$$F(s) = -\frac{1}{\beta}\ln\langle e^{-\beta W(s)}\rangle \tag{2.22}$$

where the average is performed over a set of pulling simulations starting from a Boltzmann-weighted ensemble of initial configurations with the initial collective variable s_0. This relation holds arbitrarily far from equilibrium, but the sensitivity of the external work to the initial microscopic state increases with increasing pulling velocity, so that higher statistics is required to give a correct estimate of the potential of mean force [76]. I used steered MD simulations in combination with the Jarzynski equality to compute the free energy of adsorption of BslA at an oil/water interface (section 3.2) and the binding free energies of the two BslA dimers involved in the self-assembly of the protein into an interfacial crystal (section 3.3.6).

2.3 Review of protein adsorption at interfaces from MD simulations

In recent years there has been a large number of works investigating the surface activity of proteins at fluid interfaces via MD simulations.

Many of these focused on the study of hydrophobins [11], whose surface activity resides in their native amphiphilic structure stabilised by a network of disulphide

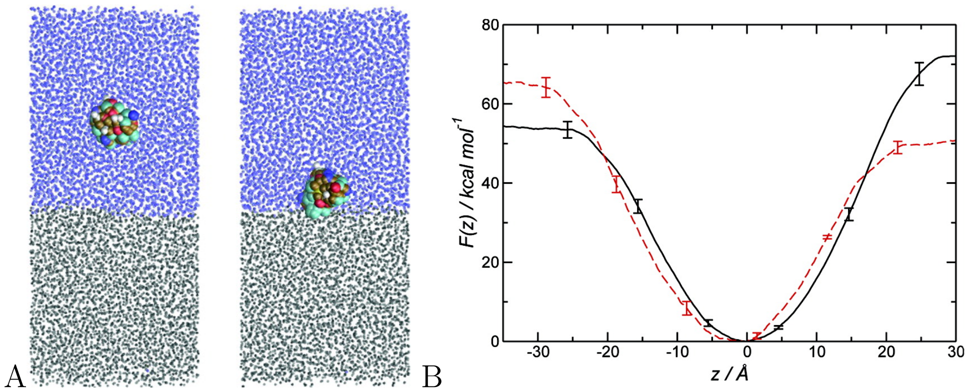


Figure 2.3 *(A) Snapshots of coarse-grained simulations of a hydrophobin adsorbing at a octane/water interface. (B) Free energies of adsorption of hydrophobin HFBI (dashed line, red) and HFBII (solid line, black) at a octane/water interface. Images reproduced from Ref. [18].*

bridges. Hydrophobins HFBI and HFBII self-assemble into interfacial crystals without undergoing significant conformational changes from their native solution structures [78]. In these cases, physics-based coarse-grained force fields such as MARTINI [58], which cannot model conformational changes, can be used to study protein adsorption, and they produce results in good agreement with all-atom simulations (at a much lower computational cost) [79]. David Cheung computed the free energy of adsorption of hydrophobins HFBI and HFBII at a octane/water interface via steered MD simulations with a coarse-grained force field [18]. His results showed that the heterogeneous distributions of polar and apolar residues on the surface of these proteins is important for the surface activity, and therefore their behaviour is different from that of uniform colloidal particles. In another interesting study on the hydrophobin EAS using a combination of all-atom and coarse-grained MD simulations [17], the authors suggested that the protein controls its self-assembly behaviour via the intrinsic disorder of its loop region. In the conclusions of the thesis, I will compare the role of disorder in EAS with those found in the proteins BslA and Rsn-2.

All-atom simulations are ideally suited for the study of protein conformational changes upon adsorption, but because of their high computational cost, there are still very few examples of such studies. Simulations of β-Lactoglobulin at a decane/water interface [19] showed that the adsorption is followed by a partial unfolding where hydrophobic residues come into contact with the interface while preserving the secondary structure content, in agreement with experimental observations showing that proteins often retain residual native structure at interfaces [12]. In Ref. [20], the authors performed replica exchange simulations of Lysozyme adsorbing at a 1,2-dichloroethane/water interface under the influence of an electric field, showing that the protein unfolds and loses secondary structure content after adsorption. However, it remains to be established whether these all-atom simulations are able to reach convergence and reproduce the equilibrium ensemble of protein conformations at the interface. The study by Deighan and Pfaendtner [80] on the adsorption of two short peptides (14 and 15 residues) at solid interfaces represents a unique example of the use of all-atom simulations to map the complete free energy landscape of an adsorption process. In this study, the authors used a particular combination of metadynamics and parallel tempering [81] to exhaustively sample the phase space of the system, computing the PMF along the distance from the interface and identifying of the most favourable adsorbed conformations.

These studies show that molecular dynamics simulations represent a powerful computational tool for the understanding of protein adsorption at the molecular level. However, despite the many recent advancements in the field, high computational cost and convergence in the simulations remain the biggest problems to be tackled, and appropriate enhanced sampling and/or coarse-graining approaches should be employed for this purpose.

Chapter 3

Adsorption and self-assembly of the protein BslA

BslA is a bacterial hydrophobin that self-assembles at interfaces and forms a hydrophobic coat around *B. subtilis* biofilms. Here I show, by means of molecular dynamics simulations, that the behaviour of BslA at interfaces is the same as that of much simpler Janus ellipsoidal colloids. However, differently from rigid colloidal particles, the protein cap region is highly plastic and it undergoes a disorder-to-order transition upon adsorption. Furthermore, I show how BslA is able to self-assemble at the interface via a set of symmetric lateral interactions based on salt-bridges.

3.1 Introduction

Bacterial colonies are often organised as interface-associated aggregates referred to as biofilms. They spontaneously form in natural or industrial environments if a surface is rich enough in nutrients, and they play an important role in global ecology, water treatment and health-related problems [82]. Within a biofilm, cells are embedded in the so-called extracellular matrix (ECM), which is mainly composed of polysaccharides and proteins. The ECM is important for the regulation of the local environment of the biofilm and the cooperation amongst the micro-organisms [82, 83].

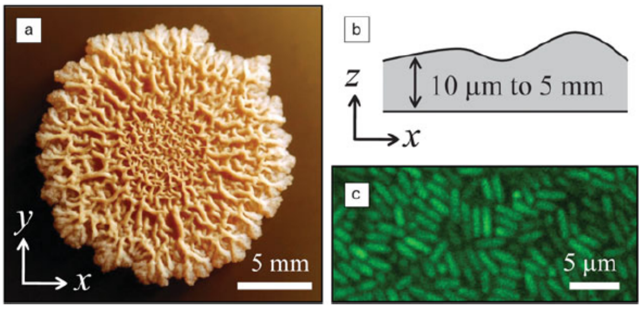


Figure 3.1 *(a) Wrinkled morphology of a B. subtilis biofilm grown at a air/agar interface. (b) Typical biofilm thickness. (c) Cells embedded in the extracellular matrix. Image reproduced from Ref. [82].*

B. subtilis is a Gram-positive soil bacterium that forms biofilms [82] (see Fig. 3.1). It represents a model organism for laboratory studies, where it can either grow on top of agar plates, or it can form floating biofilms called pellicles at the air/liquid interface [7]. The extracellular matrix is composed of three main ingredients: an exopolysaccharide (EPS) and the two proteins TasA [84] and BslA [7]. It has been proposed that TasA self-assembles into fibers whose purpose is to give structural integrity to the biofilm by binding cells together [84]. These fibers, together with EPS, may promote interaction between cells and adhesion to the surface. TasA and EPS are all necessary for the correct formation of a wild type biofilm.

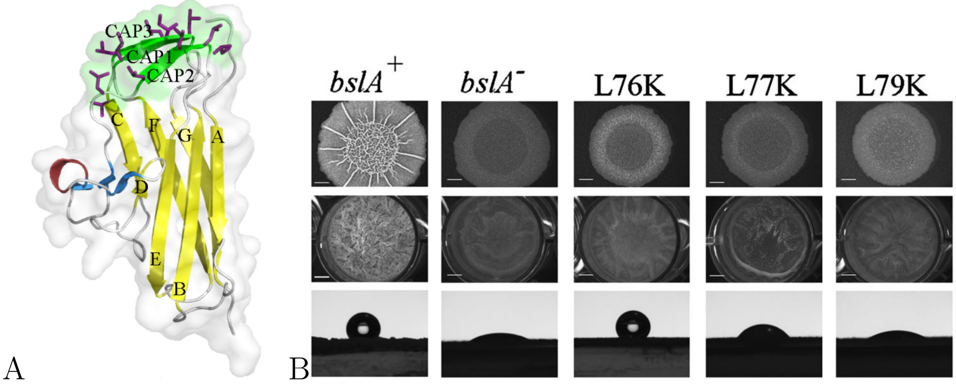


Figure 3.2 *(A) A single monomer from the BslA crystal structure, with the β-sheet cap highlighted in green, and the cap hydrophobic side chains in purple. (B) B. subtilis biofilm (top), pellicle (center) and hydrophobicity (bottom) for the following bacterial strains: BslA-positive, BslA-negative, L76K-BslA, L77K-BslA and L79K-BslA. Reproduced from Ref. [7].*

The role of BslA, which has only recently been uncovered [23], is to self-assemble into a hydrophobic layer around the biofilm, creating a protective barrier against antimicrobial agents, and stabilizing the internal conditions of the biofilm [85]. BslA is also necessary for the formation of the three-dimensional pattern of wrinkles of the biofilm [23] (Fig. 3.2B).

Thanks to the crystal structure of BslA (Fig. 3.2A), reported by Hobley et al. in Ref. [7], it was possible to link the activity of the protein to its strong amphiphilicity, which prompted the authors to call BslA the first example of a bacterial hydrophobin. BslA has a unique conformation consisting of an extended hydrophobic cap connected to an immunoglobulin fold [7]. It has been shown that mutations that reduce the hydrophobicity of the cap region (L76K, L77K, L79K; the hydrophobicity can be quantified by the hydrophobic dipole, see Fig. 3.13) have strong effects on biofilm formation, preventing the formation of the characteristic pattern of wrinkles and reducing the hydrophobicity of the biofilm itself (Fig. 3.2B). Specifically, mutations L77K and L79K generate a biofilm that is essentially indistinguishable from a BslA-negative one.

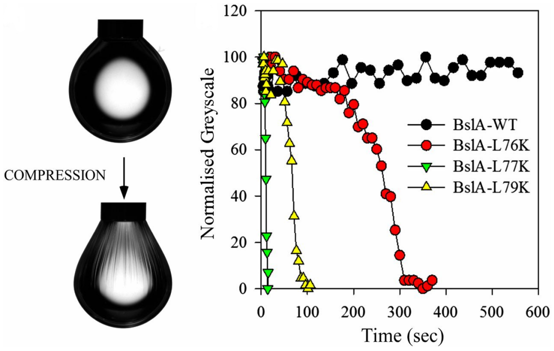


Figure 3.3 *Experimental wrinkle relaxation of BslA films at the oil/water interface for wt-BslA, L76K-BslA, L77K-BslA and L79K-BslA. Mutations in the cap region reduce the stability of the elastic film at the interface. Reproduced from Ref. [7].*

In-vitro, BslA adsorbs at air/water and oil/water interfaces forming a strongly elastic film [7]. In pendant drop experiments, the protein firstly adsorbs at the surface of a water drop surrounded by oil, forming an interfacial film. Decreasing the volume of a drop surrounded by a mature BslA film causes the formation of wrinkles, which signal the elasticity of the film. For wt-BslA these wrinkles are very stable over time, whereas for the considered cap mutants, the wrinkles relax on the order of seconds to minutes (Fig 3.3), indicating that the film is weaker and the proteins are migrating back into bulk until the film reaches the new equilibrium area. Bromley et al. [1] showed that the elasticity of BslA films lies in the ability of the protein to self-assembly at the interface into highly ordered structures, corresponding to a rectangular lattice with lattice spacings equal to 43 and 39 Å (Fig. 3.4). The L77K cap mutation causes a large decrease in the order of the film, even though small ordered patches can still be identified.

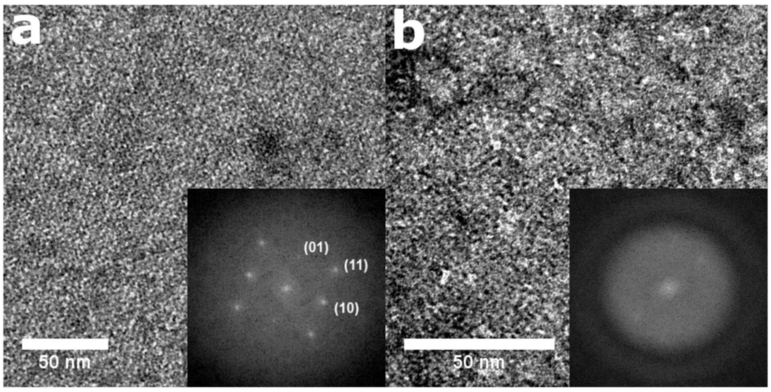


Figure 3.4 *TEM images of BslA monolayers from Ref. [1]: (a) wt-BslA and (b) L77K-BslA. Wt-BslA forms an ordered rectangular lattice with lattice spacings of 43 and 39 Å. The L77K mutation strongly reduces the order of the monolayer, even though small ordered patches can still be observed. Reproduced from Ref. [1].*

The same authors also performed circular dichroism experiments on wt-BslA in solution and at the interface, which brought evidence of a conformational change corresponding to an increase in β-sheet secondary structure content upon adsorption [1]. This may be able to explain why, despite the strong surface activity of the protein, wild type BslA does not adsorb rapidly at interfaces [1]. On the other hand, perhaps even more surprisingly, the observations on the L77K mutant are consistent with a simple diffusion-limited model of adsorption. A theoretical model for the adsorption kinetics of BslA is discussed in details in appendix A.

Despite the large number of experiments on BslA, there is still no clear picture of the molecular mechanism of BslA adsorption and self-assembly at interfaces. The aim of this work is to fill these gaps by means of molecular dynamics (MD) simulations. In the first section of this chapter, I will discuss the energetics of adsorption and the orientation of the protein with respect to the interface by drawing a parallel with Janus ellipsoidal colloids [86], and I will show that the tilting transition observed in these types of colloids is also observed in BslA from coarse-grained MD simulations. In the second section, I will use all-atom simulations to reveal the molecular details of the BslA activity starting from its solution structure up to the formation of the two-dimensional interfacial crystal. Firstly, I will show that the cap of BslA undergoes a conformational change upon adsorption, from a disordered and not very hydrophobic structure in solution to an ordered and highly hydrophobic one at the interface. The low hydrophobicity of the solution structure of the protein is perfectly consistent with the existence of a barrier in the adsorption of the protein (in appendix A). Finally, I will propose a model of BslA self-assembly starting from a set of protein-protein interactions identified in the crystal structures of BslA and YweA, another protein related to BslA.

Taken together, these results identify BslA as a biological realisation of a switchable Janus ellipsoidal colloid, whose surprising properties arise from its ability to tune the hydrophobicity of its cap region depending on the environment, and to self-assemble into an interfacial crystal via lateral protein-protein interactions.

3.2 BslA as a Janus ellipsoidal colloid: a coarse-grained study

3.2.1 Introduction

The geometry and surface chemistry of BslA, with a clear separation between the hydrophobic cap and the polar immunoglobulin domain [7], is reminiscent of that of Janus particles [45]. This similarity is also highlighted by the unit cell of the BslA crystal structure: a decameric micelle of proteins where the hydrophobic cap points towards the center of the micelle (Fig. 3.5A), which is a form of aggregation that is also present in the phase diagram of Janus particles [87]. However, differently from the spherical Janus colloids that are commonly considered in experiments and simulations [45, 87], BslA is approximately shaped as an ellipsoid with one major axis of length $l_{\mathrm{BslA,max}} = 55\text{Å}$ and two minor ones of lengths $l_{\mathrm{BslA,mid}} = 33\text{Å}$ and $l_{\mathrm{BslA,min}} = 26\text{Å}$ (as estimated from the protein's inertia equivalent ellipsoid [88]). The hydrophobic cap of BslA polarises the protein along its major axis, since there is only a small angle (~ 8 degrees) between this axis and the direction of the cap relative to the protein center of mass.

As suggested from earlier simulations of micron-size Janus ellipsoidal particles [86, 89], this type of geometry and amphiphilicity gives rise to an interesting competition between hydrophobic and interfacial forces at the interface. Since

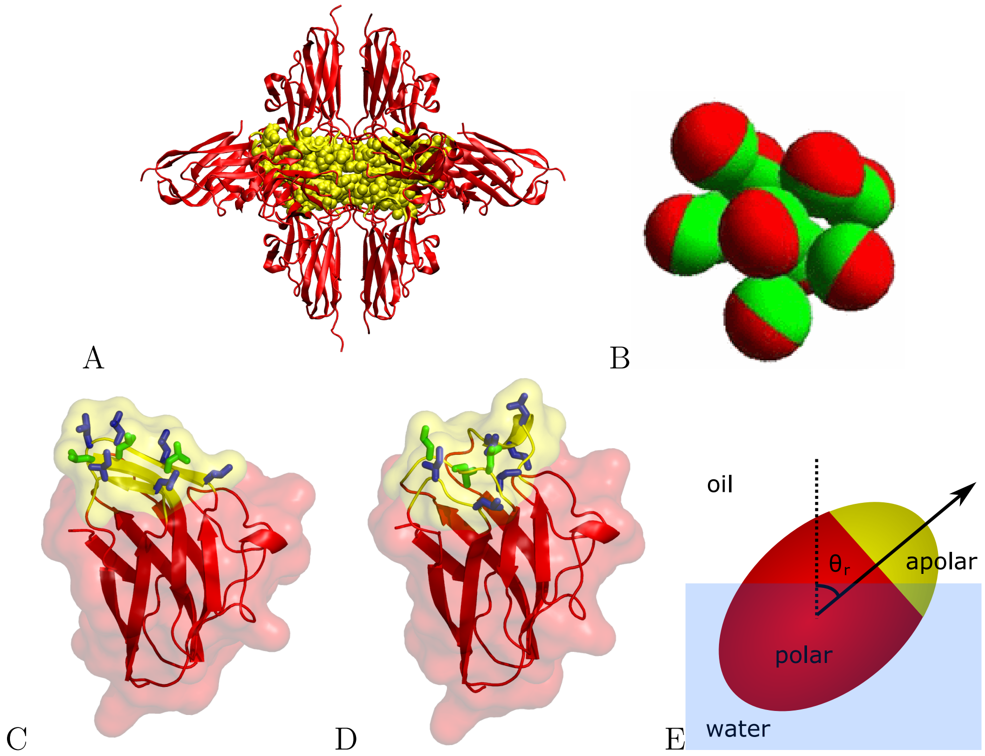


Figure 3.5 *(A) BslA crystal decamer (PDB ID: 4bhu [7]). The hydrophobic cap (residue ids 75-80,119-125,152-156) is highlighted in yellow, whereas the polar immunoglobulin domain is in red. (B) Decamer of Janus particles found in the Monte-Carlo simulations of Ref. [87]. (C-D)* L_{out} *(C) and* L_{in} *(D) conformations from the BslA crystal structure. Large hydrophobic residues in the cap are shown as blue (outer) or green sticks (inner strand). (E) Schematic of an ellipsoidal Janus colloid adsorbed at a water-oil interface, with the apolar and polar surfaces respectively in yellow and red.* θ_r *is the orientation angle with respect to the interface.*

the particle is polarised along its longest axis, an upright (or end-on) orientation optimises particle-solvent interactions, whereas a parallel (or side-on) orientation optimises the interfacial energy. A change in the hydrophobicity of the apolar cap can induce a transition between these two orientations [47, 86, 89]. At high cap hydrophobicity, the orientation of the particle is perpendicular to the interface, so that it can efficiently partition its cap into the air or oil phase, and its polar surface into the water phase. On the other hand, at low cap hydrophobicity, the particle lies on its side (tilted), so that it can increase the area occupied at the interface and minimise the interfacial energy.

The possibility to observe different orientations of BslA at an interface is suggested by the presence of markedly different conformations in the crystal structure. Eight monomers (referred to as chains A to H in the PDB file) in the BslA decamer have an ordered β-sheet cap where all hydrophobic residues are accessible to the solvent (Fig. 3.5C). On the other hand, the last two chains in the crystal micelle (chains I and J) show a disordered cap where many of the leucine residues have their hydrophobic side-chains buried into the protein core (Fig. 3.5D). I will refer to the former set of BslA conformations as $\mathrm{L}_{\mathrm{out}}$ (leucines pointing out of the cap), and to the latter set as L_{in} (leucines pointing into the cap). It is interesting to notice that most structural differences between BslA monomers are located within the hydrophobic cap (Fig. 3.6). Furthermore, a change in cap hydrophobicity may also be induced via residue mutations. Hobley et al. [7] showed that the L76K, L77K and L79K mutations in the cap region affect the behaviour of BslA both in-vivo and in-vitro; specifically, they cause a loss of the characteristic pattern of wrinkles and hydrophobicity of the biofilm, and they reduce the elasticity of the protein film obtained in pendant drop experiments (Fig. 3.2). Both cap conformational changes and mutations may provide a route to tune cap hydrophobicity and induce a transition in the BslA orientation at the interface. In turn, the precise orientation that BslA adopts at the interface may also influence protein-protein interactions and self-assembly (Fig. 3.4).

The aims of this project are to understand the effect of cap mutations on the adsorption and the role of the conformational variability observed in the BslA crystal structure, and to discuss the orientational transition suggested by the theory. To this aim, I studied the adsorption of a single BslA protein at an oil/water interface using two sets of simulations with a different level of coarse-graining. In the first part I will show, by means of Monte-Carlo simulations of a Janus ellipsoid, that for the size and geometry of BslA, a change in cap

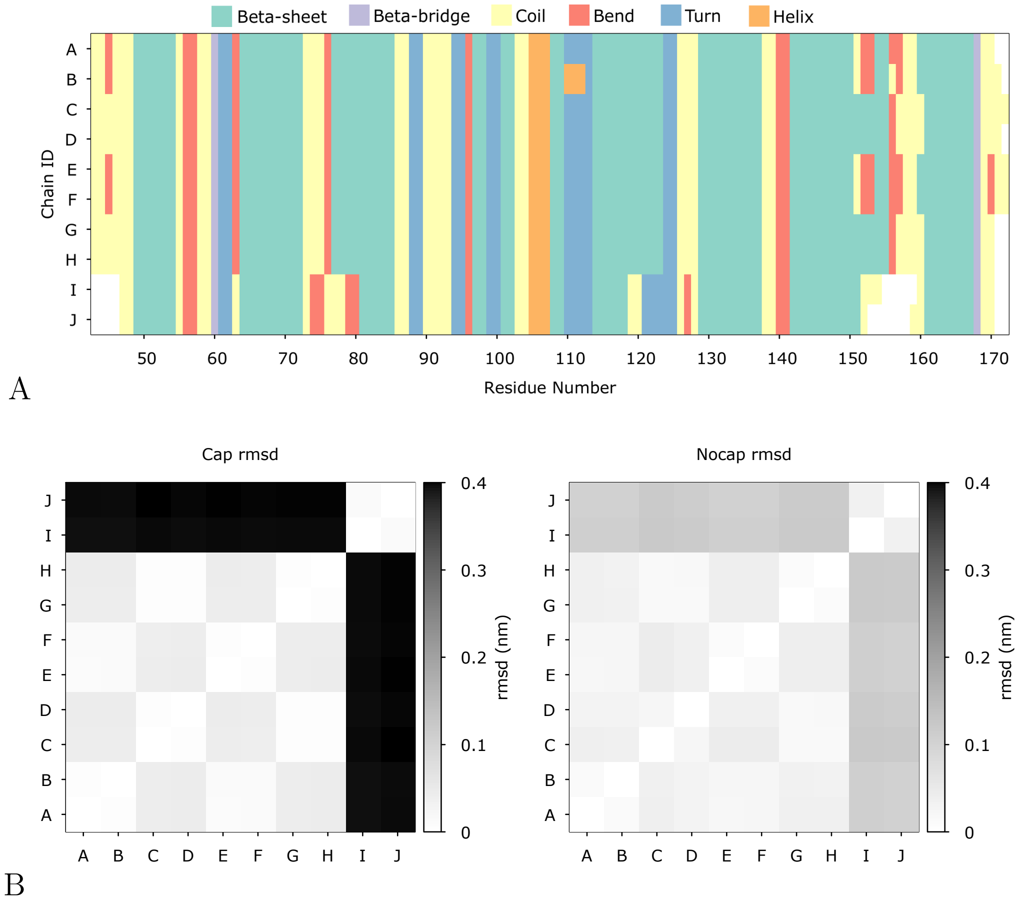


Figure 3.6 *(A) Secondary structure content of the 10 BslA monomers present in the crystal structure. (B) Root mean square displacement between the backbone atomic positions of the 10 chains: on the left I selected the residues in the cap region (residue ids 75-80,119-124,151-156), and on the right those outside of it. Conformational changes amongst the BslA monomers are localised in the cap region of the protein. We can clearly identify the two main clusters of structures:* L_{out}*, containing chains A to H, and* L_{in}*, containing chains I and J.*

hydrophobicity should induce a transition in the orientation of the protein at the interface. In the second part I will show that higher resolution coarse-grained MD simulations at a cyclohexane/water interface do reveal the expected tilting transition. Furthermore, the free energy of adsorption of different BslA variants suggests that BslA-$\mathrm{L_{in}}$ corresponds to the solution state of BslA, and that the protein undergoes a conformational change into the $\mathrm{L_{out}}$ form upon adsorption.

3.2.2 Results: adsorption of a BslA-like Janus ellipsoidal colloid

In order to test whether BslA has the potential to display a tilting transition at the interface, I study the adsorption of a BslA-like Janus ellipsoidal colloid at an oil/water interface, using the same model described in Ref. [89]. Using the approximation of a flat fluid/fluid interface, the energy of an adsorbed Janus particle relative to its bulk water state is given by [86]:

$$\begin{aligned} E &= (\gamma_{AO} - \gamma_{AW})A_{AO} + (\gamma_{PO} - \gamma_{PW})A_{PO} - \gamma A_I \\ &= \gamma(A_{AO}\cos\theta_A + A_{PO}\cos\theta_P - A_I) \end{aligned} \tag{3.1}$$

where A_{AO} and A_{PO} are the apolar and polar surface areas, respectively, exposed to the oil phase, and A_I is the area of interface covered by the colloid. γ, γ_{AO}, γ_{AW}, γ_{PO} and γ_{PW} are the surface tensions of the oil/water, apolar-side/oil, apolar-side/water, polar-side/oil and polar-side/water interfaces. θ_A and θ_P are the contact angles of the apolar and polar surfaces of the colloid (θ increases with the hydrophobicity of the surface). In this study, I consider a Janus ellipsoidal colloid whose geometry closely resembles that of BslA. The ellipsoid axes were set to $l_{\mathrm{col,max}} = 52\text{Å}$ and $l_{\mathrm{col,mid}} = l_{\mathrm{col,min}} = 28\text{Å}$, and the colloid was polarized along the longest axis, with the apolar side located on one of the ellipsoid tips, covering 18% of the total surface.

In Ref. [89], the authors considered micron-size colloids; for these large sizes, the free energy of adsorption is on the order of thousands k_BT. Therefore, entropic contributions can be ignored and the most favourable orientation of the adsorbed colloid was found by simply minimising the energy term given in Eq. 3.1 [89]. However, for the protein BslA, whose dimension is on the nm scale, fluctuations may be important, and Monte-Carlo simulations [90] are more appropriate than a simple energy minimisation. At each Monte-Carlo step of the simulation, I attempt to change the distance between the colloid center of mass and the interface, and the colloid orientation. Rotations are performed using the quaternion formalism [91] (see appendix B). The move is then accepted with a probability set by the Metropolis criterion [90]: p=min$(1, e^{-\Delta E/k_BT})$, where ΔE is the difference between the energy of the system after and before the proposed move, evaluated from Eq. 3.1. To compute the energy, the areas of the polar and apolar surfaces of the ellipsoid exposed to the oil phase are evaluated using a computational scheme where I sum over the infinitesimal elements of area of the points on the surface of the ellipsoid whose z-coordinate is greater than zero, the position of the interface. The area of interface A_I that is occupied by the ellipsoid is computed using the analytical expression [92]:

$$A_I = \pi \left(1 - \frac{d^2}{d_t^2}\right) \frac{abc}{d_t} \quad \text{, with} \quad d_t = \sqrt{a^2 n_1^2 + b^2 n_2^2 + c^2 n_3^2} \quad , \tag{3.2}$$

where d is the distance of the ellipsoid center of mass from the interface, a=$l_{\mathrm{min}}/2$, b=$l_{\mathrm{mid}}/2$, c=$l_{\mathrm{max}}/2$, and $\hat{n} = (n_1, n_2, n_3)$ is the normal to the interface in the reference frame defined by the ellipsoid axes, with $\hat{z}$ corresponding to the longest axis, $\hat{y}$ to the middle one and $\hat{x}$ to the shortest.

I performed Monte-Carlo simulations of the BslA-like ellipsoid adsorbed at an ideal oil/water interface with surface tension equal to γ=49mN/m (=12k_BT/nm^2 at ambient temperature), a value close to that of typical n-alkane/water interfaces [93]. I set the contact angle of the polar side to θ_P=60° (hydrophilic), and I vary the contact angle of the apolar side from θ_A=90° (neutral) to θ_A=180° (hydrophobic), representing the variability of the BslA cap hydrophobicity due to mutations and/or conformational changes. In Fig. 3.7A, I show that as the hydrophobicity is decreased, corresponding to a change in θ_A from 180° to 90°, there is a transition from an end-on orientation to a side-on orientation; the former orientation optimises the particle-solvent interactions, $E_{PS} = \gamma(A_{AO} \cos\theta_A + A_{PO} \cos\theta_P)$, whereas the latter one optimises the interfacial energy, $E_I = -\gamma A_I$, at the expenses of E_{PS} (Fig. 3.7B). The averages reported in Fig. 3.7 were obtained from Monte-Carlo simulations longer than at least 50 times the typical reorientation time of the colloid.

These results suggest that the geometry of BslA allows for a tilting transition via cap modifications under the typical experimental conditions. Specifically, BslA conformations with low cap hydrophobicity should adsorb with a side-on orientation parallel to the interface, whereas the conformations with high

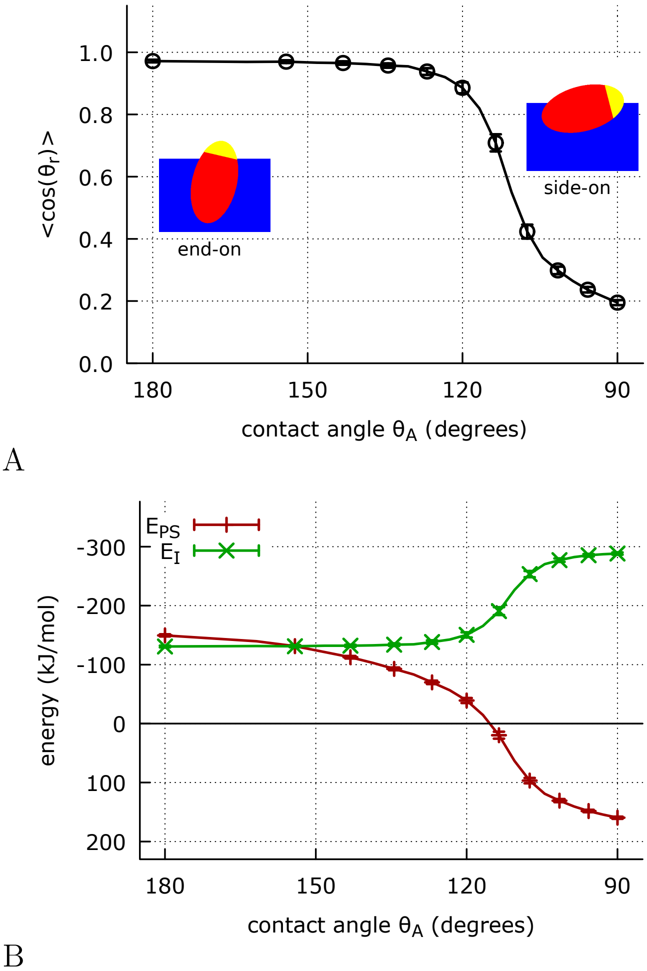


Figure 3.7 *Results of the Monte-Carlo simulations of the Janus ellipsoidal colloid. (A) Average orientation* $\cos\theta_r$ *and (B) energy contributions as a function of the apolar-side contact angle. As the hydrophobicity is decreased from* θ_A*=180° to 90° there is a transition from a end-on orientation dominated by the particle-solvent energy* $E_{PS} = \gamma(A_{AO}\cos\theta_A + A_{PO}\cos\theta_P)$*, to a side-on orientation dominated by the interfacial energy* $E_I = -\gamma A_I$*. The orientation angle* θ_r *is defined as the angle between the main axis of the colloid and the normal to the interface. In panel A, the left inset refers to the end-on orientation at high hydrophobicity, whereas the right inset refers to the side-on orientation at low hydrophobicity.*

cap hydrophobicity should adsorb with a end-on orientation perpendicular to the interface. In what follows, I will show that the proposed transition is indeed observed during higher resolution MD simulations of BslA adsorbed at a cyclohexane/water interface.

3.2.3 Methods: coarse-grained molecular dynamics simulations

MD simulations were performed in the NAP_zT ensemble using the software GROMACS 4 [94]. The system was thermostated at T=300 K using the method by Bussi et al. [66] with a relaxation time of 1 ps, and it was barostated at P_z=1 bar using the Parrinello-Rahman method [68, 95] with a relaxation time of 5 ps. In the simulations, the protein is immersed in a biphasic system consisting of water and cyclohexane (as in Fig 3.9). Both protein and solvents are represented according to the polarizable version of the MARTINI coarse-grained force field [96, 97]. This force field is called polarizable in the sense that the side-chains of polar amino acids are represented by two partial charges, so that their electric dipole can adapt to the environment, instead of a single neutral bead as in the standard MARTINI model [58]. In this model, on average, four heavy atoms in the protein are mapped onto one coarse-grained bead. One coarse-grained water molecule is made by three beads and it represents four actual water molecules. Cyclohexane molecules are made by three beads. The choice of this specific apolar phase was motivated by the fact that amino acids were parametrised from the experimental free energies of transfer from water to cyclohexane [96], and that the surface tension is in reasonable agreement with experiments (γ_{MD}=44 mN/m vs γ_{exp}=50 mN/m [98]). The equations of motion were integrated with the leap-frog algorithm [64], using a time-step equal to dt=20 fs. The global relative permittivity of the system was set to ϵ_r=2.5, and both electrostatic and van der Waals interactions are cut-off at 12 Å. These are the standard settings for the MARTINI polarizable model [96, 97].

In this study I considered the wild-type form of BslA (wt-BslA) and its L76K, L77K and L79K mutants (studied in Ref. [7]), each of them in their L_{out} and L_{in} conformations, for a total of 8 BslA variants. I identified the L_{out} (chains A to H) and L_{in} (chains I and J) clusters in the BlsA_{48-172} crystal structure (PDB ID: 4bhu [7]) by using the gromos clustering algorithm implemented in the GROMACS utility g_cluster [94, 99]. The two clusters are uniquely identified for a wide range of rmsd cut-off values (0.5-1.9 Å, with the default value being 1 Å). Chains C and I from the crystal structure were taken as representative of the L_{out} and L_{in} conformations respectively. Residues 155-159 (part of the third cap strand) are not resolved in chain I; the atomic positions of the missing residues were taken from chain C after an alignment with chain I and energy minimisation using the AMBER ff99SB-ILDN force field [50]. Mutations were applied using the Mutagenesis Wizard tool implemented in the software PyMOL [100].

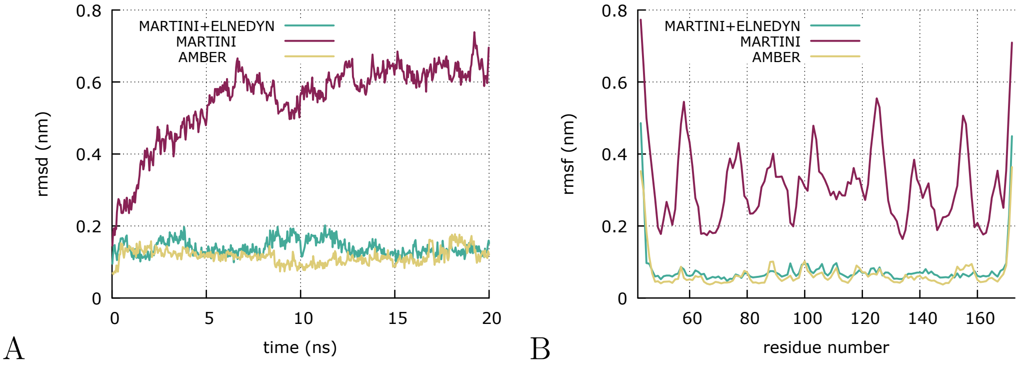


Figure 3.8 *(A) Root mean square deviation over time and (B) root mean square fluctuation per residue of BslA C_α atomic positions. I compared three different force fields: AMBER ff99SB-ILDN [50], MARTINI 2.2P [96] and MARTINI 2.2P combined with an elastic network (ELNEDYN) [60]. The spring constant and the cut-off of the ELNEDYN model are set to $K=500$ kJ/mol nm^2 and $r_c=0.9$ nm respectively. The rmsd is computed with respect to the structure at t=0, whereas the rmsf is computed with respect to the crystal structure.*

The coarse-grained topology of the protein variants were generated starting from the corresponding atomistic structures using the MARTINI tool Martinize [96]. In order to match the natural flexibility of the protein and to avoid conformational changes at the interface (which are not part of this study and will be considered later in Section 3.3 using an appropriate atomistic force field), I combined the MARTINI force field with an elastic network model (ELNEDYN), as suggested in Ref. [60]. In ELNEDYN, a set of harmonic springs are applied between any two non-consecutive C_α atoms within a certain cut-off distance. The parameters of the elastic network (spring constant k=500 kJ/mol nm^2 and cut-off distance r_c=9 Å) were chosen by optimising the root mean square deviation and the root mean square fluctuation per residue of BslA-L_{out} in water with respect to the values obtained from all-atom simulations with the AMBER ff99SB-ILDN force field [50] (Fig. 3.8).

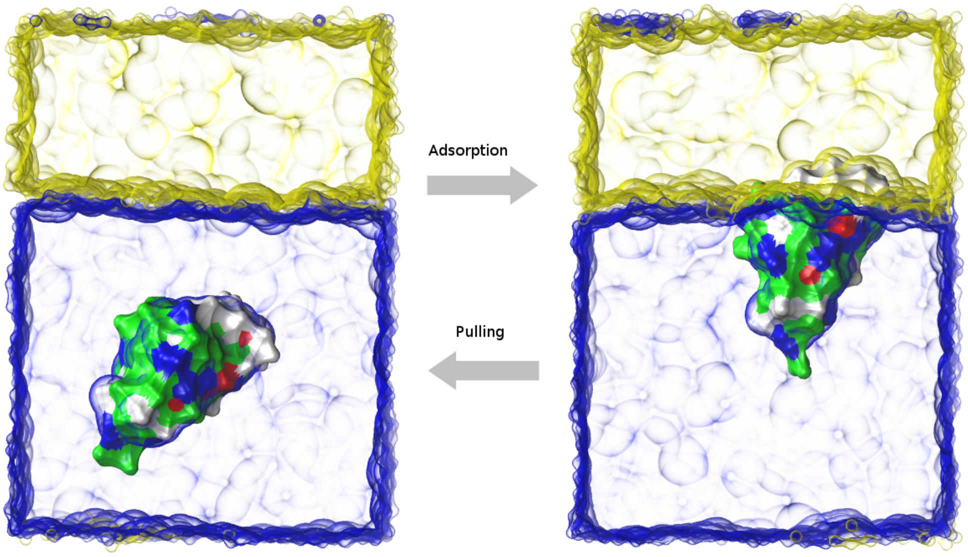


Figure 3.9 *Computational setup of the coarse-grained melecular dynamics simulations: I performed simulations of both spontaneous adsorption (left) and non-equilibrium steered MD simulations, where the protein is pulled back into the aqueous phase from its equilibrium position at the interface (right). Water is in blue, cyclohexane in yellow, hydrophobic residues in white (note the cap region), polar residues in green and charged residues in blue and red.*

In order to study the spontaneous adsorption of BslA variants at the oil/water interface, the protein is initially placed in the aqueous section of a 9x9x14 nm box filled with 65% (volume) water and 35% cyclohexane (Fig. 3.9). Due to periodic boundary conditions, two interfaces are created along the $\hat{z}$ direction. The simulation then proceeds until the protein adsorbs to the interface and the system reaches equilibrium.

The free energy of adsorption from the aqueous phase to the interface was reconstructed for each considered BslA variant from a set of non-equilibrium steered MD simulations, in which the protein is pulled from the interface back into bulk water. In each simulation the distance along $\hat{z}$ between the centres of mass of the protein and the oil phase is restrained by a harmonic spring to a target distance that moves at a constant velocity. The work profiles are then combined using the Jarzynski equality [77] (Eq. 3.3), as implemented according to Ref. [76], to give the potential of mean force (PMF) along the distance between center of mass of the protein and the center of mass of the oil phase. The harmonic spring constant is set to K=10^4 kJ/mol nm^2 and the pulling speed is set to v=0.1Å/ns. For each protein variant I performed 64 independent pulling simulations. The starting system configurations for the pulling simulations are selected every 1 ns from a MD simulation where the protein is adsorbed at the interface and the distance between protein and oil is restrained with the same harmonic spring used for pulling, but with the reference distance fixed at a certain value. The initial reference distance in the pulling simulations has to be the same as the value used to generate the ensemble of initial configurations. With the chosen settings, the stiff-spring approximation holds and the estimate of the free energy profile from the Jarzynski equality is consistent with its cumulant expansion [76]:

$$\begin{aligned} F(z) &= -k_B T \log\langle e^{-W(z)/k_B T}\rangle \qquad (3.3) \\ &\simeq \langle W(z)\rangle - \frac{1}{2k_B T}\left(\langle W(z)^2\rangle - \langle W(z)\rangle^2\right), \qquad (3.4) \end{aligned}$$

where $F(z)$ is the potential of mean force, $W(z)$ is the work performed by the moving harmonic spring as a function of the reference equilibrium distance of the spring z. The validity of the cumulant expansion is considered to be an indicator of a reliable free energy estimate [76]. The free energy of adsorption, ΔG_{ads}, was estimated as the difference between the minimum in the potential of mean force at the interface and its value in bulk.

Preferred location and orientation of BslA variants was subsequently studied using umbrella sampling simulations, resulting in the two-dimensional free energy landscape shown in Fig. 3.14. To improve the sampling of the adsorbed configurations at the interface, the distance between the protein and the oil phase was restrained by a harmonic spring of strength $K = 10^3$ kJ/mol nm^2 and for each BslA variant I used 12 windows with the reference distance spaced at regular intervals of Δz=1.67 Å (these settings allow a good overlap in the distribution of distances of neighbouring simulations). Each restrained simulation was 200 ns long, a time that significantly extends the maximum time scale of protein reorientation at the interface ($\sim$40 ns). In order to reconstruct the free energy surface along both the distance from the interface and the protein orientation $\cos\theta_r$, the two-dimensional distribution of distance and orientation was reweighted taking into account the external bias, and the different umbrella windows where combined together using a in-house implementation of the weighted histogram analysis method (WHAM) [74].

Initially, I wanted to use umbrella sampling also to compute the full free energy of adsorption (instead of the Jarzynski equality). As above, I restrained the distance between protein and oil using a spring constant of $K = 10^3$ kJ/mol nm^2. The

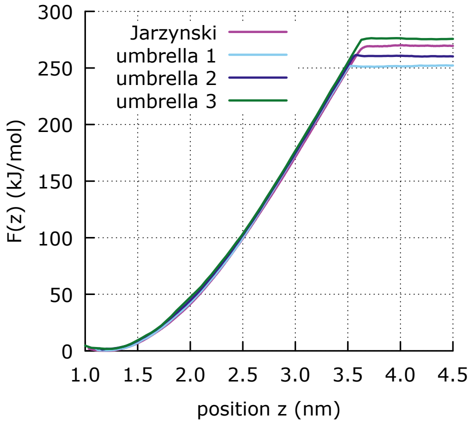


Figure 3.10 *Comparison between the free energy of adsorption of wt-BslA-L_{out} obtained from the Jarzysnki equality and from three umbrella samplings with different initial conditions. Since errors should be on the order of 1 kJ/mol, the profiles of the different umbrella samplings are not compatible and depend on the pulling simulation used to prepare the initial configurations. In the region that does not display hysteresis (below 3.5 nm), all profiles are consistent.*

initial configurations for the different umbrella windows (spaced every 1.25 Å, 100 ns each) were taken from steered MD simulations where the protein is pulled from its adsorbed state back into the water phase. The problem with this approach is that we found that the resulting potential of mean force displays hysteresis; specifically, the final profile depends on the precise location at which the protein desorbs from the interface in the initial pulling simulation (this dependence holds even for a very low pulling speed of 0.4 Å/ns). This is because around the region at which desorption occurs, the relaxation time of the system is very long, due to the fluctuations of the interface and the reorientation of the protein; therefore, the correspondent umbrella simulations do not have enough time to explore the correct equilibrium distribution within the considered simulation time. In Fig. 3.10 I compared the PMF of adsorption of wt-BslA-L_{out} obtained from the Jarzynski equality with those from three umbrella sampling profiles with different initial conditions. While the umbrella sampling results depend on the initial pulling simulation used to prepare the starting configurations, the Jarzynski equality offers a natural solution to average over multiple pulling simulations. Furthermore, the consistency of all the profiles in the region before desorption suggests that the Jarzysnki equality has been correctly applied.

The orientation of the protein is defined as the cosine of the angle θ_r between the normal to the interface and the direction of the eigenvector of the protein inertia tensor (principal moment of inertia) corresponding to the longest axis. The inertia tensor is defined as:

$$I_{\alpha\beta} = \sum_i m_i(\vec{r}_i \cdot \vec{r}_i \delta_{\alpha\beta} - r_{i\alpha} r_{i\beta}) \tag{3.5}$$

where the sum runs over the protein coarse-grained beads, m_i is the mass of bead i, $\vec{r}_i$ is its position, δ is the Kronecker's delta, and α and β correspond to the three spacial directions $\hat{x}$, $\hat{y}$ and $\hat{z}$. The inertia equivalent ellipsoid of a protein is an ellipsoidal particle with a uniformly distributed mass and with the same inertia tensor of the protein. The axes lengths of this ellipsoid can be computed from the eigenvalues of the inertia tensor using the relations [88]:

$$\begin{aligned} l_{\text{max}} &= 2\sqrt{\frac{2.5(I_{\text{mid}} + I_{\text{max}} - I_{\text{min}})}{M}} \\ l_{\text{mid}} &= 2\sqrt{\frac{2.5(I_{\text{min}} + I_{\text{max}} - I_{\text{mid}})}{M}} \\ l_{\text{min}} &= 2\sqrt{\frac{2.5(I_{\text{min}} + I_{\text{mid}} - I_{\text{max}})}{M}}, \end{aligned} \tag{3.6}$$

where $I_{\text{min,mid,max}}$ are the principal moments of inertia and M is the total mass of the protein. As discussed in the introduction, the BslA inertia equivalent ellipsoid has axes lengths: $l_{\text{BslA,max}} = 55$Å, $l_{\text{BslA,mid}} = 33$Å and $l_{\text{BslA,min}} = 26$Å.

I estimated the free energy of adsorption of the protein at the interface directly from equilibrium MD simulations as a sum of protein-solvent, E_{PS}, and interfacial, E_I, energy contributions. The protein-solvent energy was estimated as the sum of the water-to-cyclohexane free energies of transfer [61] of the residues that move from the aqueous phase to the oil phase, an approach similar to the one discussed in Ref. [28] for the calculation of the energy of a protein within a lipid membrane. The residues with buried side-chains, defined as those having a solvent accessible surface area lower than 30Å^2, were excluded from the sum. The residues that moved into the oil phase were identified from the position of the side-chain center of mass relative to the position of the interface, which was defined as the z-coordinate at which the density of water equals the density of oil. The interfacial energy was estimated from its expression in the continuum limit: $E_I = -\gamma A_I$, where γ=44 mN/m is the surface tension of the cyclohexane/water interface in the simulations and A_I is the area occupied by the protein at the interface. The area A_I was estimated as the area covered by all the coarse-grained beads (with radius 2.1 Å) that fall within 2.1 Å (the coarse-grained water radius) from the position of the interface. Fig. 3.15 reports averages over 400 ns equilibrium simulations, a time significantly longer than the reorientation time of the protein.

The hydrophobic dipole of a protein secondary structure element is defined as [28]:

$$\vec{\mu} = \sum_i H_i \hat{s}_i, \tag{3.7}$$

where the sum runs over the residues in the considered region (I used the hydrophobic cap), H_i is a measure of the hydrophobicity of residue i (I used the water-oil partitioning free energy [61]), while $\hat{s}_i$ is the unit vector pointing from the residue α carbon to the center of mass of its side chain. A large magnitude of the modulus of the dipole signals a highly anisotropic arrangement of residues within the structure, while the direction of the dipole corresponds to the average direction of the hydrophobic residues [28]. In order to account for the protein flexibility, I computed the dipole of each BslA conformation from averages over 400 ns trajectories in which the protein is adsorbed at the interface.

3.2.4 Results: coarse-grained molecular dynamics simulations

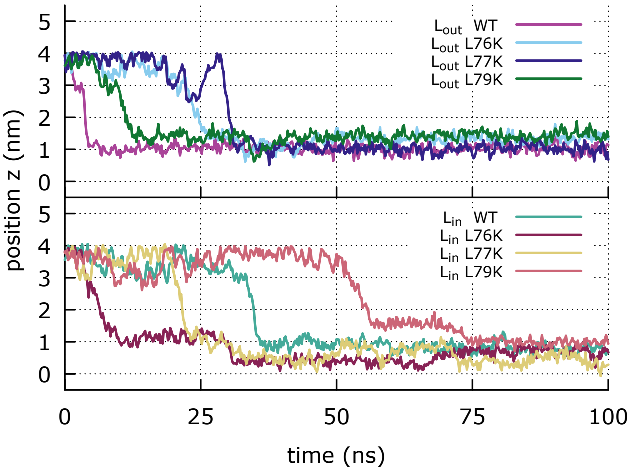


Figure 3.11 *Position of the BslA center of mass relative to the cyclohexane/water interface (at z=0) as a function of time during unbiased coarse-grained MD simulations. For all BslA variants considered, adsorption (signalled by the lack of diffusion along $\hat{z}$) occurs within ∼60 ns, and no desorption events are observed.*

Firstly, I considered the adsorption of the BslA variants at a cyclohexane/water interface during unbiased MD simulations. In all cases, the protein spontaneously adsorbed at the interface within ~60 ns, and no desorption events were observed within the considered time scale (500 ns), indicating that the free energy of adsorption is significantly higher that $k_B T$.

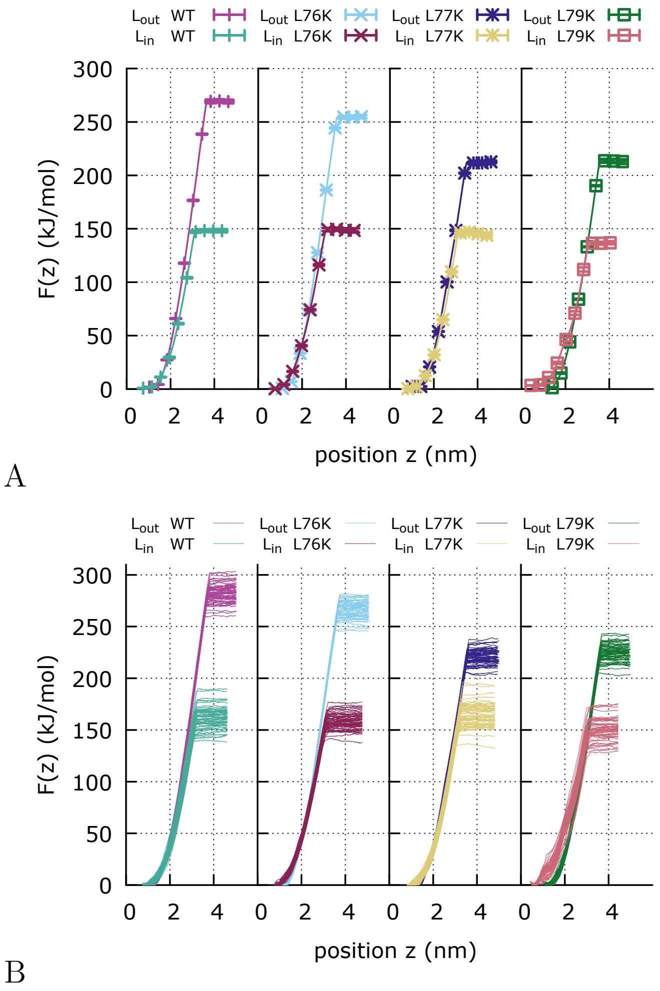


Figure 3.12 *(A) Potential of mean force for BslA variants as a function of the position of the protein center of mass relative to a cyclohexane/water interface. (B) Correspondent work profiles used to reconstruct the PMF in panel A via the Jarzynski equality. These profiles have been obtained from pulling simulations with the MARTINI coarse-grained model with the protein initially adsorbed at the interface.*

Before analysing the orientation of the protein at the interface, I wish to compute the free energy of adsorption of each variant, in order to rationalise the experimental observations by Hobley et al. [7] on the effect of the mutations (see Fig. 3.2) and to understand the nature of the conformational change between the L_{in} and L_{out} forms of the protein cap. Fig. 3.12A shows that in all cases the free energy of adsorption, ΔG_{ads}, is very high ($\gg k_BT$=2.5 kJ/mol). However, there is a dramatic change between the values for BslA-L_{in} and BslA-L_{out} (Fig. 3.12A). The L_{in} form of BslA has a ΔG_{ads} of ~148 kJ/mol that is essentially insensitive to the considered cap mutation (always within 12 kJ/mol from the wild type value). On the other hand, the adsorption free energy increases by a factor of almost two for wt-BslA-L_{out}, up to ΔG_{ads} ~270 kJ/mol, and cap mutations lower ΔG_{ads} down to ~254 kJ/mol for the L76K mutation, and to ~212 kJ/mol for the L77K and L79K mutations. Errors are on the order of 1 kJ/mol. The work profiles of the pulling simulations used to reconstruct the potentials of mean force via the Jarzynski equality are reported in Fig 3.12B.

Remarkably, the degree of perturbation of the free energy of adsorption due to mutations on BslA-L_{out} is in qualitative agreement with the experimental observations in Fig. 3.2: despite the similar position and type, the L76K mutation has a milder effect on both biofilm hydrophobicity and pendant drop wrinkle relaxation than the one caused by the L77K and L79K mutations (see below for discussion). The lack of sensitivity to mutations for BslA-L_{in} and its much lower affinity for the interface strongly suggest that BslA-L_{in} corresponds to the solution structure of the protein, and that a conformational change into the L_{out} form occurs upon adsorption at the interface. This cap conformational change is also supported by circular dichroism (CD) experiments that showed an increase in β secondary structure content upon adsorption [1].

The hydrophobicity of the BslA cap can be conveniently measured by the magnitude of its hydrophobic dipole [28]. The ΔG_{ads} values are consistent with the expectation that higher cap hydrophobicity corresponds to higher adsorption free energy (Fig. 3.13). The L_{out} cap is the most hydrophobic one, and the burial of three leucines in the L_{in} form causes a larger decrease in hydrophobic dipole than the one obtained from mutating a single hydrophobic residue into a hydrophilic one. It is interesting to notice that the minor change in adsorption free energy (and experimental behaviour) caused by the L76K mutation can be rationalised by the observation that the change in cap hydrophobic dipole is also very small. The lack of change in ΔG_{ads} from wt-BslA-L_{in} to L77K-BslA-L_{in} can be trivially explained by the fact that the L77 residue is buried in the L_{in} conformation. However, the analysis of the hydrophobic dipole still does not explain why the L79K mutation has a large effect on the free energy of adsorption for BslA-L_{out} but not for BslA-L_{in}. In order to understand this anomalous behaviour, I come back to the discussion on the orientation of the

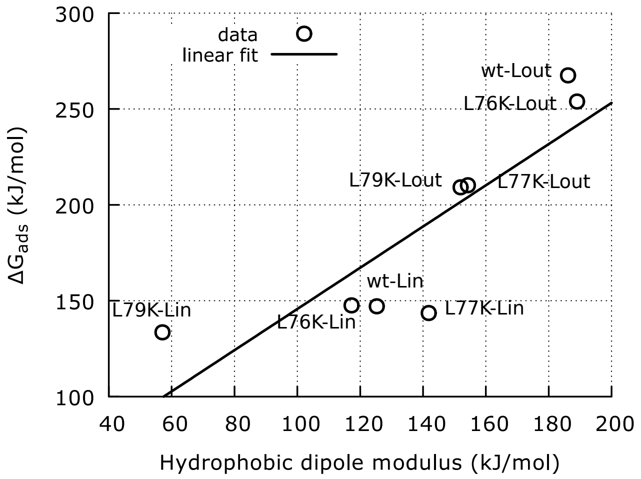


Figure 3.13 *The free energy of adsorption computed from pulling simulations is plotted against the magnitude of the hydrophobic dipole of the cap region of the protein. The linear fit shows that a decrease in hydrophobic dipole corresponds to a decrease in free energy of adsorption.*

protein at the interface.

In Fig. 3.14 I show the free energy landscape of four BslA variants adsorbed at the interface as a function of the position relative to the interface and the orientation of the protein. We notice that only BslA-L_{out} (both wild type and L79K) adopts a clear end-on configuration, with the hydrophobic cap partitioned into the oil phase and the polar domain into the aqueous phase. On the other hand, wt-BslA-L_{in} explores both end-on and side-on orientations, and its L79K mutation only adopts the side-on orientation. Hence, these coarse-grained simulations confirm the proposed orientational transition under a change in cap hydrophobicity. Furthermore, the observed orientations offer an understanding for why the L79K mutation has a milder effect on BslA-L_{in} than on BslA-L_{out}: when the protein is in its L_{in} conformation, the cap is only partially exposed to the oil phase, therefore the precise chemistry of this region should have a milder effect on the free energy of adsorption.

Finally, I wanted to test whether the continuum model used for the adsorption of a Janus colloid can explain the energetics and the tilting transition observed in the MD simulations. Fig. 3.15 shows the estimates of the protein-solvent (E_{PS}) and interfacial (E_I) energies of the BslA variants at the interface, as obtained directly from equilibrium MD simulations (see methods section). The sum of these two energy contributions is in semi-quantitative agreement with the values obtained

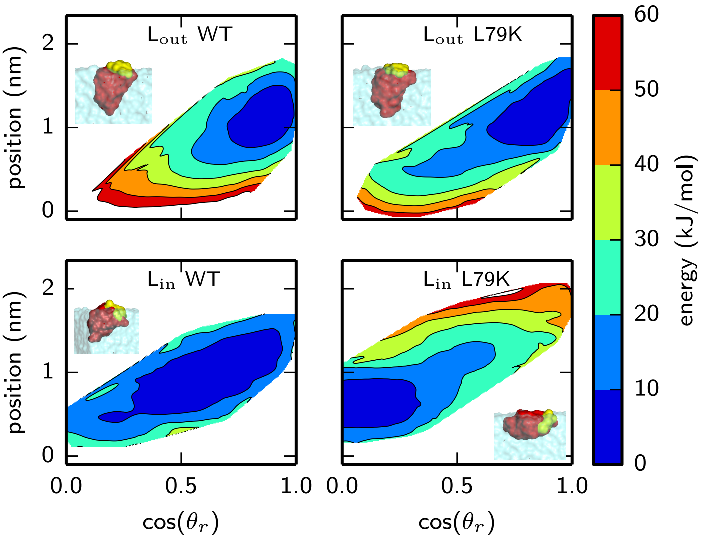


Figure 3.14 *Free energy landscape of four BslA variants (from left to right and top to bottom: wt-BslA-L_{out}, L79K-BslA-L_{out}, wt-BslA-L_{in} and L79K-BslA-L_{in}) adsorbed at the interface as a function of the position relative to the interface and the orientation of the protein, defined as the cosine of the angle between the major protein inertia axis and the normal to the interface. The inset shows the configuration corresponding the free energy minimum on the surface.*

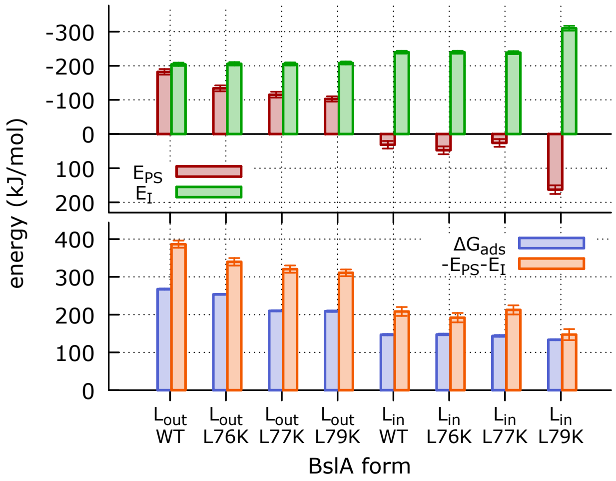


Figure 3.15 *Top: estimates of the protein-solvent (E_{PS}, red) and interfacial (E_I, green) energies of the BslA variants adsorbed at a cyclohexane/water interface, as obtained from equilibrium MD simulations. Bottom: comparison between the estimate of the free energy of adsorption obtained from non-equilibrium pulling simulations (ΔG_{ads}, indigo) and from the sum of protein-solvent and interfacial contributions ($=-E_{PS} - E_I$, orange).*

from the non-equilibrium pulling simulations, which is a more accurate but also more expensive computational method. This analysis confirms that the end-on orientation of the protein is stabilised by protein-solvent interactions, whereas the side-on orientation is stabilised by the interfacial energy. As the hydrophobicity of the cap is reduced, the interaction between the cap and the oil phase becomes less and less favourable, and the protein eventually tilts its main axis in order to occupy a larger area at the interface, despite the energy cost for generating an unfavourable interaction between the polar domain of the protein and the oil phase, which is highlighted by the positive E_{PS} energy values for BslA-L_{in}.

3.2.5 Conclusions of the coarse-grained study

To summarise, the model of BslA based on a Janus ellipsoidal colloid offers a simple understanding of the tilting transition of the protein at the interface. I showed that the hydrophobicity of the BslA cap directly controls the orientation of the protein with respect to the interface. Low cap hydrophobicity corresponds to a tilted, side-on orientation, and high hydrophobicity corresponds to an upright, end-on orientation (Fig. 3.14). This behaviour has been firstly suggested from Monte-Carlo simulations of a BslA-like Janus ellipsoidal colloid and then it has been validated via higher-resolution coarse-grained MD simulations. This tilting transition is due to the unique geometrical features of BslA, and to the possibility to tune the hydrophobicity of the cap via conformational changes and/or residue mutations. The transition is ultimately driven by the competition between protein-solvent and interfacial interactions: the formers are optimised by the end-on orientation, whereas the latter by the side-on orientation. Remarkably, the energetics of BslA adsorpbtion can be fully captured by separating these two energy contributions, giving a semi-quantitative agreement with accurate steered MD simulations.

In the next section, I will show that the orientation of the protein at the interface is important for protein-protein recognition, and that only when the protein adopts the end-on orientation corresponding to wt-BslA-$\mathrm{L}_{\mathrm{out}}$, it is able to self-assemble into a two-dimensional crystal. Therefore, the results of these coarse-grained simulations are directly relevant for the understanding of BslA self-assembly, and they offer a simple explanation for the decrease in the order of the protein film caused by the L77K mutation [1]. Even though I cannot speculate about the precise cap conformation adopted by L77K-BslA when the protein is adsorbed at the interface (it may be L_{in}, L_{out}, or some intermediate conformation), this mutation will certainly cause a decrease in the hydrophobicity of the cap region, which in turn will induce a higher tilting of the protein compared to the end-on orientation adopted by wt-BslA-L_{out}. This tilting will then prevent the proteins from coming into contact with the correct orientation necessary for binding. The group of Dr. Simon Titmuss at the University of Edinburgh performed x-ray reflectivity experiments (at the Diamond facility in Oxfordshire) on both wild type and L77K-BslA adsorbed at a air/water interface. Preliminary analysis of the reflectivity profiles showed that the protein film formed by wt-BslA has a thickness of ∼36 Å, whereas the mutation lowers this value down to 27 Å (internal communication), suggesting that L77K-BslA is indeed tilted at the interface.

Finally, in the future it may be possible to dynamically tune the hydrophobicity of the BslA cap by changing the environmental conditions (inducing a change in conformation or surface chemistry), providing a route to control the orientation of the protein at the interface and its interfacial self-assembly, which may be useful for nano-technological applications.

3.3 Atomistic mechanism of BslA conformational changes and self-assembly at interfaces

3.3.1 Introduction

Coarse-grained simulations of BslA adsorption at interfaces have provided many useful insights into the adsorption of the protein. In particular, the results suggest that the hydrophobic cap of BslA undergoes a conformational change upon adsorption, going from a conformation in solution that shields many hydrophobic residues from the aqueous phase, to a conformation that exposes all hydrophobic residues towards the apolar phase. However, this conclusion is mainly supported by the different affinity of the two considered conformations for the interface, and there is no direct evidence that the L_{in} conformation observed in the BslA crystal decamer corresponds exactly to the solution structure of the protein, and that the L_{out} conformation corresponds to the interfacial structure of the protein.

An even more important problem that was not addressed by the coarse-grained simulations regards the self-assembly of the protein at the interface. BslA forms a strong elastic film at the interface that is likely to be stabilised by protein-protein interactions. In particular, transmission electron microscopy (TEM) images obtained from drying the protein film onto a surface revealed that the film is highly ordered as a rectangular lattice with lattice spacings of 43 and 39 Å [1] (Fig. 3.4A). Unfortunately, the resolution of these images is not high enough to reveal the molecular details of the crystal and the precise protein-protein interactions that stabilise the film at the interface.

The aim of this work is to go beyond the results on the BslA adsorption mechanism obtained from coarse-grained simulations and TEM images by providing a picture of the adsorption with atomistic details. To this aim, I performed all-atom MD simulations of BslA in water, at an air/water interface, and simulations of BslA dimers. In section 3.3.3 I will discuss the ensemble of protein conformations in solution, as obtained from all-atom simulations using replica-averaged NMR restraints. These restraints enable to obtain an ensemble consistent with the experimental NMR chemical shifts of the protein atoms. The results show that the $\mathrm{L_{in}}$ conformation found in the BslA crystal decamer only corresponds to a fraction of a larger ensemble of structurally diverse conformations. It is interesting to notice that the $\mathrm{L_{out}}$ conformation is never observed in this ensemble. In section 3.3.4 I confirm the results obtained from the coarse-grained simulations showing that the $\mathrm{L_{out}}$ conformation of the protein has a higher affinity for the interface and a lower orientational freedom compared to BslA-$\mathrm{L_{in}}$. In section 3.3.5 I present replica exchange simulations of the protein adsorbed at an air/water interface which strongly suggests that the $\mathrm{L_{out}}$ conformation is the most favourable structure of the adsorbed protein. Finally, in section 3.3.6 I propose a model for the two-dimensional crystal of the protein at the interface that is consistent with the TEM images obtained by Bromley et al. [1] and I analyse the protein-protein interactions that are involved in the self-assembly by performing steered MD simulations of wild type and mutant BslA dimers.

3.3.2 Methods: all-atom simulations

All-atom MD simulations were performed using the program GROMACS 4 [94]. Unless otherwise stated, the protein was represented according to the AMBER ff99SB*-ILDN force field [50, 56, 57], which was shown to be amongst the most accurate for the simulation of protein folding [101] and was parametrised to reproduce the correct balance between the different secondary structure elements [57]. I believe that this should be of particular importance in the case of BslA, where there should be a transition between a disordered cap with a small helical content in solution ($\mathrm{L_{in}}$ conformation) to an ordered cap rich in β-sheet content at the interface ($\mathrm{L_{out}}$ conformation). For the water molecules I employed the TIP3P model [102]. All bonds were constrained to their equilibrium value using the LINCS algorithm [103], whereas the geometry of the water molecules was constrained using the SETTLE algorithm [104]. Van der Waals and short-range Coulomb interactions were both cut-off at 10 Å, whereas long-range electrostatics was treated using the particle mesh Ewald method [105] with a Fourier spacing of 1.2 Å. The equations of motion were integrated using the leap-frog algorithm [64] with a time-step of 2 fs. In all cases, the production simulation runs were performed in the NVT ensemble, using the Nose-Hoover thermostat [67, 106] with a reference temperature of 300 K and a relaxation time of 1 ps; solvent and solute were thermostated independently. The size of the simulation box is set so that the protein is always at least 12 Å away from the closest image of the box. Production runs are always preceded by an equilibration protocol consisting of an energy minimisation with the steepest descent algorithm and by a 200 ps run in the NVT ensemble with the protein's heavy atoms restrained to their initial positions by a harmonic potential with spring constant equal to 1000 $\mathrm{kJ/mol/nm^2}$. The net charge of the protein was neutralised via the addition of a minimum amount of either Na or Cl ions (depending on the system).

I performed four main sets of simulations (see below for more details):

- BslA in bulk water, using NMR chemical shifts as replica-averaged structural restraints [107];
- spontaneous adsorption of different BslA conformations at an air/water interface;
- replica exchange simulations of BslA adsorbed at an air/water interface;
- equilibrium and non-equilibrium pulling simulations of BslA dimers adsorbed at the interface.

Throughout the section, I will go beyond the binary $\mathrm{L_{in}}$-$\mathrm{L_{out}}$ classification of BslA conformations, and I will distinguish between each of the chains found in the BslA crystal structure and the conformations that are specifically observed in the simulations. The first reason for this is that even though the 8 chains from A to H found in the BslA crystal structure are all part of the L_{out} cluster, there is a small difference in the orientation of residue F159, which is hidden inside the protein core in chains A, B, E and F, and exposed to the solvent in chains C, D, G and H. This difference does not affect the hydrophobicity of the cap region nor the way the protein interacts with the interface; however, it may be relevant for the conformational change of the cap region upon adsorption and/or for protein-protein interactions at the interface. I will refer to these conformations as $L_{out,x}$, where x refers to the index of the specific chain in the BslA crystal structure (and similarly for the two L_{in} conformations from the crystal, corresponding to chains I and J). The second reason is that in the simulations in solution with NMR restraints I find conformations that can be broadly classified L_{in}, since most of the leucine residues in the cap are buried, but still displaying a high degree of conformational variability amongst them: I will refer to these conformations as $L_{in,cl\ j}$, where j refers to the index of the specific cluster to which the conformation belongs.

All-atom methods: BslA ensemble in solution

In order to understand the structure of the protein in solution, I performed all-atom MD simulations with replica-averaged structural restraints [107]. These restraints introduce a correction to the force field that forces the ensemble of explored conformations to be consistent with a target set of experimental observables. In this method, a number of replicas of the same system are run in parallel, and each of the replicas is subject to an additional energy term that restraints the averages (over the replicas) of the observables to their experimental values. In the limit of a high number of replicas, this method introduces the least biased perturbation to the force field that leads to an equilibrium probability distribution consistent with the experimental observations, i.e. in agreement with the maximum entropy principle [108, 109].

The use of experimental restraints can be particularly helpful for the study of disordered regions [55, 110–112], which are poorly represented by the current protein force-fields (including AMBER ff99SB*-ILDN) [54, 113]. This is particularly appropriate for BslA, since there is large evidence that the cap of the protein is unstructured in solution (the disorder is further confirmed a-posteriori by the MD simulations). For the simulation restraints I used the NMR (nuclear magnetic resonance) chemical shifts of the protein backbone atoms, which have been determined by Dr. Ryan Morris and Dr. Marieke Schor at the University of Edinburgh. Unfortunately, because of the poor quality of the NOESY spectra from the NMR, it was not possible to resolve the atomistic structure of the protein from the data. New NMR experiments of BslA in solution have been recently performed, and once the processing of the data will be completed, it will be possible to compare the ensemble of BslA conformations from the MD simulations with the experimental NMR ensemble.

The energy term added to the force field takes the following form [107]

$$E_{CS} = K_{CS} \sum_{i=1}^{N} \sum_{j=1}^{6} (\delta_{ij}^{\text{calc}} - \delta_{ij}^{\text{exp}})^2 \tag{3.8}$$

where the first sum runs over the protein residues and the second one over the six backbone atoms H_α, H_N, N, C_α, C_β and C'. δ^{exp} are the experimental chemical shifts and δ^{calc} are the averages of the chemical shifts over the replicas, which are calculated from the protein coordinates using the CamShift method [114]. I employed four replicas, which is regarded as the best compromise between computational efficiency and agreement with the limit of a large number of replicas [115]. In order to generate an ensemble of conformations, I used the same simulated annealing protocol described in Ref. [115]: 100 ps of run at T=300 K, 100 ps with the temperature linearly increasing up to T=400 K, 100 ps at T=400 K, 300 ps with the temperature linearly decreasing back to T=300 K. At T=300 K, the energy constant of the restraints is set to K_{CS}=24 kJ/mol, but it is set to zero in the high temperature phase of the annealing (K_{CS} is linearly increased/decreased during the cooling down/heating up). For the analysis of the protein ensemble, I only considered the configurations explored at T=300 K. This annealing protocol is necessary to allow a fast exploration of the protein conformational space; even though I was still unable to generate a converged ensemble of structures within the available computational time (Fig. 3.17B). I performed two independent sets of MD simulations: one starting from the $L_{\text{in,J}}$ conformation of BslA_{44-170} and another starting from the $L_{\text{out,A}}$ conformation. The initial atomic positions of the six residues in the cap of BslA that are not resolved in chain J were obtained from the crystal structure of the protein YweA (recently crystallised by the group of Nicola Stanley-Wall at the University of Dundee, see Fig. 3.22), which displays the same fold observed in BslA-L_{in}. The protein was solvated with 11000 water molecules in a dodecahedral box with the sides set to the average values obtained from a 100 ns long simulation in the NPT ensemble at the ambient pressure of 1 Bar. In both cases, each of the four replicas was run for 240 ns, for a total of 200 annealing cycles per replica (1920 ns of total simulation time and 1600 cycles). These simulations were run using a version of GROMACS 4 [94] patched with the software PLUMED 2 [116] and ALMOST [117]. The combined use of these three programs is implemented following the standard instructions given on the PLUMED website (www.plumed.org): firstly, I compiled ALMOST, then I compiled PLUMED with the ALMOST option, I patched the GROMACS code using PLUMED, and finally I compiled GROMACS, which is used to perform the actual MD simulations, and calls the other two programs every time step to compute the chemical shifts and apply the replica-averaged restrains.

For the clustering of the BslA ensemble, I considered the structures from both simulations (with the different starting conformations) observed at T=300 K after discarding the first 60 ns as equilibration. The clustering was based on the cap backbone root mean square deviation between the structures, using the gromos clustering algorithm [99] implemented in the GROMACS tools [94] with a cut-off distance of 1.5 Å.

All-atom methods: spontaneous adsorption at an air/water interface

I performed simulations of the spontaneous adsorption of different BslA conformations at an air/water interface. In these simulations I first placed the protein at the center of a cubic water box with sides equal to 79 Å, I increased the length of the box along the $\hat{z}$ direction by 30 Å to create the air/water interfaces, and then I let the protein diffuse until it spontaneously adsorbs at one of the two interfaces. I looked at the adsorption of four different protein conformations: one taken from the biggest cluster obtained in the bulk water simulations with the replica-averaged restraints ($\mathrm{L}_{\mathrm{in,cl1}}$ conformation), one from the fourth biggest cluster of the same ensemble ($\mathrm{L}_{\mathrm{in,cl4}}$), one taken from chain A in the BslA crystal structure ($\mathrm{L}_{\mathrm{out,A}}$) and one from chain C ($\mathrm{L}_{\mathrm{out,C}}$). In the first two conformations the cap of the protein is mostly disordered with a low hydrophobicity whereas in the other two the cap is rich in β-sheet content and all the hydrophobic residues are exposed to the solvent. I considered the adsorption of the 4th bulk cluster because in this conformation residue L77 is exposed to the solvent, contrary to what observed in the L_{in} conformations found in the crystal structure (chains H and J) and in the first three clusters from the simulation ensemble. The main difference between the $L_{\mathrm{out,A}}$ and $L_{\mathrm{out,C}}$ conformations is the orientation of residue F159, which is buried inside the cap in chain A and exposed to the solvent in chain C. For each of these conformations I performed four independent simulations with different random starting orientations, for a total of 16 runs. I stopped each of the simulations roughly 50 ns after adsorption occurs, so that I could also study the equilibrium orientation of the protein with respect to the interface as a function of its conformation.

The orientation of a rigid body (a good approximation of a protein for a fixed conformation) with respect to an interface is uniquely defined by two angles, corresponding to the spherical coordinates θ and φ of the normal to the interface in a Cartesian reference frame defined by the rigid body. I defined the protein reference frame using the coordinates of the C_α atoms of the three residues K59, F51 and D166: the $\hat{x}$ axis as the normal vector going from residue K59 to residue F51, the $\hat{y}$ axis as the direction perpendicular to both $\hat{x}$ and the vector going from K59 to D166, and the $\hat{z}$ axis defined as $\hat{x} \times \hat{y}$. The same analysis of the protein orientation is also used for the study of BslA dimers adsorbed at the interface (section 3.3.6).

All-atom methods: replica exchange simulations of BslA at an air/water interface

In order to find the equilibrium structure of BslA when the protein is adsorbed at the interface, I run replica exchange MD simulations with solute tempering [118], a variation of replica exchange where the Hamiltonian of the each replica is modified to control the effective temperature of only a small part of the system: the cap of the protein (residue ids 75-80,119-124,151-159). This technique enhances the conformational changes in the target region while reducing the computational cost compared to standard replica exchange simulations, where the number of system replicas can become prohibitively high for large systems like ours. Secondly, since there should not be relevant conformational changes outside the cap of the protein, I can selectively heat up this specific region while keeping the rest of the system at ambient temperature. In these simulations, I used twelve replicas with effective cap temperatures exponentially distributed in the range T=300-900 K and I attempted to exchange neighbouring replicas every 2 ps, achieving an acceptance rate of exchange of 0.15. I performed two simulation runs: one starting with BslA adsorbed at the air/water interface in its $L_{\mathrm{out,C}}$ conformation, and one starting from the system configuration observed after 100 ns of replica exchange at the highest effective temperature of 900 K, where the cap of the protein is unfolded. The simulations were run on a special version of GROMACS 4 [94] which has been modified to perform replica exchange simulations with solute tempering [119]. The protein force field was parametrised according to AMBER ff99SB-ILDN [50] and the results should only be considered preliminary; it would be good to to perform new replica exchange simulations with AMBER ff99SB*-ILDN [50, 56, 57], after developing a strategy to address the problem of convergence that will be discussed in the results section.

All-atom methods: simulations of BslA dimers

I studied the self-assembly of the protein at the interface by means of equilibrium and steered MD simulations of BslA dimers adsorbed at an air/water interface. I considered two BslA dimers: one corresponding to chains C and H in the BslA crystal structure (1st dimer), and another one generated from chains A and B in the crystal structure of the protein YweA (2nd dimer). The initial atomic positions of the BslA dimer based on the YweA crystal structure have been obtained by aligning BslA-$L_{\mathrm{out,C}}$ to the two conformations of YweA, which, apart from the differences in the cap region, displays the same fold (Fig. 3.22). For each BslA dimer, I compared the behaviour of the wild type protein with a set of mutants expected to affect the strength of the binding: for the 1st dimer I considered the double mutants F51A-D166K and F51I-D166K, and for the 2nd dimer I considered the double mutant D66K-N101D. Alignments and mutations were performed using the software PyMOL [100].

In the 1st BslA dimer, there is an interaction mediated by the disordered N-terminal tail. The precise conformation of these tails is expected to affect the binding energy of the dimer, but because of the high computational cost associated with their conformational sampling, I decided to truncate this region, and consider only the protein sequence from residue id 48 to residue id 170 (instead of 44 to 170, as in the other studies). This truncation may slightly lower the binding energy of the 1st dimer; however, this should not affect the conclusions of the work, since I am mainly interested in comparing the protein-protein binding affinities of wild type BslA with the mutants.

For the unbiased simulations, the two proteins are placed in a rectangular box with sides 88x85x105 Å and 94x94x105 Å for the 1^{st} and the 2^{nd} dimer respectively, and they were solvated with water (which occupies 65% of the volume) so that the two proteins are already adsorbed at the air/water interface with the same equilibrium orientation observed in the spontaneous adsorption of the BslA monomer. The system is then run in the NVT ensemble for 100 ns, unless a complete dissociation of the two monomers occurs.

For the computation of the binding free energy for each of the two binding interfaces and the different BslA variants, I performed a set of steered MD simulations. The potential of mean force (PMF) along the separation between the two monomers is then reconstructed from the work profiles using the Jarzynski equality [77], following the same protocol used to compute the free energy of adsorption of BslA [76] (in section 3.2). The initial dimer configurations were taken from the equilibrium MD runs, which were then placed in a simulation box with sides 90x77.6x105 Å where the dimer is already adsorbed at the interface and the distance vector between the two monomers is aligned to the $\hat{x}$ axis. For the pulling protocol, one of the two monomers acts as a reference, and its backbone C_α atoms are restrained to their initial positions by a harmonic potential with a spring constant of 1000 kJ/mol/nm^2, whereas the other monomer is pulled along the x component of the distance between the two proteins centres of mass at a speed of 0.2 Å/ns, using a harmonic potential with a spring constant of 10000 kJ/mol/nm^2. The high spring constant was chosen to satisfy the stiff spring approximation necessary for the method [76], whereas the low pulling speed was chosen after some preliminary tests to check that a further decrease in the speed does not cause a significant change in the work profiles. The starting system configurations for the pulling simulations are selected every 5 ns from a MD simulation with the same settings as above, but with the pulling speed set to zero. For the wild type protein I produced 32 independent work profiles for each dimer, whereas for the mutants, which have a lower binding affinity, I produced only 16 profiles. The potentials of mean force reported in Fig 3.25 correspond to the cumulant expansion of the Jarzynski equality [76] (Eq. 3.4), since this is regarded as being more robust than the Jarzynski equality itself, which is likely to suffer from a low sampling of the tails of the distribution of the work profiles [76]. The errors in the potentials of mean force have been calculated using the Jackknife algorithm [120].

The second dimer configuration is stabilised by a pair of backbone hydrogen bonds formed between the two protein loops. When one protein is pulled away from the other one, these bonds cause a large deformation of the protein structure, which in turn generates large dissipative work. In order to avoid the generation of these dissipative works that compromise the correct evaluation of the potential of mean force, I designed a set of harmonic walls that act to reduce the deformation and the orientational freedom of the protein during the pulling protocol: one wall keeps the distance between the center of mass of the protein and the loop involved in the backbone hydrogen bonds within its equilibrium basin, whereas another wall prevents the tilting of the pulled protein with respect to the interface by acting on the z component of the distance between the center of mass of the protein and its cap region. These walls do not affect the equilibrium distribution of the bound dimer nor the one of the unbound monomers, and therefore they will not modify the total free energy of binding.

3.3.3 Results: solution ensemble of BslA

The MD simulations with replica-averaged restraints from NMR chemical shifts show that the hydrophobic cap is highly disordered in solution (Fig. 3.16), and that the L_{in} conformation found in chains I and J of the crystal structure corresponds to the largest cluster of this disordered ensemble (Fig. 3.17). Firstly, we notice from Fig. 3.16 that, even when the simulation starts from the ordered $L_{\mathrm{out,A}}$ form of the cap, after a short equilibration of $\sim$60 ns, the proteins never explore conformations with a cap backbone rmsd lower than $\sim$3 Å from the $L_{\mathrm{out,A}}$ structure. On the other hand, in both simulations the protein reaches a backbone rmsd of $\sim$1 Å from the unstructured $L_{\mathrm{in,J}}$ form of the cap. However, these simulations also point to a very high structural variability of the cap region in solution. The rmsd plotted in Fig. 3.16 shows that the protein can also explore conformations that are rather far from the form of the cap found in the crystal structure, with always at least one of the four replicas in each simulation having a rmsd from $L_{\mathrm{in,J}}$ around 3 Å.

Clustering analysis (Fig. 3.17) of the simulations shows that the cap conformation found in the crystal structure corresponds to the most populated cluster, which covers 9.1% of the observed structures and whose cluster centre has a rmsd of 1.4 Å from the $L_{\mathrm{in,J}}$ form of the cap. The four biggest clusters, shown in in Fig. 3.17A, can well represent the variety of conformations explored by the protein in solution. All these clusters display the tendency to bury some of the cap

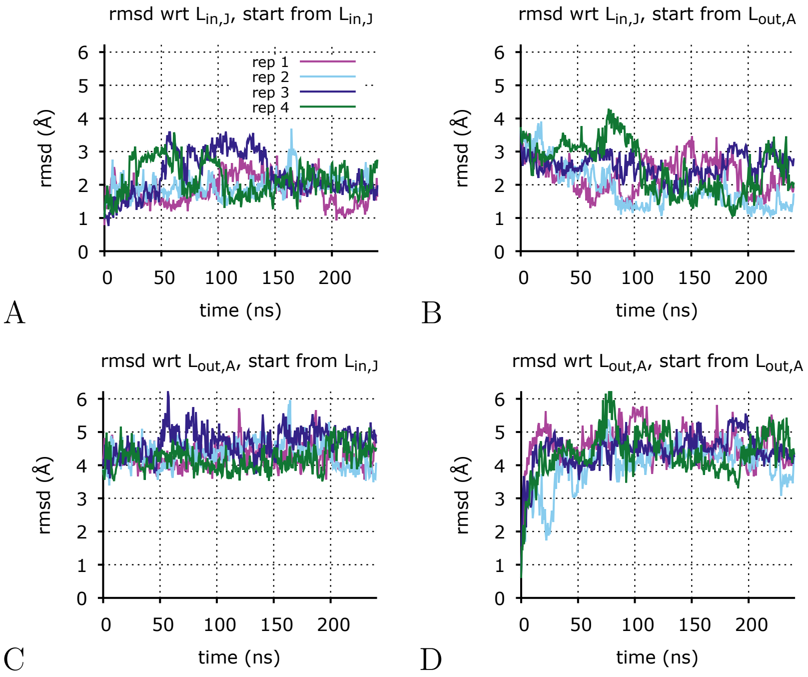


Figure 3.16 *All-atom simulations of BslA in solution using replica-averaged structural restrains from NMR chemical shifts. For each of the four replicas employed, I plot the root mean square deviation of the cap backbone with respect to the $L_{in,J}$ (A,B) and $L_{out,A}$ (C,D) conformations found in the crystal structures of the protein. Panels A and C refer to the simulation starting from the $L_{in,J}$ conformation, whereas panels B and D refer to the one starting from $L_{out,A}$.*

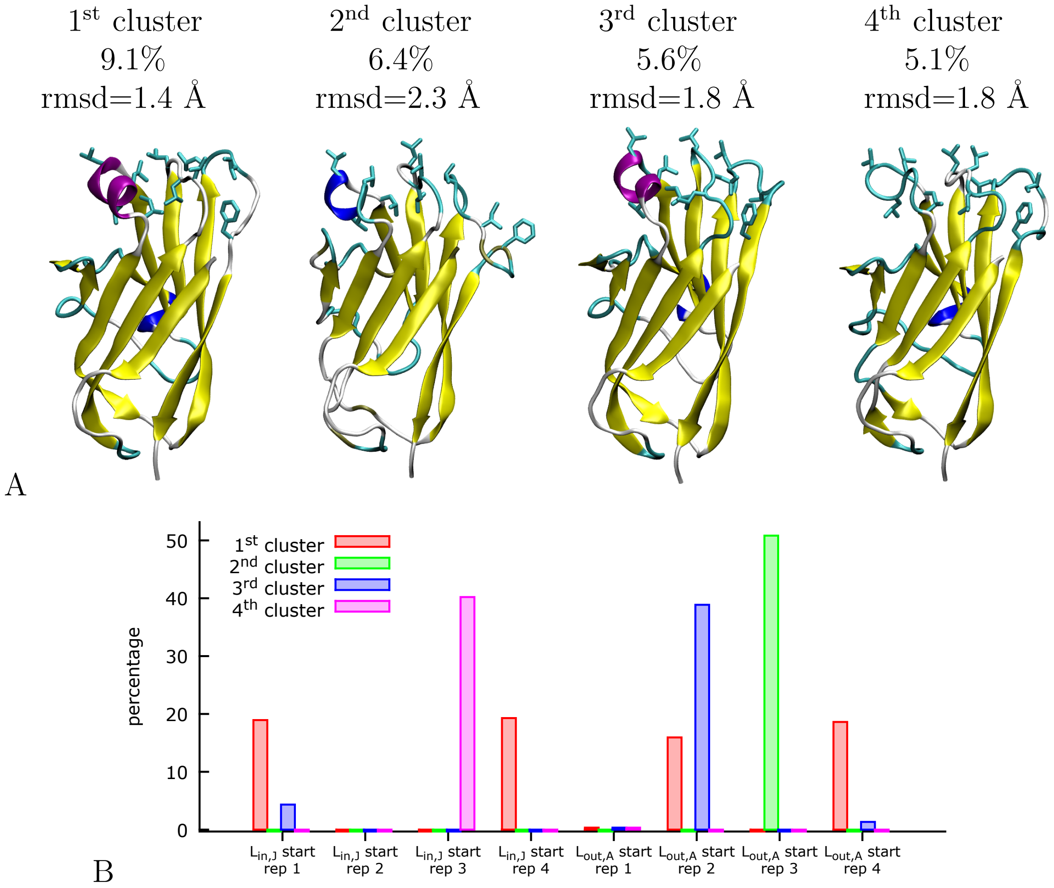


Figure 3.17 *Clustering of the structures explored in the simulations of BslA in solution with replica-averaged structural restrains. (A) Structures of the cluster centres of the four main clusters; for each of them I report the population of the cluster (as a percentage of the explored structures) and the cap backbone root mean square deviation from the* $L_{in,J}$ *conformation found in the BslA crystal structure. (B) For each of the system replicas (1 to 4) in each simulation run (starting from* $L_{in,J}$ *or from* $L_{out,A}$*) I plot the percentage of time spent in each of the four main clusters shown above.*

hydrophobic residues from the solvent (in particular, residues L119 and L123 are buried in all the structures observed in the solution ensemble). For this reason, these structures can all be broadly assigned to the L_{in} form of the cap. On the other hand, the orientation of residue L77, shows the highest degree of variability across the clusters, going from being completely buried into the protein core in the 1st and 2nd clusters, to completely exposed to the solvent in the 4th cluster. The latter conformation ($L_{\text{in,cl4}}$) may represent an intermediate step between the L_{in} and L_{out} forms of BslA found in the crystal structure, and it may play an important role in lowering the barrier to the adsorption at the interface and/or favouring the conformational change into the ordered cap upon adsorption. In the next section I will test the importance of this structure for the spontaneous adsorption of the protein at an air/water interface.

Unfortunately, the clustering analysis also points to the lack of convergence in the simulations: a necessary condition for the convergence of the ensemble is that each of the replicas explore the same clusters with the same probability. However, Fig. 3.17B shows that none of the reported clusters is equally distributed across the replicas. In fact, the 2nd cluster is only observed in a single run starting from $L_{\text{out,A}}$, and the 4th one is only observed in a single run starting from $L_{\text{in,J}}$. However, the fact that the first cluster is observed in both simulations and in multiple system replicas suggests that the L_{in} conformation found in the BslA crystal structure really is the main conformation adopted by the protein in solution. I believe that it may be possible to reach convergence using some type of enhancing technique like metadynamics [75], which has already been successfully combined with replica-averaged restraints based on NMR chemical shifts [55].

3.3.4 Results: spontaneous adsorption at air/water interfaces

In order to study the dependence of the spontaneous adsorption on the specific cap conformation of BslA, I considered the following four BslA structures: $L_{\text{in,cl1}}$, $L_{\text{in,cl4}}$, $L_{\text{out,A}}$ and $L_{\text{out,C}}$. The first two have been obtained from the simulations in bulk with replica-averaged restraints, whereas the last two are taken directly from the BslA crystal structure. In Fig. 3.18 I plot the distance of the protein center of mass from the interface as a function of time. These simulations show that the L_{in} conformation takes longer to permanently adsorb to at the interface ($\langle t_{\text{ads,in}} \rangle$=71±10 ns) compared to the L_{out} conformation ($\langle t_{\text{ads,out}} \rangle$=35±13 ns), and in one run starting from $L_{\text{in,cl4}}$ no permanent adsorption is observed within

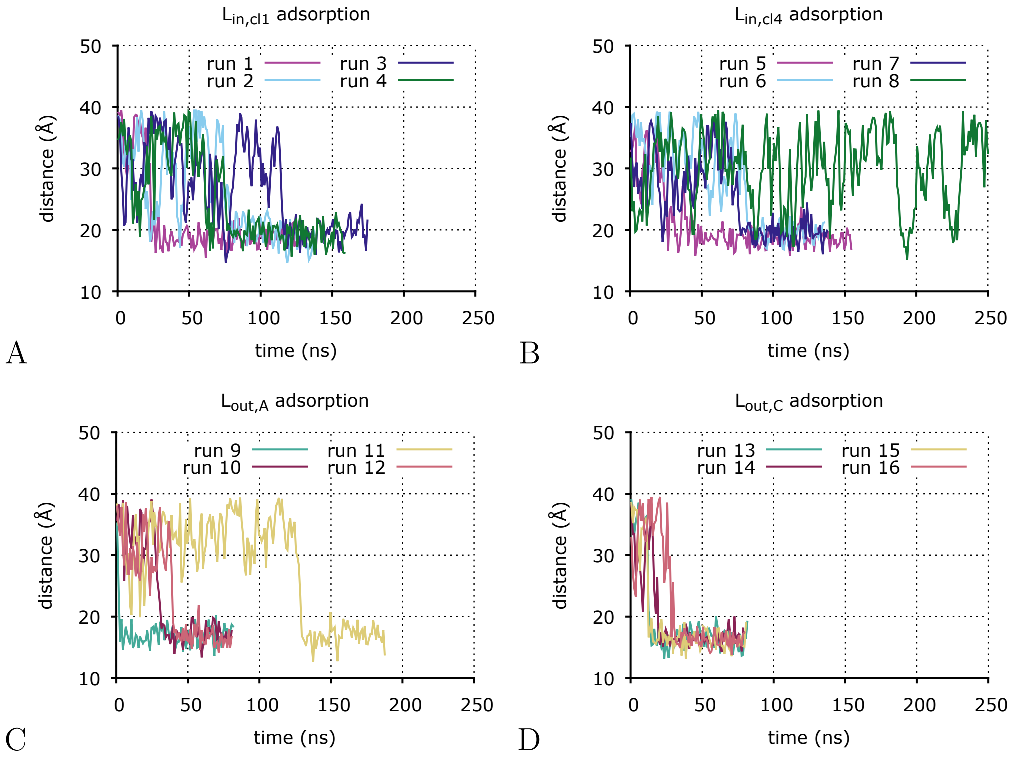


Figure 3.18 *Spontaneous adsorption of BslA at an air/water interface from all-atom simulations. I plot the distance of the protein center of mass from the interface as a function of time for different cap conformations: (A) $L_{in,cl1}$, (B) $L_{in,cl4}$, (C) $L_{out,A}$ and (D) $L_{out,C}$. For each conformation I performed four independent simulation runs with different initial random orientations of the protein with respect to the interface. The decrease in diffusion signals the adsorption of the protein, which is observed in all cases within the considered time window apart from run 8 in panel B.*

the considered time window of 250 ns (this run is not included in the average). I refer to a permanent adsorption when no desorption events are observed within 50 ns from the adsorption. Furthermore, only for the runs starting from L_{in} I observe desorption events, whereas for the L_{out} conformation the first adsorption event always corresponds to a permanent adsorption. Considering that BslA adopts the disordered cap conformation in solution, these results are in perfect agreement with the experimental slow adsorption kinetics found by Bromley et al. [1] (see appendix A). Statistics is too low to deduce a significant difference between the adsorption rates of the two L_{in} clusters, and between the two different L_{out} conformations from the crystal structure.

The variances in the distributions of the orientation angles of the L_{in} conformation with respect to the interface are larger than those of the L_{out} conformation by a factor of ∼2, which is in agreement with the enhanced orientational freedom of BslA-L_{in} observed from the coarse-grained simulations. However, there is no complete tilting of BslA-L_{in} as in the coarse-grained simulations; this may be due to the fact that I considered an air/water interface instead of an oil/water one, to the use of a different force field or to the much shorter investigated time window.

3.3.5 Results: BslA conformation at the interface

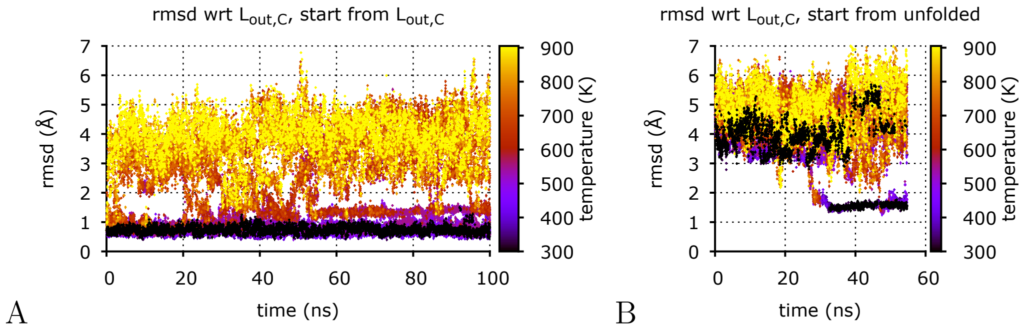
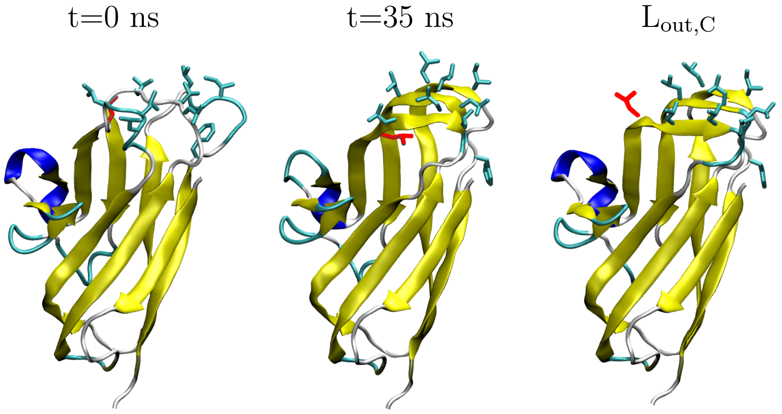


Figure 3.19 *Rmsd from the $L_{out,C}$ conformation observed from the two replica exchange simulations of BslA adsorbed at an air/water interface: (A) starting from the $L_{out,C}$ conformation, and (B) starting from a high-temperature, unfolded conformation of the cap. The different colours correspond to the different effective temperatures of the protein cap in each replica. The replica at ambient temperature (T=300 K) is in black.*

In order to find the equilibrium conformation of BslA when the protein is adsorbed at the interface, I performed replica exchange simulations with solute tempering to enhance conformational changes in the cap region of the protein. I performed two simulation runs: one starting from the ordered cap conformation found the BslA crystal structure (specifically $\mathrm{L}_{\mathrm{out,C}}$), and another one starting from a high-temperature, disordered cap conformation obtained from the first simulation run. In Fig. 3.19 I plot the cap backbone root mean square deviation from the $\mathrm{L}_{\mathrm{out,C}}$ conformation obtained from both simulation runs. We can notice that when the simulation is initiated from the $\mathrm{L}_{\mathrm{out,C}}$ conformation (Fig. 3.19A), the starting conformation is very stable and the rmsd observed at low temperatures is always within 1 Å. On the other hand, at high temperature, the cap unfolds into a very disordered structure.


t=0 ns

t=35 ns

$\mathrm{L}_{\mathrm{out,C}}$

Figure 3.20 *Cap conformational change at the interface from the replica exchange simulation starting from the high-temperature, unfolded conformation of the protein cap. I show the structure of the unfolded starting conformation (left) and the structure observed after 35 ns of replica exchange simulation (centre), which is very similar to the $L_{out,C}$ conformation found in the crystal structure of the protein (right). Residue L119, displaying a different orientation, is highlighted in red.*

It is more interesting to comment on the simulation initiated from the unfolded cap conformation, where there is a conformational change into a structure that is very close to $\mathrm{L}_{\mathrm{out,C}}$. From Fig. 3.20 we can notice that the main difference between the structure observed from the replica exchange and the one found in the crystal structure is the orientation of the hydrophobic residue L119, which is buried in the protein core in the former and exposed to the air phase in the latter. I also note that in this conformation from the replica exchange the aromatic side-chain of residue F159 is exposed to the solvent as in chain C from the crystal structure, and not as in chain A. This simulation was stopped after 55 ns because of the convergence issue discussed below. It is possible that longer time scales and an optimised computational setup would allow the observation of a complete transition from a disordered cap into L_{out}.

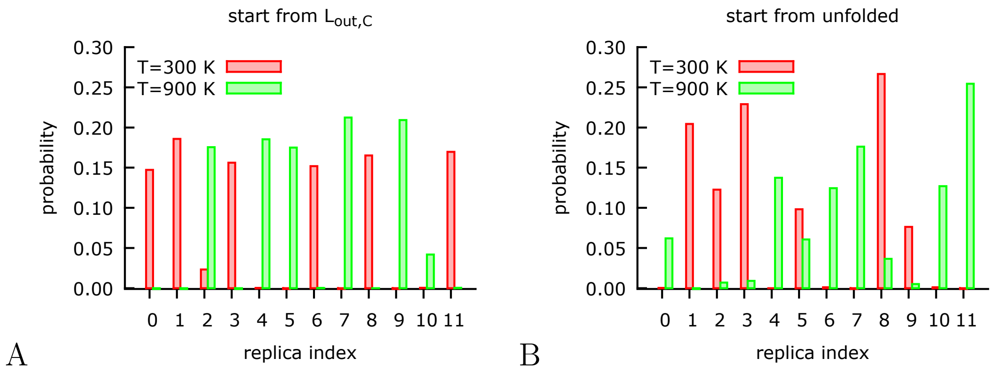


Figure 3.21 *For each system replica (0 to 11) in each simulation run (starting from $L_{out,C}$ (A) or from the high-temperature, unfolded conformation (B)) I plot the percentage of time spent at the lowest (300 K) and at the highest (900 K) temperature. For convergence, each replica should spend an equal amount of time at each temperature. However, in these simulations, the system replicas spend most of their time either at low or at high temperatures.*

By looking at the percentage of time spent at the lowest (300 K) and at the highest (900 K) temperature by each of the 12 replicas (Fig. 3.21), I can deduce a problem in the convergence of both simulations runs. In the ideal scenario, each replica should spend an equal amount of time at each temperature. On the other hand, in these simulations, the system replicas spend most of their time either at low or at high temperatures. This is particularly clear in the first simulation run, where some of the replicas undergo a conformational change from the ordered cap conformation to the disordered one within short time scales, and then they spend most of their time at high temperature without undergoing any conformational change back to the initial ordered cap conformation. This is a problem that has already been documented in replica exchange simulations with solute tempering [118]. It is possible that this technique is not the most appropriate for the study of BslA conformational changes at the interface, and in the future I may take into consideration alternative enhancing techniques such a metadynamics [75].

Nevertheless, the direct observation of a conformational change from a disordered cap to a fully β-sheet ordered cap very close to the one found in the BslA crystal structure represents a key result, which brings further evidence to the conformational change from L_{in} to L_{out} upon adsorption. This result is fully consistent with the higher surface activity of BslA-L_{out} compared to BslA-L_{in} observed from the coarse-grained MD simulations, and it explains the increase in β-sheet secondary structure content observed from circular dichroism experiments [1].

3.3.6 Results: BslA self-assembly at the interface

Modelling the interfacial self-assembly of BslA from the crystal structures of BslA and YweA

Protein crystallisation is ultimately controlled by protein-protein interactions [121]. Since this is true both in solution and at interfaces, can we learn something about BslA crystallisation at interfaces by looking at the the crystal structure of the protein in solution? Firstly, I would like to understand what are the protein-protein interactions that are involved in the self-assembly of the protein at the interface, and finally, I would like to generate a model for the two-dimensional crystal of the protein that is consistent with the experimental findings of Bromley et al. [1], i.e. a rectangular lattice with lattice spacing of 43 and 39 Å. The BslA crystal structure consists of a decameric micelle stabilised by protein-protein interactions that may be involved in the interfacial self-asembly. The group of Daan van Aalten (in collaboration with Prof. Nicola Stanley-Wall) at the University of Dundee recently crystallised the structure of YweA (the pdb file was shared with me and my collaborators at the University of Edinburgh, but it has not been uploaded on the pdb database yet), a protein with a very high sequence similarity with BslA. The crystal structure of YweA (Fig. 3.22B) consists of a tetramer formed by protein units with the same fold as BslA, where the hydrophobic cap region adopts the unstructured L_{in} conformation that can be also found in BslA. I scanned the crystal structures of both BslA and YweA to find a set of protein-protein interactions, corresponding to dimers, that may be involved in the self-assembly of BslA at the interface.

Experiments [1], coarse-grained simulations and all-atom simulations all suggest that, when the protein is adsorbed at an interface, the cap is in the ordered L_{out} conformation, and that the protein adopts an upright orientation where only the hydrophobic residues in the cap stick their side-chains into the apolar phase of the system. Therefore, the interfacial crystal should be stabilised by protein-protein interactions for which each monomer in the dimer adopts the same

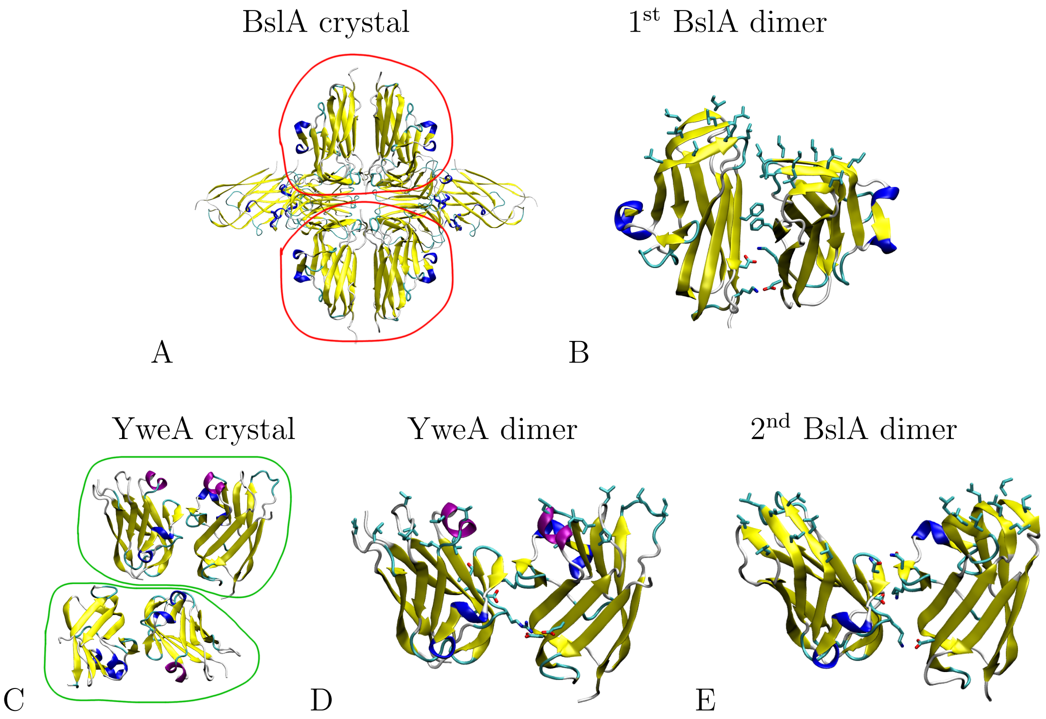


Figure 3.22 *Protein-protein interactions involved in BslA self-assembly at the interface. (A) BslA crystal structure with the dimers consistent with the orientation of the protein at the interface circled in red, corresponding to chains D and G (top), and C and H (bottom). (B) Chains C and H from the BslA crystal, which I will refer to as the 1st BslA dimer. I highlight the main residues involved in the stabilisation of the protein-protein interface: residues F51 in each chain, forming a π-stacking interaction, and residues K59 and D166, forming a pair of salt bridges. (C) YweA crystal structure with the dimers consistent with the orientation of the protein at the interface circled in green, corresponding to chains A and B (top), and C and D (bottom). (D) Chains A and B from the YweA crystal. (E) Aligning BslA (from chain C) onto the YweA dimer, I obtain the 2nd BslA dimer. I highlight the main residues involved in the stabilisation of the protein-protein interface: residues N101 in each chain, forming a side chain hydrogen bond, and residues D66 and K95, forming a pair of salt bridges. There is also a pair of backbone hydrogen bonds formed by residues G100 and A102.*

orientation as the individual adsorbed protein. By looking at the two available crystal structures, I identified two sets of dimers that are consistent with this requirement: one obtained from either chains C and H or from chains D and G from the BslA crystal (Fig. 3.22A), and another one obtained from either chains A and B or from chains C and D from the YweA crystal (Fig. 3.22B). By aligning the BslA conformation onto each of the monomers in the YweA dimer, I could generate a new BslA dimer. I will refer to the BslA dimer found in the BslA crystal as the 1st BslA dimer, and to the YweA-based dimer as the 2nd BslA dimer. These two dimers are both symmetric, i.e. the orientation of one monomer relative to the other is obtained via a rotation of 180 degrees. The protein-protein interface in the 1st dimer is mainly stabilised by the π-stacking interaction formed by residues F51 in each monomer and by a pair of salt bridges between residues K59 and D166, whereas the 2nd dimer is stabilised by a pair of backbone hydrogen bonds between residues G100 and A102 in the protein loop, by a side-chain hydrogen bond formed by residues N101 and by a pair of salt bridges between residues D66 and K95. It is interesting to notice that the 2nd dimer is stabilised by the same type of interactions at the same locations for BslA and YweA, and the only difference is the additional side-chain hydrogen bond found in BslA.

MD simulations show that both BslA dimers are very stable at the interface, since I do not observe any dissociation event within the considered time scale of 100 ns (Fig. 3.24). Moreover, the orientation of each protein in the dimers is in perfect overlap with the distribution of orientations observed for the individual adsorbed protein in the L_{out} form, and the self-assembly has the only effect of lowering the orientational freedom at the interface (see Table 3.1). This is an important requirement for the formation of the dimers after the adsorption of the protein; since an anomalous orientation with respect to the interface will effectively introduce a barrier to the formation of dimers.

By aligning one of the monomers from the 1st dimer with another monomer in the 2nd dimer I can generate a BslA trimer, and by repeating the same operation over and over I obtain a linear BslA chain where the monomer with index i has exactly the same global orientation as the monomer with index $i + 2$, and therefore I obtain a one-dimensional crystal with a unit cell formed by a BslA dimer. This is possible because the considered protein-protein interactions are symmetric, and they act on the protein's global orientation as a rotation by 180 degrees around the same axis, which is equal to the normal to the interface when

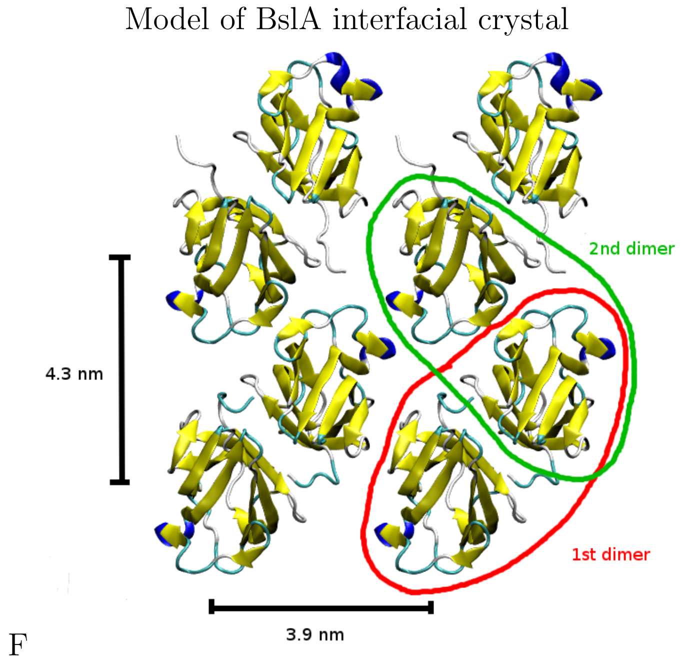


Figure 3.23 *Model of the BslA crystal at the interface obtained by firstly combining the* 1^{st} *BslA dimer with the* 2^{nd} *one into a chain of four proteins, and then by making a copy of this chain shifted by 39 Å along the direction perpendicular to the chain. In this way I obtain a two-dimensional rectangular lattice with a unit cell of sides equal to 43 and 39 Å corresponding to a BslA dimer.*

	$\cos(\theta)$	φ/rad
single	-0.52 ± 0.13	-0.16 ± 0.17
1^{st} dimer	-0.46 ± 0.09	-0.04 ± 0.10
2^{nd} dimer	-0.44 ± 0.06	-0.03 ± 0.13

Table 3.1 *Orientation of BslA with respect to the interface as defined by the averages and standard deviations of the cosine of the angle θ and the angle φ, which are defined in the methods section. I compare the individual adsorbed protein with the one of the protein within the two BslA dimers.*

the protein is adsorbed. Interestingly, the obtained one-dimensional crystal has a lattice spacing of 43 ± 1 Å (the value has been computed from a simulation of a BslA trimer adsorbed at the interface), consistently with the longest lattice spacing of 43 Å observed in the TEM images of BslA films. I generated a model of the BslA interfacial crystal by simply making a copy of the former linear chain of BslA molecules and then shifting it by 39 Å along the direction perpendicular to the chain. In this way I obtain a two-dimensional crystal with P2 symmetry and a rectangular dimeric unit cell with sides of 43 and 39 Å. The model is depicted in Fig. 3.23. Notably, this operation does not generate significant clashes between the atoms, and the neighbouring chains are close enough to interact.

Unfortunately, it was not possible to identify exactly the third additional contact that is required to stabilise the interfacial crystal. However, from an observation of the proposed model, I speculate that two neighbouring chains may form a cross-β-sheet between the third strands of two hydrophobic caps. There could be also a π-stacking interaction involving residue F159, which is flexible, as it displays two alternative orientations in the $L_{out,A}$ and $L_{out,C}$ conformations found in the bulk crystal structure. The exact conformation of this 3^{rd} BslA dimer may be obtained using a protein-protein docking algorithm such as HADDOCK [122], or experimentally via solid state NMR of BslA films [123].

Stability and binding free energies of BslA dimers

In order to test the model of BslA self-assembly at interfaces against experiments, I designed a set of double mutants that are expected to lower the stability of the two protein-protein binding interfaces (one at the time, either the 1^{st} or the 2^{nd} dimer), and therefore prevent the self-assembly. For the 1^{st} dimer I considered the double mutants F51I-D166K and F51A-D166K, which eliminate the pair of salt bridges and the π-stacking interaction. Initially, I only considered the double mutant F51I-D166K, but I found the replacement of the phenylalanine with an isoleucine to be only partly effective (since it replaces a π-stacking with a hydrophobic interaction), so I decided to study the additional mutant F51A-D166K. For the 2$^{\text{nd}}$ dimer I considered the double mutant D66K-N101D, which eliminate the pair of salt bridges and the side-chain hydrogen bond.

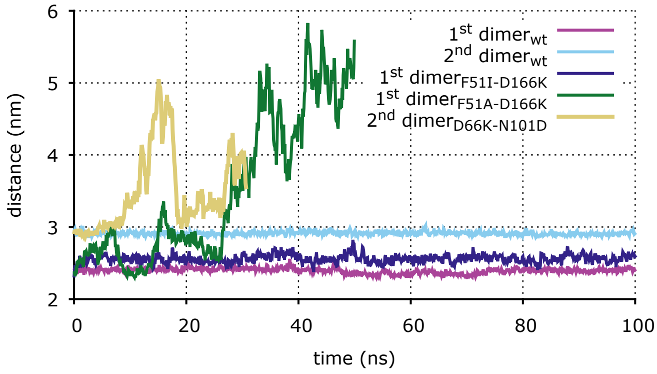


Figure 3.24 *Stability of BslA dimers from all-atom unbiased MD simulations. I plot the distance between the centres of mass of the two monomers as a function of time for the following dimers: the 1st dimer of wt BslA and the double mutants F51I-D166K and F51A-D166K, and the 2nd dimer of wt BslA and the double mutant D66K-N101D.*

In Fig. 3.24 I plot the distance between the two monomers in the dimers of wt BslA and the considered mutants, as observed from unbiased MD simulations. The double mutation F51I-D166K only causes an increase in the equilibrium distance of the 1$^{\text{st}}$ BslA dimer, presumably because of the charge repulsion, but I do not observe any dissociation event within the considered time scale of 100 ns. On the contrary, the mutations F51I-D166K and D66K-N101D have a larger effect and they cause the complete dissociation of the two monomers.

In order to quantify the energetics of adsorption, I computed the potential of mean force along the distance between the monomers for each of the considered BslA dimers (Fig. 3.25). The results show that the 1$^{\text{st}}$ BslA dimer has the strongest protein-protein interface, with a free energy of binding (ΔG_{bind}) of 43±6 kJ/mol, whereas for the second dimer I obtain ΔG_{bind}=31±4 kJ/mol. All the considered mutations are able to lower the strength of the corresponding protein-protein interface. The double mutations F51A-D166K and D66K-N101D are so effective that they respectively lower the free energy of binding of the 1$^{\text{st}}$ and 2$^{\text{nd}}$ dimers down to a value consistent with zero (taking into account the error on the PMF). As suggested by the unbiased MD simulations (Fig. 3.24),

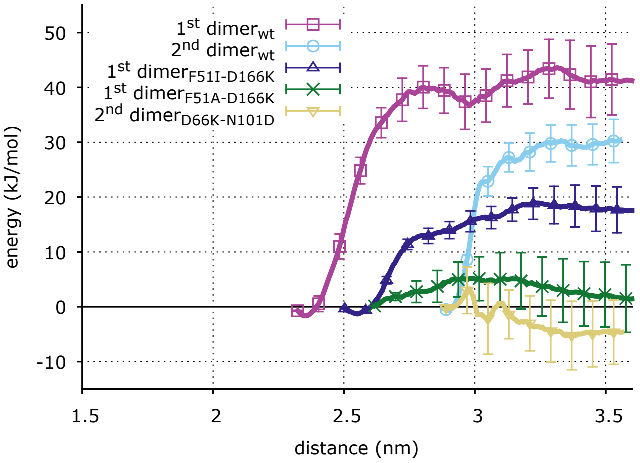


Figure 3.25 *Binding free energy of BslA dimers from non-equilibrium pulling simulations. I plot the potential of mean force along the distance between the centres of mass of the two monomers for the following dimers: the 1st dimer of wt BslA and the double mutants F51I-D166K and F51A-D166K, and the 2nd dimer of wt BslA and the double mutant D66K-N101D.*

the F51I-D166K mutation is less disruptive than the F51A-D166K mutation, but it still lowers the free energy of binding by more than a half down to ΔG=19±4 kJ/mol.

The results of pendant drop experiments performed by Dr. Marieke Schor at the University of Edinburgh, which are shown in Fig. 3.26, confirmed that all the considered mutations reduce the stability of the BslA film at the interface. The D66K-N101D mutation has the strongest effect, whereas the F51I-D166K mutation has the mildest, in agreement with the binding free energies from the simulations.

3.3.7 Conclusions of the all-atom study

In summary, the all-atom MD simulations of BslA brought further evidence to the proposed conformational change of the protein upon adsorption. In solution, MD simulation with replica-averaged restraints based on the NMR chemical shifts revealed a highly disordered ensemble of cap conformations. All the structures of this solution ensemble display the tendency to bury some of the cap hydrophobic residues, and therefore they can all be broadly identified as L_{in} folds. However, some of them are rather different from the L_{in} cap found in the

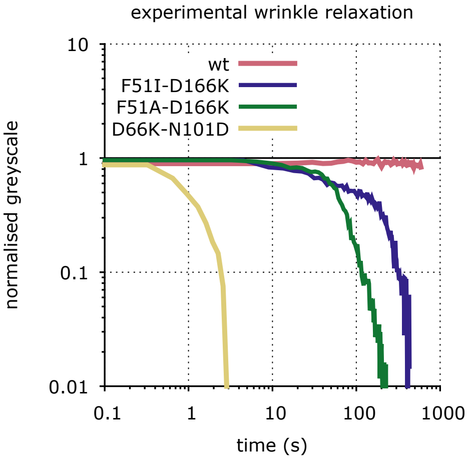


Figure 3.26 *Wrinkle relaxation observed during pendant drop experiments of BslA films: wild type vs mutants targeting self-assembly (F51I-D166K, F51A-D166K and D66K-N101D). The strength of the film is reduced in all cases, with the D66K-N101D mutation having the strongest effect and the F51I-D166K mutation having the mildest.*

BslA crystal structure, which corresponds to the largest cluster of the solution ensemble. When the protein is adsorbed at a air/water interface, replica exchange simulations showed a conformational change from a disordered high-temperature cap conformation to a β-sheet cap very close to the L_{out} structure found in the BslA crystal, in agreement with the increase in β-sheet content from circular dichroism experiments [1]. Furthermore, all-atom simulations of spontaneous adsorption confirmed that BslA-L_{in} has a lower adsorption rate than BslA-L_{out}, offering an explanation for why the kinetics of BslA surface activity is slower than diffusion-limited [1] (appendix A).

These simulations enabled to understand the working mechanism of the protein with atomistic details, highlighting features that are rather unique in the context of surfactant proteins. Despite being often compared to hydrophobins [7, 11], the adsorption mechanism of BslA is actually very different. Hydrophobins display a stable amphiphilic structure stabilised by a network of disulphide bridges in the protein core [11], where their hydrophobic cap maintains its high hydrophobicity both in solution and at the interface. On the other hand, BslA is able to tune the hydrophobicity of its cap depending on the environment, using a conformational change that essentially inverts the orientation of the cap's leucines side chains from an inward orientation (L_{in}) in solution to an outward orientation (L_{out}) at the interface. The residual cap hydrophobicity in the solution ensemble of BslA is high enough to allow the protein to identity the interface and adsorb to it; but at the same time, the reduced hydrophobicity compared to the L_{out} form is advantageous to prevent unwanted clustering in solution (which is indeed experimentally not substantial [1]).

Furthermore, the simulations reveal the importance of disordered regions for the adsorption mechanism of a surfactant protein. Many proteins contains intrinsically disordered regions that fold into ordered structures upon binding to their target sites [124]. If we consider the interface as the target site of BslA, the similarity with intrinsically disordered proteins is very clear: the cap of BslA is a disordered region that folds into an ordered structure upon binding to the interface. The first question to ask would be whether the adsorption and conformational change of BslA corresponds to a conformational selection or an induced fit mechanism [125], or a mixture of both. Since the L_{out} conformation is not observed in the solution ensemble obtained from the simulations with replica-averaged restraints, the adsorption mechanism of BslA broadly falls into the induced fit class, i.e. the protein is always in its L_{in} state in solution, and it undergoes a conformational change into its L_{out} state only after adsorption. Hoverer, given the large conformational variability of the solution ensemble, it is very likely that some conformations have a stronger affinity for the interface than others, introducing an element of conformational selection into the process. Unfortunately, all-atom simulations are very expensive, and I was not able to perform enough simulations of spontaneous adsorption to show whether there is any difference in the affinity of the different clusters of L_{in} conformations.

Finally, the crystal structures of the proteins BslA and YweA have been employed to generate a model of the BslA two-dimensional interfacial crystal consistent with the TEM images of the film obtained by Bromley et al. [1]: a rectangular lattice with lattice spacings of 43 and 39 Å. This model shows that the correct upright orientation of the protein with respect to the interface is fundamental for the formation of the lateral interactions between the adsorbed proteins, offering an explanation for why the mutations that lower the hydrophobicity of the cap region and cause the protein to tilt at the interface also reduce the order of the interfacial film [1]. Thanks to this model I identified the precise interactions that stabilise the two-dimensional interfacial crystal, providing a starting point to tune the surface chemistry of the protein to either decrease (as it was done with the considered double mutations) or increase the strength of the BslA film. Furthermore, using the binding free energies of BslA dimers computed from steered MD simulations, it will be possible to develop a very coarse-grained patchy-particle model [126] to study the self-assembly kinetics of thousands of BslA molecules at the interface.

Chapter 4

Adsorption mechanism of the surfactant protein Ranaspumin-2

Rsn-2 is a surfactant protein involved in the formation of the foam nest of a tropical frog. I used all-atom molecular dynamics simulations to study the surface activity of the protein at oil/water interfaces. In the simulations, the protein first captures the interface via its disordered N-terminal tail, and then it unfolds exposing the hydrophobic residues in the core towards the oil phase. Surprisingly, interfacial unfolding preserves secondary structure elements. To analyse the complete free-energy landscape of protein adsorption, I developed a new structure-based coarse-grained model that is computationally efficient and provides results in agreement with all-atom simulations.

4.1 Introduction

Ranaspumin 2 (Rsn-2) is a surface active protein and a fundamental component of the foam nest of the tungara frog (*Engystomops pustulosus*) [8]. In recent years this protein has received significant attention for its use in the development of a foam for artificial photosynthesis [22]. Surprisingly, the structure of Rsn-2, which has been resolved by NMR spectroscopy (PDB id: 2WGO), does not provide any strong hint to its surface activity. Many surface active proteins, most notably hydrophobins [11], display an extensive hydrophobic patch exposed to the solvent. On the contrary, in Rsn-2 the vast majority of hydrophobic

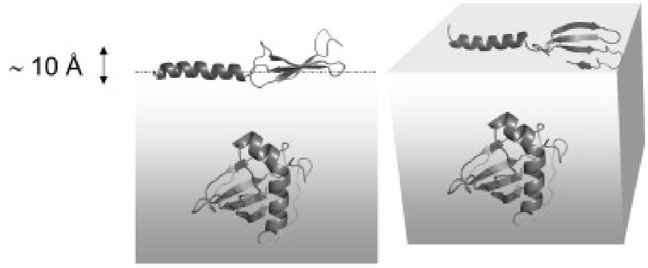


Figure 4.1 *Clamshell model of the Rsn-2 surface activity proposed by MacKenzie et al. [8]: the α-helix is closed on the β-sheet in solution, whereas it opens upon adsorption, exposing the hydrophobic residues of the core to the apolar phase.*

residues are buried in the core of the protein, which is formed by an α-helix folded on top of a β-sheet [8]. The protein also contains a negatively charged C-terminal tail (...DDDDDDGY), and a N-terminal tail containing both charged and hydrophobic residues (LILDGDLLKDKLKLPVI...). The N-terminal domain is highly flexible (unresolved in the NMR structure) and it has been speculated that the first contact that the protein makes with the interface is precisely regulated by the hydrophobic residues present in this region [8]. Furthermore, neutron reflectivity and polarised IRRAS measurements showed two interesting features: first, the protein forms a ~10 Å thick layer at the interface, which is thinner than what would be expected for the globular native conformation, and second, the α-helix and the β-sheet lie on the plane of the interface [8]. These results pointed towards a clamshell mechanism (Fig. 4.1) where the protein unhinges at the linker between the α-helix and the β-sheet, exposing the previously buried hydrophobic residues of the core to the interface [8]. Circular dichroism experiments performed by Ryan Morris at the University of Edinburgh showed that the interaction with the interface does not affect the secondary structure of the protein, consistently with the proposed clamshell mechanism [3].

In order to test the suggested mechanism and to provide further details on the pathway and the energetics of adsorption, I performed both all-atom (Section 4.2) and coarse-grained (Section 4.3) MD simulations of the protein near an oil/water interface. After the description of the computational methods employed, I will present the computational results. All-atom simulations enable us to observe the dynamics of adsorption in great detail without any external input other than the initial protein structure. However, they come at a high computational cost, and the large size of the system prevents a full characterisation of the statistics of the process. On the other hand, the structure-based coarse-grained model I developed for the study of Rsn-2 is less accurate but highly efficient and it enables the computation of the full free energy landscape of the adsorption process. Both computational models give further support to the clamshell model of interfacial unfolding suggested by MacKenzie et al. [8] and they suggest that the hydrophobic N-terminal tail of the protein is responsible for the first step of the adsorption at the interface. These conclusions are fully supported by recent experiments performed in my group at the University of Edinburgh, which will be discussed at the end of the chapter.

Throughout the text, I will refer to the native conformation of the protein as the closed or folded form, and to the partially unfolded conformation at the interface as the open or unhinged form. Moreover, I will refer to the state of the whole system where the protein is adsorbed at the interface in its closed conformation as state 1, and to the state where the protein is partially unfolded exposing the hydrophobic residues of the core towards the oil phase as state 2. The state where the protein is not adsorbed at the interface will be called state 0.

4.2 All-atom study of Rsn-2 adsorption

4.2.1 Methods: all-atom simulations

All-atom MD simulations of Rsn-2 were performed using the software GROMACS 4 [94]. I employed the force field combination tested in Ref. [127], where the protein is represented according to the all-atom AMBER ff99SB-ILDN force field, the water according to the TIP3P model [102], and the cyclohexane molecules were parametrised starting from the Berger model for lipids [128]. This choice was mainly motivated by the fact that this force field combination was tested against the experimental amino-acid partitioning free energies from water to cyclohexane [127]. As we saw from the coarse-grained simulations of BslA, these are amongst the most important parameters that control the adsorption of proteins at interfaces. I also computed the surface tension of the cyclohexane/water interface from a MD simulation of a biphasic system with no protein, obtaining a value of γ=43 mN/m. This is acceptably close to the experimental value of 50 mN/m [98], especially considering the difficulty of all-atom force fields at reproducing the correct surface tensions [129].

In this study I considered two Rsn-2 variants: the wild type (wt), and a N-terminal deletion mutant (d1-15), which lacks the first 15 residues of the N-terminal tail. For each protein variant I performed four independent simulation runs (referred to as 1, 2, 3 and 4) starting from frames 6, 14, 16 and 1 (respectively) of the NMR ensemble deposited in the protein data bank (PDB id: 2WGO). These four protein conformations correspond to the cluster centres of the four clusters identified in the 25 conformations of the NMR ensemble, using the gromos clustering algorithm implemented in the GROMACS tools [94, 99]. To prepare the starting system configuration, the protein was solvated with water in the center of a cubic box with sides 81 Å and 67 Å for the wt and the d1-15 variants respectively. Then I added 2 Na ions to neutralise the charge of the system. Finally, I expanded the size of the box along the z direction by 40 Å and filled the available space with an equilibrated cyclohexane phase.

Before starting each production run I performed the following equilibration steps: energy minimisation with a steepest descent algorithm until the maximum force in the system is lower than 1000 kJ/mol/nm, 50 ps of simulation in the NVT ensemble and 100 ps of simulation in the NAP$_z$T ensemble. In the last two steps, the positions of the heavy atoms of the protein are restrained at their initial positions with a set of harmonic springs with spring constant equal to 1000 kJ/mol/nm^2. Each production run consists of 200 ns of simulation in the NAP$_z$T ensemble, with the exception of run 1 for wt Rsn-2, where the length of the simulation was extended until 350 ns in order to let the protein conformation equilibrate after the observed partial unfolding. The temperature of the system is maintained at 300 K using the Bussi thermostat [66], and the pressure of the system along the z direction was kept at 1 Bar using the Parrinello-Rahman method [68]. The Newton's equations of motion are integrated using the leap-frog algorithm [64] with a time-step of 2 fs. The length of hydrogen bonds was constrained using the LINCS algorithm [103] whereas the whole geometry of water was constrained using the SETTLE algorithm [104]. As in Ref. [127], van der Waals interactions were smoothly switched to zero from 8 to 9 Å. Short-range Coulomb interactions were cut-off at 10 Å whereas long-range electrostatics was treated using the particle mesh Ewald scheme [105] with a Fourier grid spacing of 1.5 Å.

Following the successful characterisation of the competition between protein-solvent and interfacial interactions for the protein BslA [2], I repeated the same analysis here to quantify the role of the N-terminal tail and the unhinging transition at the interface. The contribution to the free energy of adsorption due to protein-solvent interactions, E_{PS}, was estimated as the sum of the water to cyclohexane free energies of transfer of the side-chains exposed to the oil phase [61]. The sum includes the residues for which the position of the center of mass of the side-chain is located above the interface into the oil phase. The position of the interface is estimated as the z-coordinate at which the density of water equals the density of cyclohexane. I estimated the interfacial contribution to the free energy of adsorption from $E_I = -\gamma A_I$, the product of the cyclohexane/water surface tension and the area of interface covered by the protein. The latter should not be confused with the molecular area of the protein, or area per molecule, A_{mol}, which is generally higher than A_I and corresponds to the area per protein of a monolayer with 100% coverage. In order to compute A_I, I defined a regular grid of points on the xy-plane at the z-coordinate corresponding to the interface. A_I was then estimated as the sum of the elements of area (dA=1 Å^2) for which the corresponding grid point is closer than a water van der Waals radius (1.4 Å) to any protein atom. The molecular area was estimated using that same strategy but projecting the protein atomic coordinates on the xy-plane, and therefore ignoring the distance from the interface.

For the analysis of the unhinging transition, the number of hydrophobic contacts in the system were estimated using the following expression:

$$N_{\mathrm{contacts}} = \sum_{ij} \frac{1-(r_{ij}/r_0)^{nn}}{1-(r_{ij}/r_0)^{mm}} \quad , \tag{4.1}$$

where the sum runs over a set of possible hydrophobic contacts, r_{ij} is the distance between the two hydrophobic groups and r_0 is set to 5 Å. Setting nn=6 and mm=12, the fraction inside the sum takes a value close to 1 when $r_{ij} \lesssim r_0$ and close to zero when $r_{ij} \gtrsim r_0$, therefore signalling the formation of a contact between the considered groups. This collective variable is often used when one is interested in identifying the transition between two states of the system (in my case the closed and open conformations of the protein) in combination with metadynamics [130]. To quantify the unfolding of the hydrophobic core of the protein, the sum runs over the possible combinations of contacts between the C$_\gamma$ atoms of the hydrophobic residues in the α-helix and those in the β-sheet of the protein; whereas to quantify protein-oil interactions, the sum runs over the contacts between the C$_\gamma$ atoms of the hydrophobic residues in the core of the protein (α-helix plus β-sheet) and the centres of mass of the cyclohexane molecules.

4.2.2 Results: all-atom simulations

Adsorption of wild type Rsn-2

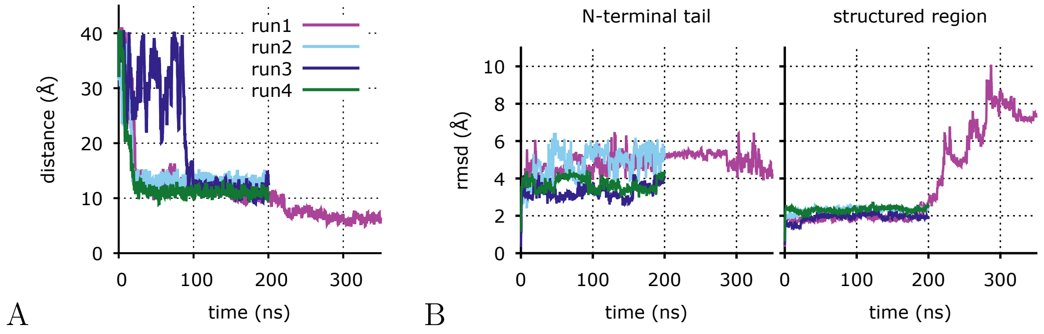


Figure 4.2 *Adsorption of wild type Rsn-2 at a cyclohexane/water interface from all-atom simulations: (A) distance between the center of mass of the protein and the interface, and (B) backbone root mean square displacement (rmsd) from the initial protein conformation for each simulation run. I considered separately the rmsd of the N-terminal tail (up to residue P15) and that of the structured part of the protein (residue ids 16 to 88). The length of the first simulation run was extended up to 350 ns to allow the system to equilibrate after the observed conformational change.*

In all the all-atom simulation runs, wild type Rsn-2 readily adsorbs at the cyclohexane/water interface within short time scales (∼100 ns), which is consistent with its known surface activity [8]. In Fig. 4.2A, the adsorption is highlighted by the lack of diffusion along the direction perpendicular to the interface. From Fig. 4.2A we can also notice that the distance between the protein and the interface right after adsorption is around 12 Å in all simulation runs, but that in the first run the distance decreases down to ∼7 Å starting from ∼200 ns. This decrease in the distance corresponds to a large-scale conformational change that involves the globular part of the protein: Fig. 4.2B shows that in the first run, for residues 16 to 88, the backbone root mean square displacement (rmsd) from the initial configuration increases from ∼2 Å before 200 ns (the equilibrium value for runs 2, 3 and 4) up to ∼8 Å towards the end of the simulation. On the other hand, the N-terminal tail is highly flexible, and its rmsd increases up to values in the range 3-6 Å in all simulation runs right after the beginning of the simulation. It is interesting to notice that the secondary structure content of the protein is not particularly affected by the interaction with the interface (Fig. 4.3); the β-sheet is conserved in all cases and the large-scale conformational change only causes a

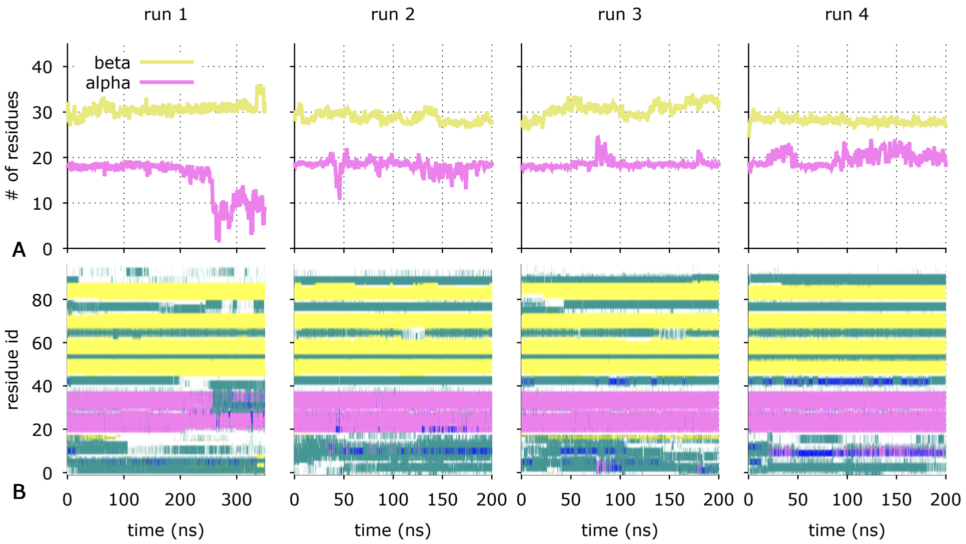


Figure 4.3 *Secondary structure content of wt Rsn-2 during adsorption at a cyclohexane/water interface as a function of time: (A) number of α-helical (purple) and β-sheet (yellow) residues, and (B) secondary structure of each residue (magenta α-helix, blue 3/10 helix, yellow β-sheet, green turn). Secondary structure assignment has been performed using the program VMD [131].*

partial loss of α-helical content.

Visual inspection of the simulations shows that the conformational change in the first run corresponds to the unhinging mechanism proposed in Ref. [8]. Fig. 4.4 shows snapshots of system configurations corresponding to the key steps of Rns-2 adsorption in the first simulation run, and represents well the general interfacial behaviour of the protein. From its bulk configuration (Fig. 4.4A), the protein makes its first contact with the interface through part of its N-terminal tail (Fig. 4.4B), and the whole tail becomes adsorbed within ∼24 ns from the beginning of the simulation (Fig. 4.4C). By looking at the configuration of the system from the direction perpendicular to the interface, we can notice that initially both the β-sheet and the α-helix are oriented perpendicularly to the interface (Fig. 4.4D). However, starting from ∼100 ns the protein changes orientation and the two secondary structure elements are parallel to the interface (Fig. 4.4E). Then, around 200 ns we can observe the increase in the distance between the α-helix and the β-sheet; within a short time, these two elements are far apart, and they expose the side-chains of their hydrophobic residues into the oil phase (Fig. 4.4F).

More insights into the adsorption process can be obtained by looking at the

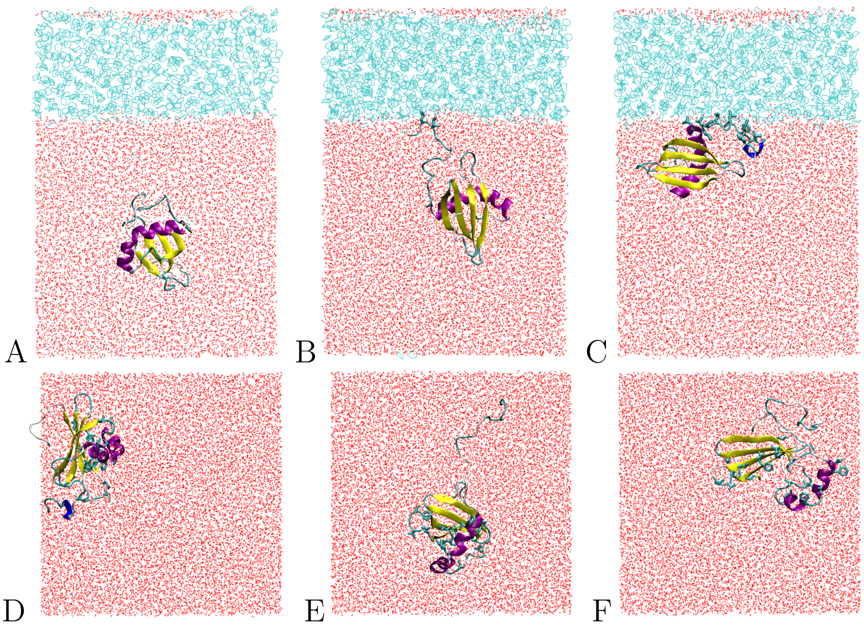


Figure 4.4 *Snapshots of wt Rsn-2 adsorbing at an air/water interface from the first all-atom simulation run: side views at (A) 0 ns, (B) 12 ns and (C) 24 ns, and top views at (D) 24 ns, (E) 131 ns, and (F) 263 ns. The β-sheet is in yellow and the α-helix is in purple.*

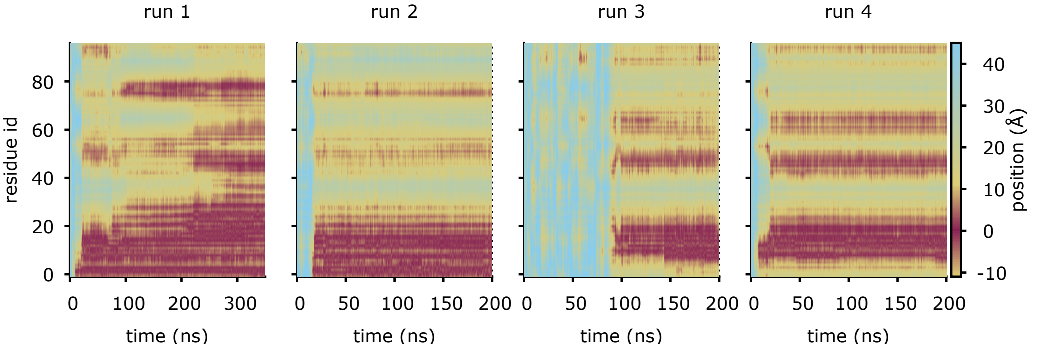


Figure 4.5 *Position of the center of mass of each residue relative to the interface as a function of time from all-atom simulations of wt Rsn-2 adsorption.*

position of each residue relative to the interface as a function of time (Fig. 4.5), allowing to separate the different adsorptions stages and compare the different simulation runs. Firstly, the importance of the N-terminal tail of the protein in the first stages of the adsorption is evidenced by the fact that in all simulation runs this flexible region is always close to the interface from the very first contact. This is likely due to the high hydrophobicity of the region and to its flexibility (Fig. 4.2B), which enables a fast exploration of the possible contacts with the interface. In runs 1 and 2, all the hydrophobic residues in the N-terminal tail have their side-chains oriented towards the oil phase, specifically residues L1, I2, L3, L7, L8, L12, L14, P15, V16 and I17; whereas in the runs 3 and 4 the first three hydrophobic residues are buried in the core of the protein. Secondly, there are also other parts of the protein that interact with the interface, and the specific set of contacts depends on the simulation run. In run 1, during the time window that precedes the unfolding of the hydrophobic core (from ∼100 ns to ∼200 ns), and in run 2, during the whole simulation run after adsorption, the hydrophobic residues at the beginning of the α-helix (L20 and F21) and residues V78 and P79 on the β-sheet are in contact with the oil phase. On the other hand, residues V78 and P79 do not interact with the interface in the third and fourth simulation runs; instead, there is a contact formed by residue L47, and the protein adopts an orientation perpendicular to the interface (rather than the parallel orientation observed in most of run 1 and in run 2). This difference is reflected in the pattern of residue-interface positions reported in Fig. 4.5. Finally, the unhinging transition in run 1 is evidenced by an increase in the number of residues (on both the α-helix and the β-sheet) interacting with the interface during the time window that goes from ∼200 ns until ∼260 ns.

Adsorption of the N-terminal deletion mutant

In order to study the importance of the N-terminal tail for the adsorption, I performed simulations of an Rsn-2 mutant where I truncated residues 1 to 15 (d1-15 Rsn-2). This mutant has also been studied in experiments by Dr. Ryan Morris at the University of Edinburgh; I will discuss his results at the end of the chapter. In the four simulation runs performed, the d1-15 mutant also adsorbs at a cyclohexane/water interface within short time scales (Fig. 4.6A), and the rmsd of the structured part of the protein with respect to its initial configuration (Fig. 4.6B) shows that in the third simulation run the protein undergoes a large-scale conformational change. This change is the same clamshell opening

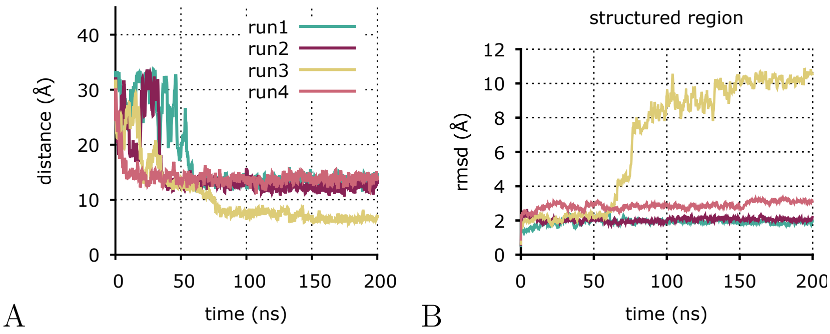


Figure 4.6 *Adsorption of d1-15 Rsn-2 at a cyclohexane/water interface from all-atom simulations: (A) distance between the center of mass of the protein and the interface, and (B) backbone root mean square displacement (rmsd) from the initial protein conformation for each simulation run. The rmsd includes only the structured part of the protein (residue ids 16 to 88).*

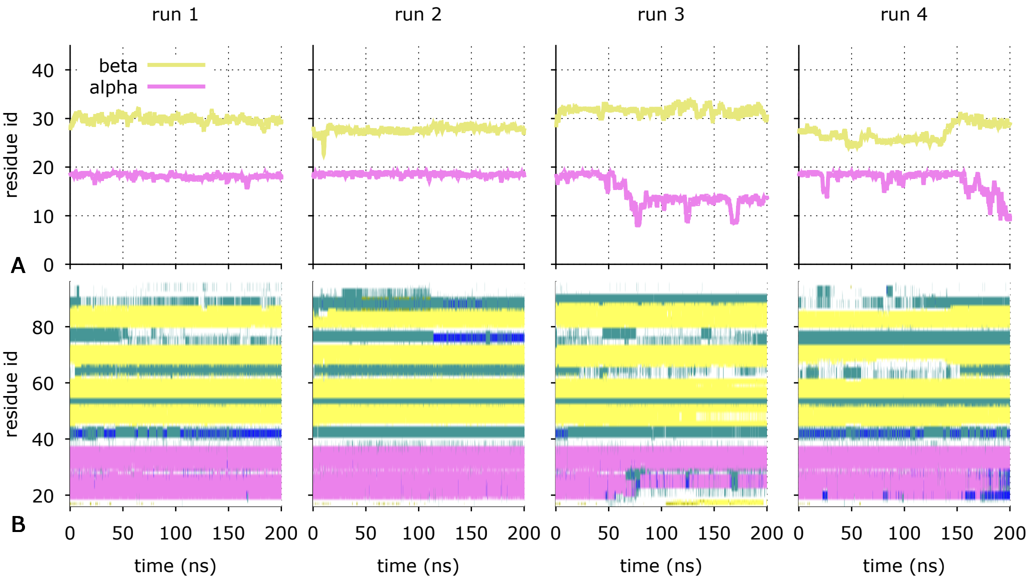


Figure 4.7 *Secondary structure content of d1-15 Rsn-2 during adsorption at a cyclohexane/water interface as a function of time: (A) number of α-helical (purple) and β-sheet (yellow) residues, and (B) secondary structure of each residue (magenta α-helix, blue 3/10 helix, yellow β-sheet, green turn). Secondary structure assignment has been performed using the program VMD [131].*

experienced by wt Rsn-2 in the first run; indeed, even in this case the secondary structure is mostly conserved, and there is only a minor decrease in the helical content (Fig. 4.7).

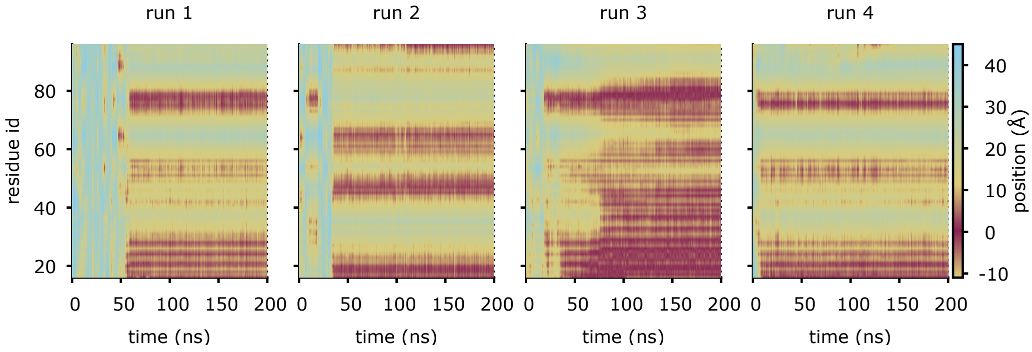


Figure 4.8 *Position of the center of mass of each residue relative to the interface as a function of time from all-atom simulations of the adsorption of the d1-15 Rsn-2 mutant.*

Given the evidence of the importance of the N-terminal tail for the adsorption, it may be surprising that the d1-15 mutant is still able to adsorb at the interface. The reason is that there are still many hydrophobic residues with side-chains exposed to the solvent that form contacts with the interface: residues V16 and I17 in all simulation runs; L20, F21, V78 and P79 in runs 1, 3 and 4; and residue L47 in run 2 (Fig. 4.8).

Clustering of the protein-interface configurations

From the patterns of residue-interface separations (Figs. 4.5 and 4.8) and from visual inspection of the simulation snapshots it appears that when the protein is folded (state 1) it can interact with the interface in two possible ways: one where the α-helix and β-sheet are parallel to the interface (see the snapshot in Fig. 4.9D), with enhanced protein-interface interactions around residues V78 and P79, and another one where the secondary structure elements are perpendicular to the interface (see the snapshot in Fig. 4.9C), with a contact formed by residue L47. In both cases the N-terminal tail interacts with the interface via its hydrophobic side-chains. From now on, I will denote these system configurations as state $1_{\parallel}$ and state $1_{\perp}$ respectively.

In order to characterise the observed patterns of protein-interface contacts quantitatively, I performed a clustering of the simulations frames based on the

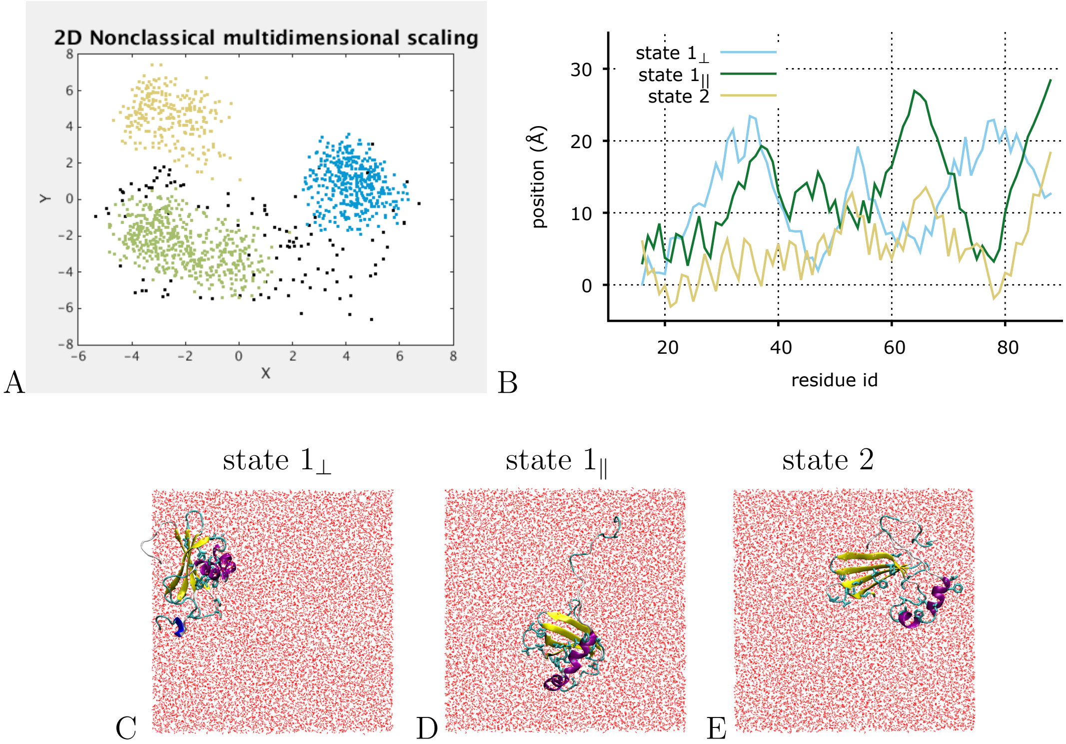


Figure 4.9 *Clustering of residue-interface separations observed in all-atom simulations of Rsn-2 adsorption. Three main clusters have been identified: state* $1_\perp$ *(blue), state* $1_\parallel$ *(green) and state 2 (yellow). (A) The configurations are plotted as a function of two variables defined via non-classical multidimensional scaling (produced with MATLAB [132]). (B) Profiles of the residue-interface separations corresponding to the three cluster centres. (C-E) Top views of wt Rsn-2 adsorbed at an air/water interface from the first all-atom simulation run. We show one frame for each of the identified clusters: at 24 ns (C, state* $1_\perp$*), 131 ns (D, state* $1_\parallel$*) and 263 ns (E, state 2). The* β*-sheet is in yellow and the* α*-helix is in purple.*

positions of the residues centres of mass relative to the interface (reported in Figs. 4.5 and 4.8). For each pair of configurations, I defined the distance between them as the norm of the vector of the differences between the two sets of residue positions:

$$d_{ab} = \sqrt{\sum_{\text{res i}} (z_{ai} - z_{bi})^2} \quad , \tag{4.2}$$

where the sum runs over the considered residues and the indices a and b refer to the two protein-interface configurations. For the clustering I considered all the frames of all simulation runs (every 1 ns) after adsorption, and I only included the residues corresponding to the structured part of the protein (ids from 16 to 88, this is necessary to include both wt and d1-15 Rsn-2). I employed the recently developed algorithm of clustering by search and find of density peaks [133], using a distance cutoff of 1.25 Å (the results are rather robust with respect to this parameter).

The method identifies three main clusters (Fig. 4.9): one corresponding to state $1_{||}$ (folded protein, enhanced interactions around residues V78 and P79), one to state $1_{\perp}$ (folded protein, enhanced interactions around residue L47) and one to state 2 (unhinged conformation, uniform interactions involving both the α-helix and the β-sheet). The profiles of the residue-interface separations corresponding to the three cluster centres are reported in Fig. 4.9B. The clustering also enables analysis of the path of the system through the different states in each simulation run. Because of the metastability of the states, when the protein adsorbs in a certain configuration, it usually stays in the same one for the whole time until the end of the simulation. However, I could observe one orientational transition from state $1_{\perp}$ to state $1_{||}$ in run 1 for wt Rsn-2, and two unhinging transitions from state $1_{||}$ to state 2 in run 1 for wt Rsn-2 and in run 3 for the d1-15 mutant. The system only explores state $1_{||}$ in run 2 for wt Rsn-2 and runs 1 and 4 for d1-15 Rsn-2, and it only explores state $1_{\perp}$ in runs 3 and 4 for wt Rsn-2 and run 2 for d1-15 Rsn-2.

Driving forces of adsorption

In this section I discuss the energetic contributions that drive the adsorption of Rsn-2 at the oil/water interface, distinguishing between the two main types of interactions: the interfacial energy given by the product of the oil/water surface tension and the area of interface covered by the protein (E_I=$-\gamma A_I$), and the

			units					
			ns	nm^2		kJ/mol		
	run	state	Δt_{window}	A_{mol}	A_I	E_I	$E_{PS,\text{1-15}}$	$E_{PS,\text{16-96}}$
wt	1	$1_{\parallel}$	100-150	12.3	2.6	-67	-88	3
	1	2	300-350	14.9	5.3	-136	-85	-47
	2	$1_{\parallel}$	100-200	11.1	2.7	-69	-100	-13
	3	$1_{\perp}$	100-200	10.8	2.1	-54	-20	5
	4	$1_{\perp}$	100-200	10.2	2.4	-62	-59	2
d1-15	1	$1_{\parallel}$	100-200	8.6	0.8	-21		1
	2	$1_{\perp}$	100-200	8.7	1.2	-31		2
	3	$1_{\parallel}$	40-60	9.0	1.4	-36		-3
	3	2	100-200	11.9	4.0	-103		-80
	4	$1_{\parallel}$	100-200	8.9	0.9	-23		-16

Table 4.1 *Area per molecule* A_{mol}*, interfacial area* A_I*, interfacial energy* $E_I = -\gamma A_I$ *and protein-solvent energies* E_{PS} *for the two protein variants (wt and d1-15) during different time windows. The protein-solvent energy has been separated between the contribution from the N-terminal tail (residues 1 to 15),* $E_{PS,1\text{-}15}$*, and the one from the rest of the protein (residues 16 to 96),* $E_{PS,16\text{-}96}$*. The third column of the table refers to the conformational states of the adsorbed protein (see clustering). Errors on the averages are on the order of 0.1 nm*^{2} *for areas and 2 kJ/mol for energies.*

protein-solvent energy due to the transferring of residues from the water phase to the oil phase. Table 4.1 shows the average energy contributions to the adsorption for the different simulation runs during the different stages of adsorption: states $1_{\parallel}$, $1_{\perp}$ and 2. This analysis shows that the unfolding of the protein optimises both the interfacial energy (caused by an increase in interfacial area) and the protein-solvent energy contribution. The latter is caused by the hydrophobic residues in the core of the protein exposing their side-chains into the oil phase and it does not affect the interaction of the N-terminal tail with the interface. The increase in interfacial area is accompanied by an increase in the area per molecule, which compares favourably with the coarse-grained simulations and experimental results (see section 4.4). The less favourable protein-solvent energy observed for wt Rsn-2 in the third and fourth runs compared to runs 1 and 2 is caused by the fact that not all the residues in the N-terminal tail are interacting with the interface; this configuration is likely to represent only a metastable state, and the protein should change its orientation from perpendicular to parallel to the interface before the unhinging transition. This analysis is also important to understand the effect of the N-terminal tail deletion. When the protein is folded at the interface, the protein-solvent interaction between the tail and the oil phase

($E_{PS,1-16}$) represents the main driving force to the adsorption. When the first 15 residues of the protein are absent, the protein-solvent contribution is essentially zero, and the first adsorption state is controlled by the interfacial energy E_I; this contribution is still significantly higher that k_BT ~2.5 kJ/mol, which could explain why the adsorption of the mutant is observed in the simulations.

These values are only useful to compare the relative importance of the interactions for the different systems, but they cannot be used to estimate the total free energy of adsorption of the protein. This is mainly because I did not attempt to estimate the change in the free energy of the system due to the unfolding of the hydrophobic core of the protein. However, Fig. 4.10 shows that the breakage of the hydrophobic interface between the α-helix and the β-sheet occurs simultaneously with the “invasion” of cyclohexane molecules into the core of the protein. The formation of new hydrophobic contacts between side-chains and cyclohexane may help to reduce the barrier to the unfolding of the protein core.

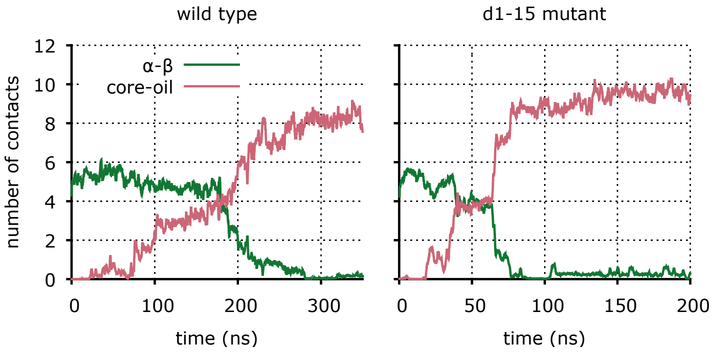


Figure 4.10 *Unfolding of the Rsn-2 hydrophobic core from all-atom simulations. For the two simulation runs were the unhinging of the protein core occurs (run 1 for wt Rsn-2 and run 3 for d1-15 Rsn-2) I plot the number of hydrophobic contacts (see methods) between the α-helix and the β-sheet of the protein (green curve), and the contacts between the core of the protein (which excludes the N- and C-terminal tails) and the cyclohexane molecules (pink curve).*

These results suggest that the adsorption of Rsn-2 at interfaces is a two-step process: in the first step, the folded protein moves from bulk to the interface, whereas in the second step the hydrophobic core of the protein collapses, and the α-helix and the β-sheet maintain their native fold while exposing their apolar side towards the oil phase. The first step is driven by the favourable interaction between the hydrophobic residues on the N-terminal tail and the interface. The flexibility of the N-terminal domain at the interface suggests that the kinetics of the first adsorption step is controlled by diffusion. On the other hand, the second step of the adsorption is only occasionally observed in the simulations, and the free energy barrier is probably due to the breakage of the hydrophobic contacts in the core of the protein. Since the second step of the adsorption does not involve a transfer of residues from water to oil, but simply a switch between hydrophobic interfaces, this step is probably driven by the large increase in the interfacial area occupied by the protein at the interface, which corresponds to a decrease in the interfacial energy of the system.

4.3 Coarse-grained study of Rsn-2 adsorption

4.3.1 Methods: a new structure-based coarse-grained model for protein adsorption at interfaces

Here I describe a structure-based coarse-grained model [30] developed for the study of Rsn-2 adsorption at interfaces. The method is general and it could be applied to the study of other surface active protein; I will discuss this point in more details at the end of the chapter. The reason for the development of this model is that all-atom simulations, because of their high computational cost, do not allow a complete study of the free energy landscape of the system. In particular, in the study of Rsn-2 surface activity, once the protein adsorbs at the interface with a certain orientation, it is often stuck in the same initial metastable state ($1_{\parallel}$ or $1_{\perp}$) for the whole length of the simulation run, and the unhinging conformational change is only observed in 2 runs out of 8. This implies that there could be other stable conformations that are simply not observed because of the low statistics or because of the free energy barrier to access them. Furthermore, even though these simulations are extremely informative and very detailed, I cannot extract the free energy differences between the observed states of the system, which would be particularly useful for the comparison between the wild type protein and the d1-15 mutant. My coarse-grained model is computationally efficient and it will fill the gaps in the results obtained from all-atom simulations.

Contrary to physics-based models, where the interaction potential of the protein depends solely on its amino-acid sequence, in structure-based models (which are also know as Gō models [134]) interactions are defined by the native conformation of the protein, as obtained from the crystal structure or from NMR spectroscopy [135]. In general, any pair of residues interact via an attractive non-bonded potential when there is a contact between them in the native structure of the protein, and via a purely repulsive term otherwise. Despite their extremely low computational cost (the number of beads in the system is usually on the order of 10^2, compared to 10^4 for all-atom simulations), structure-based models have been successful at predicting the folding pathway of many proteins [30], because the number of native contacts represents a good collective variable for describing protein folding [136]. Furthermore, these models can be easily extended to the study of more complex scenarios, many of which are currently beyond the possibilities of all-atom simulations. Some notable examples include: coupled binding and conformational changes [26, 125, 137, 138], protein-dna interactions [32, 33, 139, 140] and finally protein adsorption at interfaces [141].

In Ref. [141], the authors developed a structure-based model to study the adsorption of proteins at interfaces. In their model, each residue of the protein is coarse-grained into a single bead centred on the C_α atom, and residues interact with a flat interface via an energy term proportional to the hydrophobicity of the amino-acid, so that hydrophilic residues tend to stay in the aqueous phase, whereas hydrophobic ones tend to partition into the apolar phase. The model was used to study the adsorption of various proteins, their conformational change upon adsorption and the behaviour of protein monolayers under shear. However, the authors did not provide a comparison with relevant experimental results. The main criticism to this approach regards the coarse-graining of residues onto one single backbone bead centred on the C_α atom. In my work on BslA (section 3.2), I found that the conformation of the hydrophobic cap greatly affects both the free energy of adsorption and the orientation of the protein at the interface. This is because the conformational change involves the burial of many hydrophobic side-chains that will not be able to participate to the interaction with the interface; such a change would not be captured by a coarse-grained model that neglects the position of side-chains. Furthermore, protein secondary structure elements generally possess a strong hydrophobic dipole [28], i.e. they possess well-defined polar and apolar sides. For instance, when Rsn-2 is in solutions the apolar sides of its β-sheet and α-helix are closed on top of each other and shielded from water, whereas in the unhinged conformation at the interface these sides are exposed to the oil phase. In general, a secondary structure element should adsorb at the interface with the hydrophobic side facing the apolar phase, but again this would not be necessarily the case within a model where side-chains are omitted.

In order to address these issues, I developed a structure based model where each residue is coarse-grained onto a backbone bead and a side-chain bead, similarly to what is done in the MARTINI force field [142], and where adsorption is modelled by an extra interaction term between each side-chain and an ideal interface. We will see in the results section that this model faithfully reproduces the adsorption of Rsn-2 in agreement with both all-atom simulations and experiments. For each residue, the backbone coarse-grained bead is placed at the position of the C_α atom, whereas the side-chain bead, which is omitted for the amino acid glycine, is placed at the center of mass of the side-chain. To simplify the notation and improve the transferability of the model, I will employ the units of ϵ for energies and m for mass: ϵ represents the interaction energy between two residues in contact (see later for more details) whereas m is the mass of a each coarse-grained bead (all equal). The protein potential energy function depends solely on the native conformation, and it is given by:

$$\begin{aligned} V_{\text{protein}} = & \sum_{\text{bonds}} \epsilon_r (r_i - r_0)^2 + \sum_{\text{angles}} \epsilon_\theta (\theta_i - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} \epsilon_\phi \{[1 - \cos(\phi_i - \phi_0)] + 1/2\,[1 - \cos(3(\phi_i - \phi_0))]\} \\ & + \sum_{\text{native}} \epsilon_c \left[\left(\frac{\sigma_c}{r_i} \right)^{12} - 2 \left(\frac{\sigma_c}{r_i} \right)^{6} \right] + \sum_{\text{non-native}} \epsilon_{nc} \left(\frac{\sigma_{nc}}{r_i} \right)^{12} \quad , \end{aligned} \tag{4.3}$$

where the equilibrium values of bond lengths r_0, angles θ_0 and dihedrals ϕ_0 are equal to their corresponding values found in the NMR or crystal structure of the protein. The energetic constants ϵ_r, ϵ_θ, ϵ_ϕ and ϵ_{nc} do not depend on the native conformation of the protein and are set to ϵ_r=100ϵ, ϵ_θ=20ϵ, ϵ_ϕ=0.5ϵ and ϵ_{nc}=0.01ϵ. The attractive non-bonded interactions (native) are defined between each pair of coarse-grained beads for which any two atoms form a contact in the native structure. For example, for the residues within the Rsn-2 β-sheet that interact via a backbone hydrogen bond I define an attraction amongst the corresponding coarse-grained backbone beads. Similarly, for the residues of the β-sheet that interact with another residue in the α-helix via a hydrophobic contact in the protein core I define an attraction between the two coarse-grained side-chain beads. The all-atom native contacts are obtained from the coordinates of the heavy atoms in the NMR structure using the shadow map method [143, 144] with a cut-off radius of 6 Å and a shadowing radius of 1 Å. In general, for each pair of residues i and j there are 4 possible coarse-grained interactions: backbone with backbone (bb), side-chain with side-chain (ss), backbone of residue i with side-chain of residue j (bs) and side-chain of i with backbone of j (sb). For each pair of residues, the strength of the interaction energy $\epsilon_{c,xx}$ for the interaction type xx is given by:

$$\epsilon_{c,xx} = \epsilon \frac{n_{xx}}{n_{bb} + n_{ss} + n_{bs} + n_{sb}} \tag{4.4}$$

where n_{bb}, n_{ss}, n_{bs} and n_{sb} are the number of all-atom native contacts that are involved in that interaction type. For instance, let us consider two residues i and j; if there are two contacts between the atoms of the two side-chains and one contact between the atoms in the backbone, then the coarse-grained interaction energies are: $\epsilon_{c,bb}$=1/3ϵ, $\epsilon_{c,ss}$=2/3ϵ, $\epsilon_{c,bs}$=0 and $\epsilon_{c,sb}$=0. This means that the total interaction energy between two residues that form at least one native contact is always equal to ϵ, and that the contact energies are distributed amongst residues in the same way as in the popular structure-based model developed by Clementi et al. [30]. The equilibrium distance of the attractive interaction potential (σ_c) is equal to the distance between the coarse-grained beads in the native structure. Disordered regions are not involved in the formation of native contacts and there is no dihedral potential biasing the configuration to a particular secondary structure element. This is the case for the N-terminal domain of Rsn-2, where the equilibrated structure deposited on the PDB is only used for the calculation of the equilibrium bonds and angles. For the non-native non-bonded interactions, the potential is purely repulsive, and the size σ_{nc} is set to the sum of the radii of the two beads involved in the interaction, which are defined as $r_{nc}=2(n_{\text{heavy}}/4)^{1/3}$Å, where n_{heavy} is the number of heavy atoms that form the coarse-grained bead (equal to 4 for the backbone bead and variable for the side-chain).

For the study of Rsn-2, the structure-based parameters in Eq. 4.3 were extracted from the first configuration of the protein NMR ensemble (PDB id: 2WGO [8]); given the similarity between the 25 conformations of the ensemble, the precise choice does not affect the results. In order to test the new model, I performed a simulation of the protein at the folding temperature, and I compared it against the structure-based model by Clementi et al. [30]. In both cases, the protein jumps between two well-defined folded and unfolded states following the same pathway (see below, Fig. 4.19). The observed two-state folding behaviour is consistent with experimental results [8]. The predictions are robust with respect to changes in the protein potential energy function, e.g. the relative importance of non-bonded and dihedral interactions or the precise form of the non-bonded interaction used.

In order to model protein adsorption, I represent the interface as an ideal wall that attracts or repels the amino acids side-chains depending on their hydrophobicity.

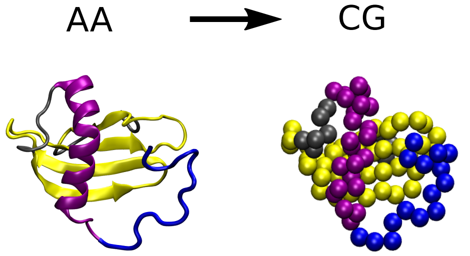


Figure 4.11 *Schematics of the coarse-graining of Rsn-2: all-atom (left, highlighting secondary structure features) and coarse-grained (right, showing the backbone beads, side-chain beads are omitted for simplicity) structures. The N-terminal tail is in blue, the α-helix in purple, the β-sheet in yellow, and the linker and the C-terminal tail are in grey.*

The interaction takes the following form:

$$V_{\text{interface}} = \sum_{\text{hydrophobic}} \epsilon_i \left[\frac{1}{2}\left(\frac{\sigma_w}{r_{iz}}\right)^9 - \frac{3}{2}\left(\frac{\sigma_w}{r_{iz}}\right)^3 \right] + \sum_{\text{hydrophilic}} \epsilon_i \left(\frac{\sigma_w}{r_{iz}}\right)^9 \quad , \tag{4.5}$$

where the thickness of the interface is set to σ_w=8 Å, the two sums run over hydrophobic and hydrophilic residues, and r_{iz} is the distance between the side-chain of residue i and the interface (the backbone bead of glycine is included in the sum as a side-chain). The backbone beads interact with the interface through a potential equal to the one for hydrophobic residues, but with a cut-off at σ_w, which makes the interaction purely repulsive. A residue is considered hydrophobic/hydrophilic if its water to cyclohexane partitioning free energy [61] is lower/greater than zero; the interaction energy ϵ_i is proportional to the absolute value of this partitioning free energy. In the simulations of Rsn-2, the proportionality constant is set so that the ratio between $E_{\text{I,max}}$, the sum of the interaction energies between the hydrophobic residues and the interface, and E_{folding}, the sum of the contact energies ϵ_c that control the folding of the protein, is equal to k=$E_{\text{I,max}}/E_{\text{folding}}$=0.65. The interfacial strength k is a controllable parameter that is necessary to define whenever a structure-based model of a protein is used to study the binding process to another macromolecule or interface, e.g. another protein [138] or DNA [33, 139]. In my case, the choice was motivated by the observation of results consistent with the experiments; specifically, the clamshell unfolding of the hydrophobic core upon adsorption without loss of secondary structure. These results are robust in the range from k=0.6 to 0.8. However, when the interface is too weak, no unhinging transition is observed, whereas when the interface is too strong the protein undergoes a complete unfolding. These scenarios may occur under certain experimental conditions that have not been explored yet, but I decided to limit my study to the conditions relevant to the available experiments.

4.3.2 Methods: details of the coarse-grained simulations

MD simulations where performed using the software LAMMPS [145] in combination with PLUMED 2 [116] when the latter was needed for the free energy calculations. The equations of motion are integrated using Langevin dynamics with an integration time-step of dt=$0.001\sqrt{m/\epsilon}d_0$ and a relaxation time equal to τ_{rel}=$2.6\sqrt{m/\epsilon}d_0$, where d_0=3.8 Å is the typical bond length between two consecutive backbone beads. Simulations of adsorption are performed at the ambient temperature of T=$0.9T_f$, where T_f is the folding temperature of the protein. The ambient temperature is chosen so that the native conformation of the protein is stable in bulk and no unfolding events are observed. The interface described above was defined perpendicularly to the z-direction, and another purely repulsive wall was defined 12 nm away from the interface to avoid the diffusion of the protein in the direction opposite to the interface. I employed periodic boundary conditions along both $\hat{x}$ and $\hat{y}$.

As before, I refer to the state where the protein is folded and solvated in the aqueous phase as state 0, to the one where it is folded and adsorbed at the interface as state 1, and to the one where it is unfolded and adsorbed as state 2. Consistently with the all-atom simulations, the adsorption of Rsn-2 proceeds via a two-step mechanism where first the protein grabs the interface through its N-terminal tail, and then undergoes the unhinging transition. To study the spontaneous adsorption of Rsn-2, I place the protein in its native structure 6 nm away from the interface, and I observe the molecular dynamics of the system until adsorption occurs and the protein reaches its equilibrium conformation. I also computed the free energy landscapes for the two adsorption steps separately: in the first step (state 0 to state 1) the protein can adsorb and desorb from the interface but it always remains folded in its native conformation, whereas in the second step (state 1 to state 2) the protein is kept at the interface and is allowed to transition between its closed and open conformations. I considered three different protein variants: the wild type protein (wt), the same N-terminal tail deletion considered in the all-atom study (d1-15) and a mutant with two extra disulphide bridges between residue ids 19 and 46 and between residue ids 32 and 81 (2C-C).

Despite the low computational cost of the new structure-based model, the interaction energy with the interface was very high, and it was not possible to reconstruct the free energy profile from unbiased MD simulations, since adsorption is irreversible within the accessible time scales. The free energy profiles were computed using the well-tempered metadynamics method [75], a technique that adds an external history-dependent bias that converges to a potential proportional to the free energy of the system along the set of collective variables used for the bias [146] (see background section 2.2). In order to build a useful understanding of the target process, the system should be biased along a set of collective variables that are able to efficiently discriminate between the different metastable states of the system; in my case between states 0 and 1 for the first adsorption step, and between states 1 and 2 for the second adsorption step. Overall, I defined the following collective variables:

- $d_{\text{interface}}$, the distance between the center of mass of the globular part of the protein (Rsn-2_{16-96}, which excludes the N-terminal tail) and the interface.
- d_{tail}, the z-component of the distance between the center of mass of Rsn-2_{16-96} and that of the N-terminal tail (I included residue ids 1 to 15 for wt and 2C-C variants, and residue id 16 only for the d1-15 variant). The use of this variable was useful to help the protein to reorient itself as it approaches the interface and it was necessary to avoid hysteresis in the reconstruction of the free energy profiles.
- N_{native}, the number of native contacts of the protein, which are defined using the same switching function that I employed for the the analysis of the all-atom simulations:

  $$N_{\text{native}} = \sum_{\text{native}} \frac{1-(r_i/r_0)^{nn}}{1-(r_i/r_0)^{mm}} \quad , \tag{4.6}$$

  where the sum runs over the pairs of coarse-grained beads that interact via an attractive non-bonded potential, nn=8 and mm=10, $r_0 = 1.2\sigma_c$, where σ_c is the equilibrium distance of the considered non-bonded interaction, and r_i is the distance between the two beads in the contact. This variable takes a value close to zero for a completely unfolded protein and a value close to $N_{\text{native}}^{\max}$=224 for a folded protein. In the analysis, in order to separate the folding of the secondary structure elements from the folding of the hydrophobic core, I used the same function, but the sum was limited respectively to the native contacts that are involved in the formation of the α-helix and the β-sheet, and to those that form the protein core (the contacts between the α-helix and the β-sheet).

To reconstruct the free energy of the first step of the adsorption with metadynamics, I biased the system along the variables $d_{\text{interface}}$ and d_{tail}. For this adsorption step, in addition to the metadynamics bias, I also added a constant harmonic wall that keeps the number of native contacts higher than the value $N_{\text{native}}^{\min}$=168; this allows the natural fluctuations of the protein structure while keeping it in its folded state. For the second step of the adsorption, the system was biased along the variables $d_{\text{interface}}$ and N_{native}, and the complete desorption of the protein was prevented by placing a Lennard-Jones wall acting on all coarse-grained beads 6 nm away from the interface. For the folding of the protein in bulk at the transition temperature T_f I applied the metadynamics bias along the number of native contacts N_{native}. The free energy surfaces in Fig. 4.19 as a function of the secondary structure native contacts and the hydrophobic core native contacts were reconstructed from the corresponding metadynamics simulations using the reweighting scheme discussed in Ref. [147].

In all well-tempered metadynamics simulations, I set the initial height of the hills to $w_0 = 0.5 k_B T$ and the frequency of hill deposition to 500 times the integration time-step (see background section for more details on metadynamics). The temperature parameter ΔT was varied between $5T$ and $19T$ depending on the system (high ΔT is needed to cross high free energy barriers, but it also increases the error on the free energy profile): $\Delta T_{\text{wt,step1}} = \Delta T_{\text{2C-C,step1}} = \Delta T_{\text{2C-C,step2}} = 19T$, $\Delta T_{\text{wt,step2}} = \Delta T_{\text{d1-15,step1}} = \Delta T_{\text{d1-15,step2}} = 9T$, $\Delta T_{\text{wt,bulk}} = 5T$. The errors on the reported free energies are always within $\sim 1.5 k_B T$; convergence was assessed by looking at the difference in free energy between the two relevant metastable states. The width of the hills was set equal to the standard deviation of the collective variable in the considered system: $\sigma_{\text{wt,step1}}^{d_{\text{interface}}}$=$\sigma_{\text{d1-15,step1}}^{d_{\text{interface}}}$=$\sigma_{\text{2C-C,step1}}^{d_{\text{interface}}}$=0.1Å, $\sigma_{\text{wt,step2}}^{d_{\text{interface}}}$=$\sigma_{\text{d1-15,step2}}^{d_{\text{interface}}}$=$\sigma_{\text{2C-C,step2}}^{d_{\text{interface}}}$=0.05Å, $\sigma_{\text{wt,step1}}^{d_{\text{tail}}}$=$\sigma_{\text{2C-C,step1}}^{d_{\text{tail}}}$=0.15Å, $\sigma_{\text{d1-15,step1}}^{d_{\text{tail}}}$ = 0.2Å, $\sigma_{\text{wt,step2}}^{n_{\text{native}}}$=$\sigma_{\text{d1-15,step2}}^{n_{\text{native}}}$=$\sigma_{\text{2C-C,step2}}^{n_{\text{native}}}$=$\sigma_{\text{wt,bulk}}^{n_{\text{native}}}$=6.

4.3.3 Results: coarse-grained simulations

Adsorption of wt Rsn-2

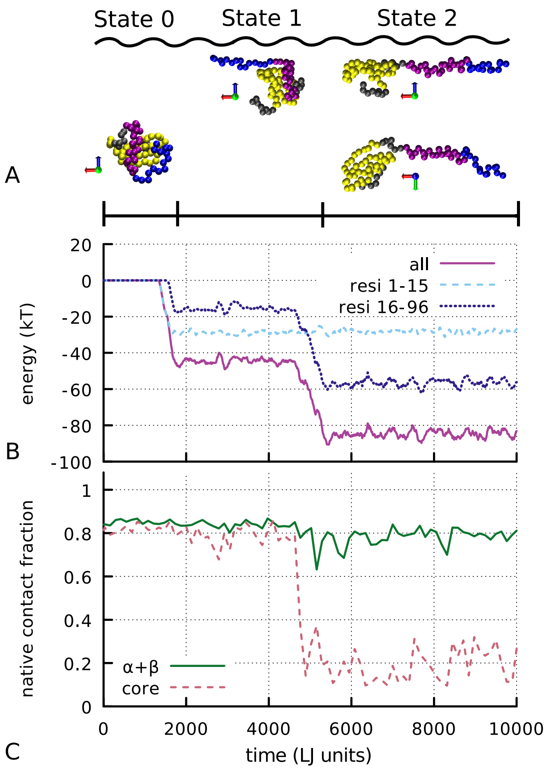


Figure 4.12 *Adsorption of wt Rsn-2 observed in the coarse-grained simulations. (A) cartoon of the adsorption states of the protein: folded in bulk (state 0), folded at the interface (state 1), and unhinged at the interface (state 2). (B) Interaction energy between the protein and the interface, computed directly from the expression of the potential energy (Eq. 4.5). The total protein-interface energy (purple line) has been separated between the contribution from the N-terminal tail (residues 1 to 15, dashed cyan line) and from the rest of the protein (residues 16 to 96, dotted blue line). (C) Fraction of native contacts corresponding to the secondary structure elements (α-helix and β-sheet, green line) and to the hydrophobic core of the protein (core, dashed pink line).*

Using the structure-based model for protein adsorption described above, I performed MD simulations of wt Rsn-2 near an oil/water interface. The protein spontaneously adsorbs in all performed simulation runs (5 in total) via the same two-step route (Fig. 4.12A) observed in all-atom simulations: from its bulk state (state 0) it adsorbs at the interface interacting via its hydrophobic N-terminal tail (state 1), then it undergoes the unhinging structural change where both the α-helix and the β-sheet expose their hydrophobic sides lying parallel to the interface (state 2). Fig 4.12B shows the interaction energy between the protein residues and the interface: the first adsorption state is indeed dominated by the interaction with the N-terminal tail, whereas the second one is dominated by the residues in the structured part of the protein. Moreover, consistently with the all-atom simulations and with the experiments, the unhinging transition is not associated with a loss of secondary structure content (Fig 4.12C).

Clustering of the protein-interface configurations

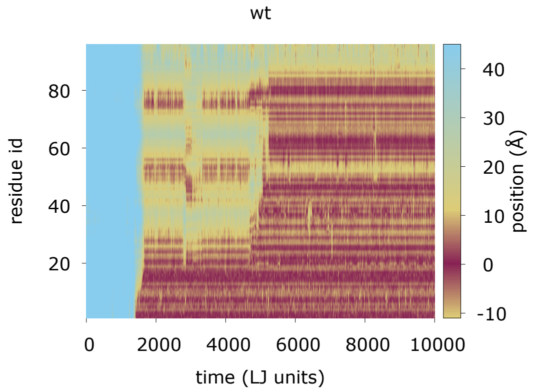


Figure 4.13 *Positions relative to the interface of the center of mass of each residue as a function of time from a coarse-grained simulation run of wild type Rsn-2. Other simulation runs display identical behaviours.*

In the all-atom simulations, the folded protein can interact with the interface in two different orientations: one parallel (state $1_{\parallel}$), with enhanced interactions around residues V78 and P79, and one perpendicular (state $1_{\perp}$) to the interface, with enhanced interactions around residue L47. It would be interesting to know whether the coarse-grained simulations give the same results, or a different set of metastable orientations. In Fig. 4.13 I show the position relative to the interface of the centres of mass of each residue as a function of time for a single simulation run (other runs display identical behaviour). This plot further confirms that the first adsorption step is mainly characterised by the interaction with the N-terminal tail, and that the clamshell opening of the core widen the interaction to the rest of the protein. Furthermore, it also shows that when the protein is folded it can interact with the interface via two slightly different sets of additional contacts in the structured part of the protein: formed either by the region around residues V78 and P79 (state $1_{\parallel}$), or by the one near residue L47 (state $1_{\perp}$).

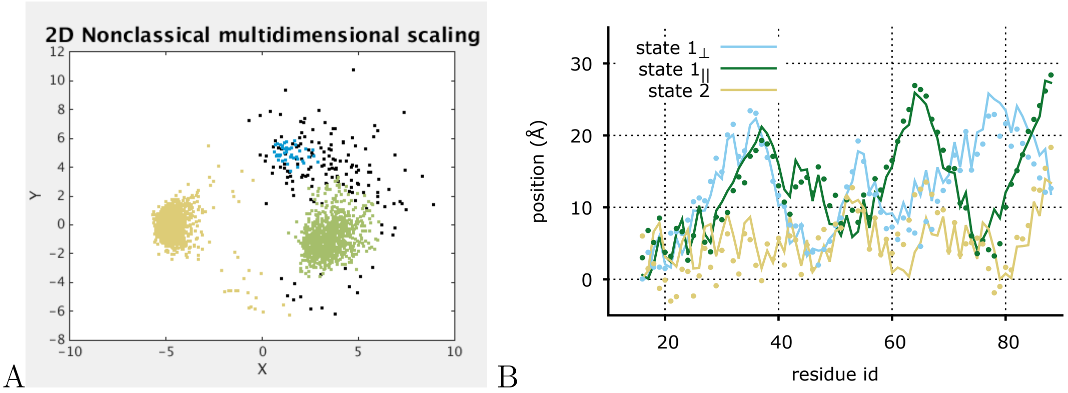


Figure 4.14 *Clustering of residue-interface separations observed in coarse-grained simulations of Rsn-2 adsorption. Three main clusters have been identified: state $1_{\perp}$ (blue), state $1_{\parallel}$ (green) and state 2 (yellow). (A) The configurations are plotted as a function of two variables defined via non-classical multidimensional scaling (produced with MATLAB [132]). (B) Profiles of the residue-interface separations corresponding to the three cluster centres (continuous lines) and comparison with the corresponding results from all-atom simulations (filled circles).*

I performed a clustering of the residue-interface separations observed during all the performed coarse-grained simulation runs of wt Rsn-2, using the same algorithm employed for the all-atom simulations. This analysis highlights exactly the same three clusters observed in the all-atom simulations, and the residue-interface separations corresponding to the cluster centres are in quantitative agreement with those previously found (Fig. 4.14). While the protein is folded and adsorbed at the interface, there are many jumps between the two orientations, and therefore there is no need to study this transition using metadynamics: state $1_{\parallel}$ is ~10 times more populated than state $1_{\perp}$, implying a free energy difference of ~2.3 k_BT between the two states.

Adsorption of Rsn-2 mutants

The study of protein mutants is important to test our understanding of the adsorption mechanism against experimental data. In order to affect the two adsorption steps of Rsn-2 independently, I considered two mutants: one where I deleted the N-terminal tail up to residue P15 (d1-15 mutant), which was also studied via all-atom simulations, and one where I added two extra disulphide bridges between residue ids 19 and 46 and between ids 32 and 81 (2C-C). The aim of the first mutation is to decrease the affinity of the protein for the interface during the first step of adsorption, whereas the second one is supposed to improve the stability of the protein hydrophobic core and therefore prevent the unhinging transition.

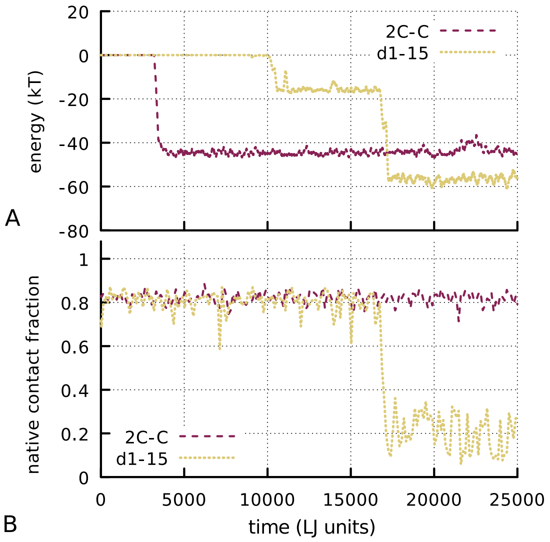


Figure 4.15 *Adsorption of d1-15 and 2C-C Rsn-2 mutants observed in the coarse-grained simulations. (A) Interaction energy between the protein and the interface for d1-15 Rsn-2 (dotted yellow line) and 2C-C Rsn-2 (dashed red line). (B) Fraction of native contacts corresponding to the hydrophobic core of the protein for d1-15 Rsn-2 (dotted yellow line) and 2C-C Rsn-2 (dashed red line).*

In Fig. 4.15 I show the protein-interface interaction energy and the fraction of hydrophobic core native contacts during the spontaneous adsorption of the mutants. Similarly to what observed in the all-atom simulations, despite the N-terminal deletion, the d1-15 mutant is still able to adsorb at the interface, but the interaction energy with the interface in the first adsorption state is reduced compared to the wild type from ∼-45 k_BT to ∼-16 k_BT. The deletion of the tail does not affect the clamshell mechanism of the protein, highlighted by the loss of hydrophobic core native contacts in Fig. 4.15B. As expected, the addition of the two disulphide bonds prevents the unhinging transition at the interface, and it does not affect the first adsorption step. The patterns of residue-interface separations for the two mutants (Fig. 4.16) are very similar to those observed for the wild type protein. In particular, when the protein is folded, there are even in this case two distinct patters of interactions: one with enhanced interactions around residues V78 and P79 (state $1_{\parallel}$, the most populated state) and another with enhanced interactions around residue L47 (state $1_{\perp}$, see for instance the run

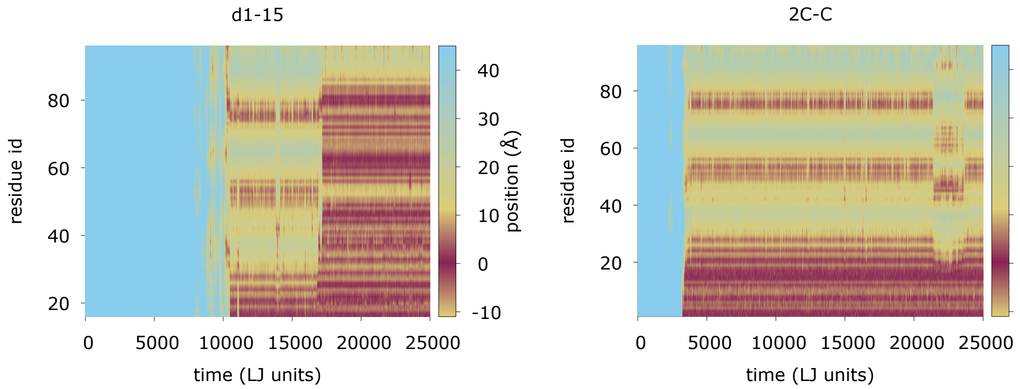


Figure 4.16 *Positions relative to the interface of the center of mass of each residue as a function of time from a coarse-grained simulation run of d1-15 (left) and 2C-C (right) Rsn-2. Other simulation runs display identical behaviours.*

of 2C-C from t=22000 to t=24000 LJ time units).

Free energy landscape of Rsn-2 adsorption

In order to offer a clear understanding of the adsorption mechanism of Rsn-2 and the effect of mutations, I computed the free energy landscapes of the two adsorption steps independently, using the well-tempered metadynamics method [75].

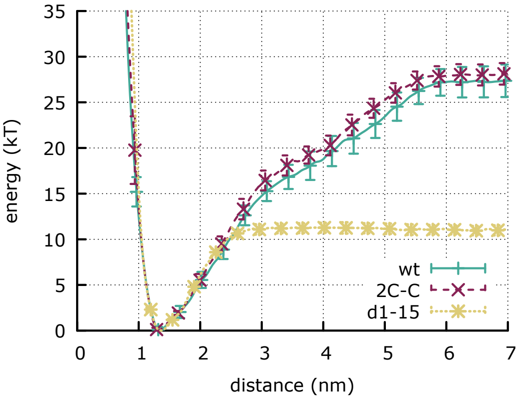


Figure 4.17 *Free energy profile of the first adsorption step of Rsn-2 from coarse-grained simulations computed via matadynamics calculations. The potential of mean force is given as a function of the distance between the center of mass of Rsn-2*$_{16-96}$ *and the interface. I compare the profiles of wt (green line), d1-15 (dotted yellow line) and 2C-C (dashed brown line) variants of Rsn-2.*

In Fig. 4.17 I report the potential of mean force as a function of the distance of the protein from the interface ($d_{\text{interface}}$) while the protein is kept folded in its native state. The plot shows that for wt Rsn-2 the free energy of the first adsorption state is very large ($\sim 27k_BT$). The 2C-C mutation does not affect the potential of mean force, whereas the d1-15 tail deletion decreases the free energy of adsorption by a factor of $\sim$2. Moreover, the deletion of the tail reduces the interaction range of the profile from 4.5 nm down to 1.5 nm, suggesting the importance of the N-terminal tail to "capture" the interface efficiently.

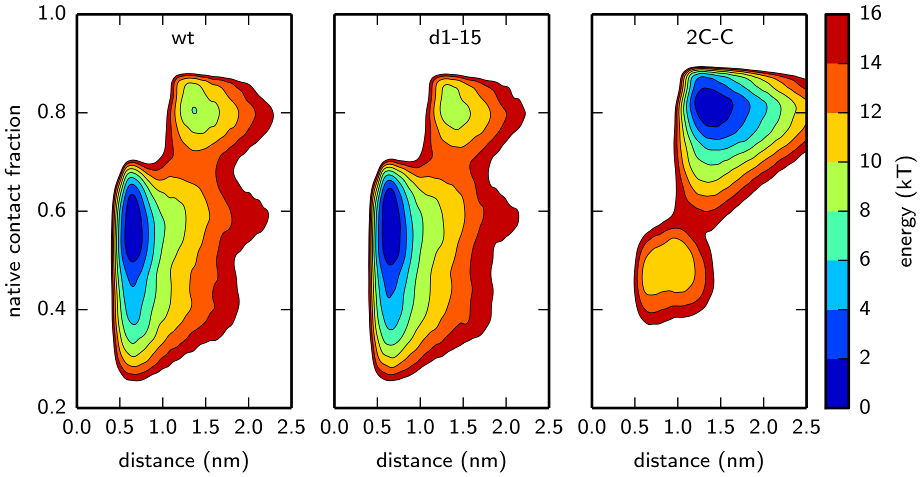


Figure 4.18 *Free energy profiles of the second adsorption step of Rsn-2 from coarse-grained simulations computed via matadynamics calculations. The free energy surface is given as a function of the distance between the center of mass of Rsn-2_{16-96} and the interface, and the fraction of native contacts formed by the protein. I compare the profiles of wt (left), d1-15 (centre) and 2C-C (right) variants of Rsn-2.*

In Fig. 4.17 I show the free energy landscape relative to the second step of the adsorption as a function of the distance from the interface and the fraction of the protein native contacts, for each of the considered protein variants. The landscape of the wild type protein is characterised by a local minimum corresponding to the folded protein adsorbed at the interface (state 1), and by a global minimum corresponding to the unhinged, partially unfolded, conformation of Rsn-2 (state 2). The free energy barrier to go from state 1 to state 2 is $\sim 6k_BT$, which is low enough to observe the transition in the unbiased simulations; but the barrier in the opposite direction is $\sim 13k_BT$, which causes the protein to be trapped in its open conformation. As suggested from the analysis of the spontaneous adsorption, the deletion of the N-terminal tail does not modify the free energy landscape of the unhinging transition. On the other hand, the addition of the disulphide bonds introduces a very high free energy barrier to the unfolding of the protein (on the order ∼15 k_BT); now the unfolded state of the protein is only a local minimum in the free energy landscape, and the conformation of this local minimum is also rather different from the unhinged state observed in wt and d1-15 Rsn-2.

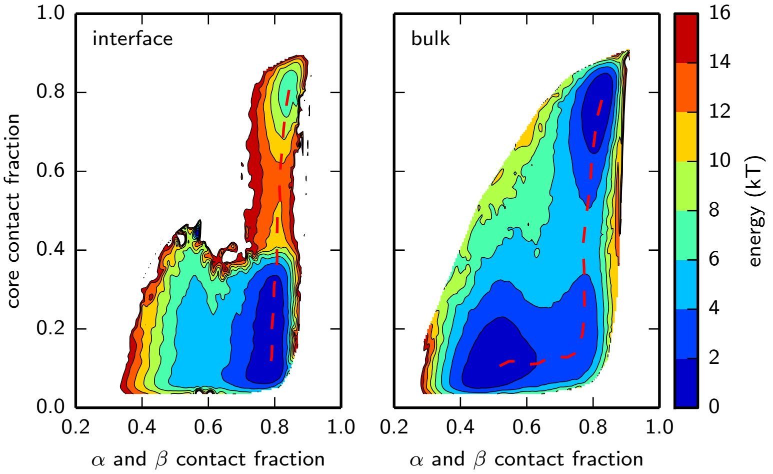


Figure 4.19 *Comparison between the free energy landscape of the unfolding of Rsn-2 at the interface at ambient temperature and in bulk at the transition temperature from coarse-grained simulations. The free energies are given as a function of the fraction of native contacts corresponding to the secondary structure elements (α-helix and β-sheet) and to the hydrophobic core of the protein. The dashed red line represents the minimum energy path between the two local free energy minima of the landscape, and it represents the average pathway between the folded and unfolded states of the protein.*

Finally, I reweighted the free energy landscape of the second adsorption step of wt Rsn-2 as a function of the fraction of secondary structure native contacts and hydrophobic core native contacts, and I compared it to the folding landscape of the protein in bulk at the transition temperature T_f (Fig. 4.19). It is interesting to notice that the unhinging transition at the interface proceeds via the same pathway as the unfolding in bulk, with the difference that in bulk the unfolding corresponds to a loss of both hydrophobic and secondary structure contacts, whereas at the interface both the α-helix and the β-sheet are fully preserved. Therefore, the unhinged conformation observed at the interface can be considered as an arrested state on the unfolding pathway of the protein in bulk, and the perturbation to the protein free energy landscape introduced by the interface is similar, but not exactly the same as an increase in temperature.

4.4 Comparison between all-atom simulations, coarse-grained simulations and experiments

Both all-atom and coarse-grained simulations reveal the same adsorption mechanism of Rsn-2 at interfaces, characterised by a first adsorption step where the protein captures the interface via the hydrophobic residues located on its disordered N-terminal tail, and a second step where the protein undergoes a unhinging conformational change where it opens its hydrophobic core towards the interface while maintaining its secondary structure elements intact. This mechanism is in perfect agreement with the experimental results obtained by Dr. Ryan Morris at the University of Edinburgh. In Fig. 4.20 I report the change in surface tension observed in pendant drop experiments during the adsorption of the protein at an air/water interface. In this type of experiments, the profile of the surface tension as a function of time is characterised by three regimes: regime I, where the there is no significant change in surface tension compared to the reference value of the clean interface, regime II, where there is a steep change in surface tension, and regime III, where the system is essentially in equilibrium and there is only a mild change in surface tension [9]. Regime I consists of a lag-time required for the proteins to reach a significant proportion of the maximum surface coverage (which is usually between 50% and 99% depending on the protein), whereas regime II is characterised by a combination of protein conformational changes, protein-protein interactions and further adsorption. The length of regime I for the adsorption of wild type Rsn-2 is in agreement the with the simple diffusion-limited model of adsorption by Ward and Tordai [13], confirming that the N-terminal tail of the protein provides an efficient mechanism to capture the interface and strongly bind to it. On the other hand, experiments show that the d1-15 N-terminal deletion significantly slows down the first stages of adsorption (Fig. 4.20). This is consistent with the finding that this mutant has a lower affinity for the interface and a shorter interaction range, causing both a decrease in the rate of adsorption and an increase in back-diffusion of proteins into bulk. The 2C-C mutation does not affect the length of regime I compared to the wild type protein, but it slows down the rate of change in surface tension observed during Regime II. This is because the 2C-C mutant is trapped at the interface in its first adsorption state, and since the protein is not able to increase its molecular area via the unhinging transition, more proteins have to adsorb at the interface to reach the maximum surface coverage. This shows that the clamshell mechanism is exploited by Rsn-2 to rapidly cover the available interfacial area using a smaller amount of proteins, and therefore improve its efficiency as a surfactant.

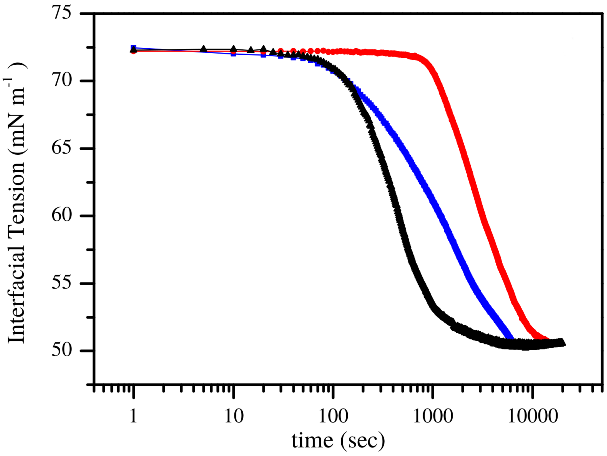


Figure 4.20 *Surface tension during pendant drop experiments of Rsn-2 adsorption; comparison between wt (black), 2C-C (blue) and d1-15 (red) protein variants.*

The increase in the area of the protein at the interface A_{mol} following the unhinging conformational change was further confirmed experimentally. The calculation of the area per molecule in Rsn-2 monolayers using a Langmuir trough gives ∼14 Å^2 for the wild type protein and ∼7 Å^2 for the 2C-C mutant. For the wild type, the estimates of the molecular area from the simulations are in good agreement with the experiments: ∼14.9 Å^2 from the all-atom simulations (Table 4.1) and ∼13.3 from the structure-based model. For the closed conformation of the adsorbed protein, corresponding to wt Rsn-2 in state 1 or to the 2C-C mutant at equilibrium, the estimates of A_{mol} are slightly higher than the one from the experiments: ∼10.9 Å^2 from the all-atom simulations (Table 4.1) and ∼9.9 from the structure-based model. Nevertheless, all-atom and coarse-grained simulations are consistent with each other and they are both able to capture the experimental observations.

Most experiments were performed using an air/water interface, whereas in the simulations the protein adsorbs at a cyclohexane/water interface. Dr. David Cheung, who collaborated with me on this project, performed all-atom simulations of Rsn-2 at air/water interfaces. His simulations confirmed the importance of the N-terminal tail in capturing the interface in the first adsorption step. However, differently from my observations, the N-terminal deletion completely prevents the adsorption of the protein at the interface, and even for the wild type protein no unhinging events were observed within 200 ns. Both differences are likely to be due to the lower affinities of hydrophobic residues for the air phase compared to the oil phase [9, 61]. At the air/water interface unhinging should still occur (as suggested by experiments); however, given the decreased core-interface interaction, the energetic barrier of this conformational change is probably too high to observe it during an all-atom simulation. Despite these differences, all simulations and experiments point towards a robust mechanism of adsorption that does not depend on the specific type of fluid/fluid interface.

One of the most surprising observation is the ability of the new structure-based coarse-grained model to describe quantitatively the patterns of residue-interface separations, highlighting exactly the same three clusters observed from the all-atom simulations. In particular, both simulations find that when the protein is in its closed form, it can interact with the interface via two alternative sets of contacts: state $1_\perp$, where the α-helix and the β-sheet are perpendicular to the interface, and the protein interacts via the N-terminal tail and residue L47; and state $1_\parallel$, where the secondary structure elements now lie parallelly to the interface, and the contact with residue L47 is substituted by those with residues V78 and P79. This shows that the new structure-based model represents a valid complement (or even an alternative) to all-atom simulations for the study of the interaction between proteins and interfaces quantitatively.

4.5 Conclusions

In this work I have used multi-scale MD simulations to study the surprising mechanism of surface activity employed by the protein Rsn-2. Both all-atom and coarse-grained simulations strongly suggest that the adsorption of Rsn-2 proceeds via a two-step mechanism, where the first contact with the interface is made by the flexible and hydrophobic-rich N-terminal tail without any rearrangement of the protein native fold, and then the protein increases its interfacial area via a conformational change where the α-helix and the β-sheet unhinge further apart. Consistently with circular dichroism experiments [3], the unhinging of the protein is not accompanied by a loss of secondary structure content. This is possible because both the α-helix and the β-sheet have a strong amphiphilic character, with clearly distinct polar and apolar sides, allowing these structural elements to optimally partition their hydrophobic residues into the oil phase while maintaining their native conformation.

In the literature there are other instances of proteins that adsorb at the interface via a rearrangement of the tertiary structure while preserving the secondary structure content. A notable example are the apolipoproteins, whose structure consists of a helix bundle of amphipathic α-helices [148]. These proteins bind to lipid membranes by reorienting their helices to bury the hydrophobic residues into the membrane [149]. In an analogous way to Rsn-2, where the hydrophobic core contacts are replaced with equivalent protein-oil interactions, apolipoprotein core contacts are replaced by equivalent helix-lipid interactions. Moreover, Rsn-2 shares significant structural similarities to cystatins [8], and it should not be surprising that the clamshell mechanism of Rsn-2 adsorption is reminiscent of the domain-swapping mechanism of cystatin dimer formation [150]. During domain-swapping, the α-helix of the first monomer unhinges from its β-sheet and closes on top of the β-sheet of the second monomer, and vice-versa; this operation does not involve a change in secondary structure content, but, similarly to Rsn-2 adsorption, simply a swap of hydrophobic interfaces. It would be interesting to investigate whether Rsn-2 evolved its surfactant properties directly from a domain-swapping mechanism.

Finally, I want to comment on the transferability of the new structure-based model. In its current form, the model can be used to study the adsorption of proteins at interfaces or lipid membranes in all cases where we expect unfolding with no formation of new secondary structure elements. This is because the model is based on the native conformation of the protein and it cannot be used to predict new residue-residue contacts. However, this model could be used in combination with multiple-basin energy landscapes [26] to study the interplay between protein adsorption and conformational changes between known conformations with different secondary structure content. A system where such an approach could be tested is BslA, where the adsorption of the protein induces a conformational change from the $\mathrm{L_{in}}$ to the $\mathrm{L_{out}}$ form of the cap (see chapter 3).

Chapter 5

Colloidal rods at an interface: flipper dynamics and bilayer formation

I describe molecular dynamics simulations of interfacial monolayers of rods under compression. The results reveal how aspect ratio affects ordering and rod flipping at the interface, comparing favourably with the experimental results.

5.1 Introduction

In the first results chapter we have seen how the adsorption of the protein BslA at interfaces is governed by the same physics that controls the adsorption of a Janus ellipsoidal colloid. For their simplicity, colloidal particles represent the ideal model system for the study of complex soft matter systems [34], with implications ranging from protein crystallisation [121] to protein monolayers at interfaces [4]. I performed coarse-grained molecular dynamics (MD) simulations to study the influence of aspect ratio (AR) on the behaviour of interfacial monolayers formed by rod-like colloids under compression. In doing so, I slightly diverge from the main topic of proteins at interfaces, but I stress that a monolayer of anisotropically-shaped colloids represents a model system for the study of protein fibrils assembled at interfaces [16, 151], and these findings may be relevant for future research on proteins.

The adsorption of colloids with high aspect ratio at interfaces has firstly relevance for the theoretical study of the isotropic-nematic phase transition in two dimensions [151, 152], but also for practical applications: shape anisotropy enhances the stability of Pickering emulsions compared to those formed by spherical particles [153, 154] and it enable the emergence of rich self-assembly behaviours [42–44, 155]. A significant amount of research has been devoted to study the self-assembly of cylindrical and ellipsoidal particles at liquid interfaces [43, 44]. Because of their shape, the contact line of these particles at the interface is no longer flat (as for spheres), and the deformation of the interface induces capillary interactions that drive a pair of cylinders into a end-to-end configuration and a pair of ellipsoids into a side-to-side one [43].

Experimentally, particle-laden interfaces are studied under compression in a Langmuir trough to infer the properties of more complex systems such as Pickering emulsions [39, 154, 156]. Experiments on monolayers of ellipsoidal colloids showed an interesting behaviour as the area of the through is decreased: particles can release the compression stress by flipping from their favourable orientation parallel to the interface to a perpendicular orientation [156]. However, the precise features of this flipping behaviours are not yet very clear. In this chapter, I consider an interfacial monolayer formed by rod-like particles with neutral wetting. Similarly to spherical particles, the contact line of spherocylinders is flat, and there are no capillary interactions. Because of the lack of attraction, which may interfere with the aforementioned flipping dynamics, this represents a perfect model system for the study of film compression. To this aim, I used coarse-grained MD simulations to characterise quantitatively how the behaviour of the film is affected by the aspect ratio of the spherocylinders, with particular emphasis on the flipping dynamics and the formation of bilayers. In these simulations, the rods are initially adsorbed at the interface between two immiscible fluids (representing water and oil), the interface is then compressed, forcing the rods to change their orientation, desorb or form bilayers. For simplicity, through the chapter I will refer to the spherocylinders as rod-like particles, or simply rods. In the discussion, I will compare the computational results with the experimental observations obtained by Li Tao, PhD student under the supervision of Dr. Paul Clegg at the University of Edinburgh. MD simulations enable us to capture microscopic details that cannot be observed in the experiments, due to the limitations in the resolution of the optical microscope.

5.2 Methods

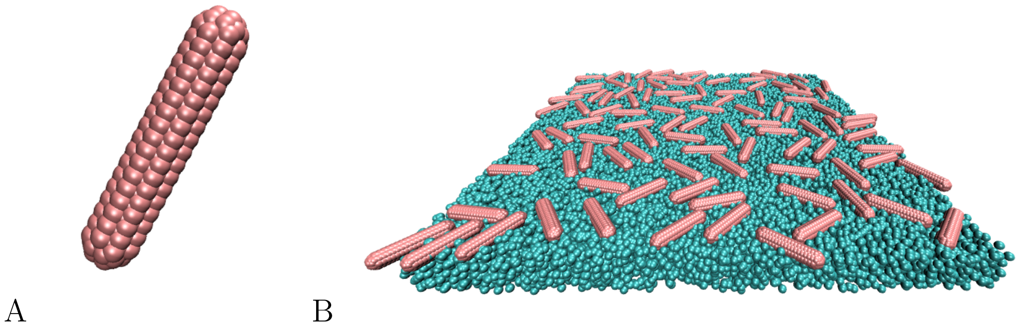


Figure 5.1 *(A) Geometry of the individual colloidal rod with aspect ratio aspect ratio (AR) 6, formed by coarse-grained beads arranged in a fcc lattice. (B) Initial configuration of the system for AR=6; with the colloids in pink and the water fluid phase in cyan. The oil phase is omitted for the sake of clarity.*

The MD simulations of adsorbed colloidal rods have been performed using the dissipative particle dynamics (DPD) model [157, 158]. This computational strategy has been motivated by the fact that DPD has been already successfully applied to the study of colloidal particles at fluid/fluid interfaces [159–163]. In principle, Lattice-Boltzmann simulations [164] could have been a valuable alternative; however, we are not aware of any Lattice-Boltzmann software that includes spherocylindrical colloids, whereas this can be done straightforwardly with any molecular dynamics software that implements DPD (for instance LAMMPS [145], see below). In DPD, the fluid is represented by a set of point-like beads, each of them representing a meso-scale fluid element. Each pair of beads interact via the sum of conservative (F^C), dissipative (F^D) and random (F^R) forces [158]:

$$\begin{aligned} \vec{F} &= (F^C_{ij} + F^D_{ij} + F^R_{ij})\hat{r}_{ij} \quad \text{for} \quad r_{ij} < r_c \\ F^C_{ij} &= \chi_{ij} w(r_{ij}) \\ F^D_{ij} &= -\Gamma w^2(r_{ij})(\hat{r}_{ij} \cdot \vec{v}_{ij}) \\ F^R_{ij} &= \sigma w(r_{ij})\alpha/\sqrt{\Delta t} \\ w(r) &= 1 - r/r_c \end{aligned} \tag{5.1}$$

where r_c is the interaction cutoff, r_{ij} is the distance between beads i and j, $\hat{r}_{ij}$ is the unit vector from j to i, and $\vec{v}_{ij}$ is the velocity of i relative to j. $w(r_{ij})$ is a weighting function that increases linearly from 0 at $r_{ij} = r_c$ to 1 at $r_{ij} = 0$, so that the conservative force is soft and purely repulsive. χ_{ij} is a positive parameter that sets the strength of the repulsion, α is a Gaussian random number with zero mean and unit variance, and Δt is the time-step of the MD integration. Γ and σ are parameters related via the fluctuation-dissipation theorem by $\sigma^2 = 2k_BT\Gamma$, where k_B is the Boltzman constant and T is the temperature [157]. To simulate colloidal particles, I use the approach described in Refs. [159–161, 165], where each colloid is represented as a rigid body built by fluid beads arranged in a face-centred-cubic (fcc) lattice with the desired shape, using an fcc lattice constant of $0.35r_c$.

In the following, I will make use of the units: r_c for space, k_BT for energy, m for mass and $\tau = \sqrt{mr_c^2/(k_BT)}$ for time. In the simulations I set $\Gamma = 4.5m/\tau$ and $\Delta t = 0.01\tau$. The number density of the water and oil phases is kept constant at $\rho = 3/r_c^3$. In the notation for the parameters, the suffixes W, O and C refer respectively to water, oil and colloid beads. The mass of the fluid beads is set to $m_W = m_O = 1m$, whereas the mass of the colloid beads is set to $m_C = 0.1m$; since the number density of the beads inside the colloids is higher than the one of the fluid phases, using the same mass would give the colloids high inertia and this would slow down the relaxation of the system under compression. The strength of the repulsion depends on the types of the interacting beads and I employed the same settings used by Hore and Laradji in Ref. [160]: $\chi_{WW} = \chi_{OO} = \chi_{CC} = 25k_BT/r_c$, $\chi_{WO} = 100k_BT/r_c$, and $\chi_{CW} = \chi_{CO} = 9.2k_BT/r_c$. The considered colloids are spherocylinders with a constant diameter of $D_C = 1.08r_c$ and varying aspect ratios of AR=3, 6, 9 and 15. The equations of motion are integrated using a modified version of the Velocity-Verlet algorithm [158, 166]. All simulations are performed with the program LAMMPS [145].

The parameters for the individual fluid phases are set so that the fluid has the same compressibility of water at ambient temperature [158], and the higher repulsive interaction between water and oil ensures phase separation. The free energy penalty for the creation of the interface between the two fluids is equal to the product of the surface tension γ and the area of the interface between the two fluids. I computed the surface tension [167] of the water-oil interface from an equilibrium DPD simulation without colloids, finding a value of γ=4.2 k_BT/r_c^2. Since the interaction between any of the two fluids and the colloidal particles is the same, the colloids do not have any preference to stay in one of the two fluid phases, i.e. the colloids have a contact angle of 90 degrees. When a colloid is adsorbed, it reduces the area of the interface between the two fluids, lowering the free energy of the system. In order to compare favourably with the experiments, the rods have to be able to adsorb at the interface irreversibly in the absence of external forces. The free energy of adsorption of a colloid can be estimated as the product of the surface tension and the maximum area of interface covered by the colloid (which is achieved when the rod lies parallel to the interface):

$$\Delta G_{\mathrm{ads}} = -E_I = \gamma A_{\mathrm{rod}} \tag{5.2}$$

For rods of aspect ratios AR=3, 6, 9 and 15 I find A_{rod}=6.0, 11.9, 16.6 and 28.4 r_c^2, and ΔG_{ads}=25.4, 50.1, 69.8 and 119.2 $k_B T$ respectively; therefore adsorption is expected to be irreversible in the absence of compression (within the timescales considered in the simulations). The area A_{rod} has been computed considering the effective hard-core size of the rod, estimated from the minimum distance between the surfaces of two rods under compression; this increases both colloid diameter and length by 0.6 r_c.

For each aspect ratio, I compress along the $\hat{x}$ and $\hat{y}$ directions a system with water and oil phase-separated along the $\hat{z}$ direction, and with $N_{\mathrm{start}} = 100$ rods initially adsorbed at one of the two interfaces (I use periodic boundary conditions). The 100 rods are placed at time $t = 0$ on the interface at random but non-overlapping orientations, with their centres of mass on the plane $z = 0$, and with their axes parallel to the interface. The initial box side along $\hat{z}$ is set to L_z=10 r_c for all ARs, whereas along $\hat{x}$ and $\hat{y}$ the box sides ($L_x = L_y$) are set so that the initial surface coverage of the interface is $\theta_0 = N_{\mathrm{start}} A_{\mathrm{rod}}/(L_x L_y) = 0.3$. Before starting the compression, the system is equilibrated for a time $t_{\mathrm{eq}} = 3000\tau$, corresponding to the typical reorientation time of the longest rod, which is estimated as the typical time required to diffuse over and angle of 90 degrees. I compress the box along $\hat{x}$ and $\hat{y}$ using a constant strain rate with a velocity equal to v_{comp}=0.0003 r_c/τ, which is low enough to allow every rod to diffuse and reorient during the compression. The box is also expanded along $\hat{z}$ to keep the volume, and therefore the density, constant (this is a requirement of DPD simulations). The compression proceeds until I observe either a significant desorption of rods from the interface or the formation of a bilayer. The simulations are computationally very expensive, and for each aspect ratio, I report data relative to a single individual compression. In the analysis of the simulations, I define a rod to be adsorbed at the interface if the distance between the rod and the interface is lower than the cutoff distance d_{ads}=2.2 r_c, where the position of the interface is estimated as the average position along $\hat{z}$ of the adsorbed rods.

In order to characterise the ordering of the rods at the interface I define a convenient quantity that measures the formation of a contact C between two rods:

$$C = \frac{1-(d_{\mathrm{cm}}/d_0)^6}{1-(d_{\mathrm{cm}}/d_0)^{12}} \frac{1-((1-\cos\alpha)/(1-\cos\alpha_0))^6}{1-((1-\cos\alpha)/(1-\cos\alpha_0))^{12}} \tag{5.3}$$

where d_{cm} is the distance between the rods centres of mass, α is the angle between their orientations, and d_0 and α_0 are parameters set to d_0=3 r_c and α_0=30 degrees. This function is close to one only when $d_{\mathrm{com}} \lesssim d_0$ and $\alpha \lesssim \alpha_0$, i.e. only when the two rods are close to each other and have the same orientation, and it gives a value close to zero otherwise. The value of C is then used for clustering the rods at the interface (see Fig. 5.4): two rods i and j are in the same ordered domain only if $C_{ij} > 0.55$. For the clustering it would be sufficient (and simpler) to define distance and angle cutoffs below which two rods are assigned to the same cluster. However, the smoothness of the contacts C defined above is convenient for the analysis of the locations at which desorption occurs, that I will describe in the results section (Fig. 5.9).

5.3 Results

In order to compare the behaviour of the system under compression as the aspect ratio is changed from AR=3 to AR=15, I define the quantity $\theta_{\mathrm{max}} = N_{\mathrm{start}} A_{\mathrm{rod}}/(L_x L_y)$, where $N_{\mathrm{start}} = 100$ is the number of adsorbed rods at the beginning of the simulation, A_{rod} is the area occupied by a single rod when it lies parallel to the interface, and $L_x L_y$ is the area of the interface between water and oil. θ_{max} is a convenient coordinate to describe the progression of the compression; regardless of the considered rod, for $\theta_{\mathrm{max}} \lesssim 1$ the 100 rods are able to stay adsorbed at the interface and keep their optimal parallel orientation, whereas for $\theta_{\mathrm{max}} \gtrsim 1$ the rods have to rearrange in at least on of the following ways: desorb from the interface, tilt out of the plane of the interface (flipping) or form multiple layers. Fig. 5.2A shows that indeed, as θ_{max} approaches the value of 1, rods start to desorb from the interface. Desorption starts earlier for the rods of aspect ratio AR=3, which have the lowest free energy of adsorption, and it is not observed within the considered time-scale for AR=15. In Fig. 5.2B I plot the effective surface coverage $\theta = N_{\mathrm{ads}} A_{\mathrm{rod}}/(L_x L_y)$, defined as the surface coverage that would be achieved by the rods adsorbed at the interface when

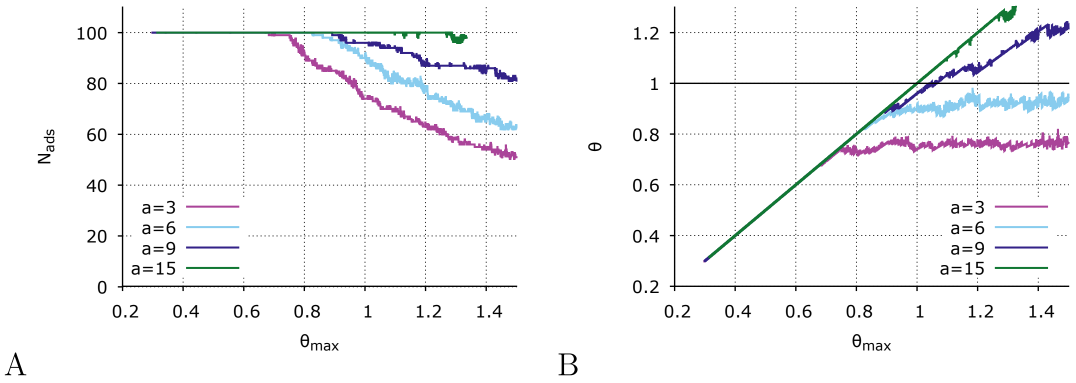


Figure 5.2 *(A) Number of rods adsorbed at the interface* N_{ads} *and (B) surface coverage* θ *as a function of the maximum surface coverage* θ_{max} *for the different aspect ratios during a single compression simulation. The maximum surface coverage is defined as* $\theta_{max} = N_{start}A_{rod}/(L_x L_y)$*, whereas the surface coverage is defined as* $\theta = N_{ads}A_{rod}/(L_x L_y)$*. The definition of* θ *assumes that the rods lie parallel to the interface, which is a good approximation (see snapshots from Figs. 5.5,5.6,5.7 and 5.8); a value of* θ *greater than one signals the formation of a bilayer, which occurs for AR=9 and AR=15.*

they lie parallel to the interface. We can notice that only for AR=3 and AR=6 θ reaches an equilibrium value lower than 1, whereas for AR=9 and AR=15 θ becomes higher than 1. Since most of the adsorbed rods keep their parallel orientation (see Fig. 5.3), $\theta > 1$ signals the formation of a multiple layers of rods at the interface, which can be observed towards the end of the compression in the simulation snapshots reported in Figs. 5.7 and 5.8 and is discussed later on.

Another characteristic feature of the system is the formation of distinct ordered domains under compression. Fig. 5.4 shows the average size of the domains (or clusters) formed by the rods adsorbed at the interface. For AR=3 there is no evidence for clustering, whereas at higher aspect ratios clusters start to form from $\theta_{\text{max}} = 0.6$. In Figs. 5.5, 5.7, 5.7 and 5.8 I show several simulation snapshots of the four considered rods under compression. Together with the former quantitative analysis, these snapshots allow to compare the different stages of compression for the considered aspect ratios.

- (AR=3) During compression rods start to desorb from the interface around $\theta_{\text{max}} = 0.7$ and there is no domain formation. At all times there is no overlap of rods at the interface, i.e. no bilayer formation.

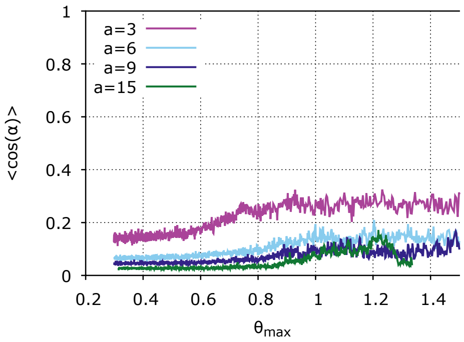


Figure 5.3 *The cosine of the angle* α *between the orientation of the rod and the interface normal* $\hat{z}$ *averaged over the rods adsorbed at the interface, as a function of the maximum surface coverage* $\theta_{max} = N_{start}A_{rod}/(L_xL_y)$. *No large-scale tilting transition is observed under compression.*

- (AR=6) Rods start to desorb from $\theta_{\mathrm{max}} = 0.8$ and, similarly to what observed for AR=9 and AR=15, domain formation starts around $\theta_{\mathrm{max}} = 0.6$. There is no formation of a bilayer. After the start of desorption, we can notice from the snapshots that on the interface there are some rods perpendicular to the interface localised at the boundary between ordered domains. This perpendicular orientation usually anticipates the complete desorption of the rod.

- (AR=9) Desorption starts from $\theta_{\mathrm{max}} = 0.9$. However, from around $\theta_{\mathrm{max}} = 1.1$ we can observe from the snapshots the overlap between the rods at the interface, i.e. the formation of a bilayer. Here perpendicular rods are also occasionally at the boundary between domains, but these events are much fewer than those found for AR=6.

- (AR=15) Rod desorption does not occur; the system responds to compression by forming immediately multiple layer domains from $\theta_{\mathrm{max}} = 1$.

I wanted to characterise more carefully the relationship between rod desorption and ordered domains. From the visual analysis of the snapshots of the simulations for AR=6 and AR=9 there is evidence that adsorption occurs via a tilting of the rod perpendicularly to the interface, and that this flipping occurs at the boundary between other ordered domains, or equivalently in a locally disordered region. In order to test this hypothesis quantitatively, I consider the following scenario: considering a rod i that desorbs from the interface at a certain time t_i, I would like to know if, at the certain time in the past $t_i - t$, rod i was located within an

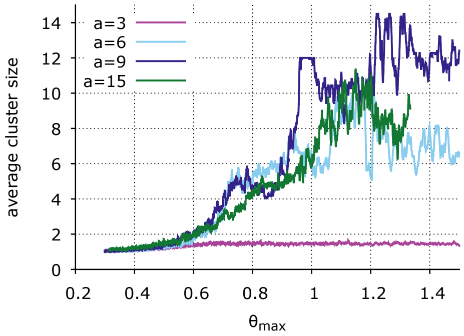


Figure 5.4 *Average cluster size for the different aspect ratios as a function of the maximum surface coverage* $\theta_{max} = N_{start}A_{rod}/(L_xL_y)$. *Fluctuations are high, especially towards the end of the compression, therefore I reported running averages over* $\Delta t_{ave} = 200\tau$. *For AR=6, 9 and 15, domain formation starts to occur around* $\theta_{max} = 0.6$, *but it does not occur for aspect ratio AR=3. See the methods section for the definition of a cluster.*

ordered or disordered region of the monolayer at the interface. To answer this question I can define the quantity:

$$\Delta c(t) = \frac{\langle C_i(t_i - t) - \langle C(t_i - t)\rangle_{\text{adsorbed}}\rangle_{\text{desorption events}}}{\sigma_{C_i - \langle C\rangle}} \tag{5.4}$$

$C_i(t) = \sum_j C_{ij}$ is the sum of the contact values defined in Eq. 5.3 between rod i and all other rods adsorbed at the interface. C_i will take a high value if rod i is located within an ordered domain, and a low value otherwise. The quantity $C_i(t_i - t) - \langle C(t_i - t)\rangle_{\text{adsorbed}}$ represents the difference between the order parameter C for rod i, and the average order parameter of the rods adsorbed at the interface at time $t_i - t$. The outer average runs over the rods that desorbed from the interface at any time during the compression (i.e. the desorption events); the result is also rescaled by the standard deviation. If rods desorb from the interface at completely random locations, Δc will be close to zero. On the other hand, if rods that desorb were on average located within an ordered region at a time t before the desorption, then $\Delta c(t)$ will be positive, and it will be negative otherwise. In Fig. 5.9 I plot the quantity Δc for AR=3, 6 and 9 (the low number of desorption events prevents the evaluation for AR=15). The results show that for aspect ratios AR=6 and AR=9, desorption is indeed correlated with disorder. The absence of correlation for AR=3 is probably due to the lack of ordering for this aspect ratio.

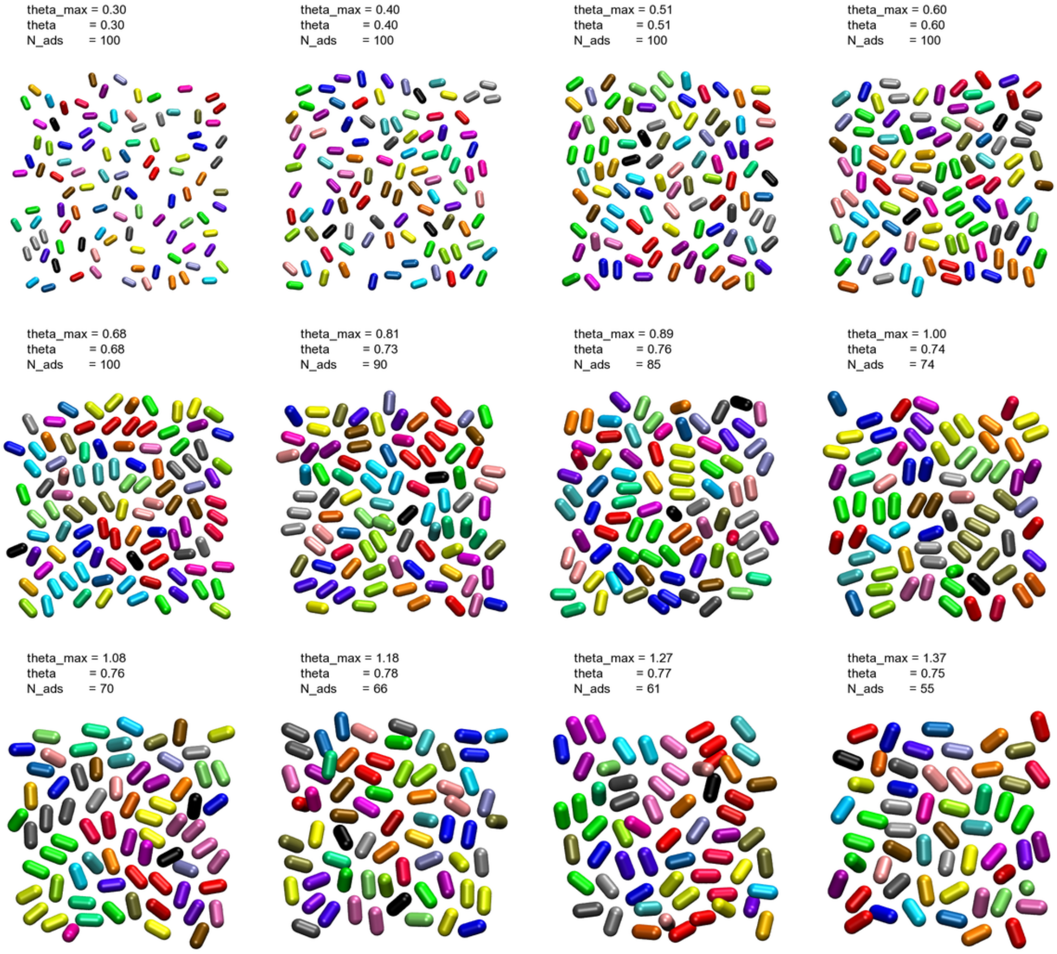


Figure 5.5 *Snapshots of the rods at the interface for rods of aspect ratio AR=3. Time increases from left to right and from top to bottom. For each frame I indicate: maximum surface coverage* $\theta_{max} = N_{start}A_{rod}/(L_xL_y)$, *surface coverage* $\theta = N_{ads}A_{rod}/(L_xL_y)$ *and number of adsorbed rods* N_{ads}. *Rods that belong to the same cluster are represented using the same colour.*

5.4 Comparison with experiments

Tao Li performed experiments using micron-sized rod-like particles with aspect ratios of 3, 9 and 15 (Fig. 5.10). The experiments were performed in a Langmuir trough by firstly depositing the rods on an air/water interface, and then by compressing the monolayer parallely to the interface using two moving barriers. The changes in the organisation of the rods in the monolayer were then recorded using an optical microscope.

The relatively large size gives these rods a free energy of adsorption significantly

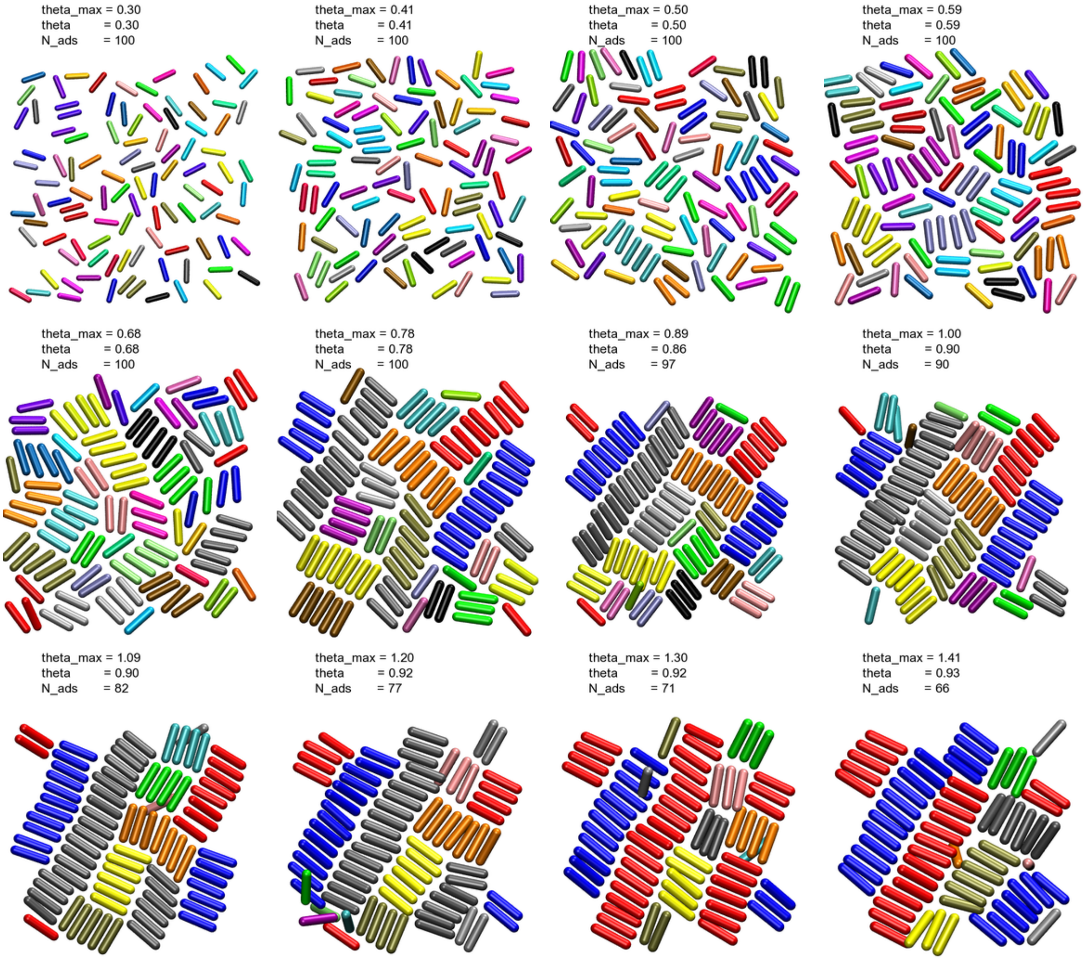


Figure 5.6 *Snapshots of the rods at the interface for rods of aspect ratio AR=6. Time increases from left to right and from top to bottom. For each frame I indicate: maximum surface coverage* $\theta_{max} = N_{start}A_{rod}/(L_xL_y)$, *surface coverage* $\theta = N_{ads}A_{rod}/(L_xL_y)$ *and number of adsorbed rods* N_{ads}. *Rods that belong to the same cluster are represented using the same colour.*

higher than that of the rods considered in the simulations; for this reason, flipping events are not followed by the complete desorption of the particle. Instead, at very large compression, film buckling occurs, highlighted by large regions of monolayer moving out of the focus of the optical microscope. For aspect ratio 3, the orientation of the rods cannot be captured by the images, but the buckling of the film is clear from the darker areas in the last snapshot in Fig. 5.11A. For AR=9, experiments highlight flipping events (Fig. 5.11B) similar to those found in the past experiments using ellipsoids [156] and in the simulations for aspect ratios 6 and 9 (Fig. 5.6). For AR=15, no flipping events were observed, and, as the area of the trough is decreased, some bilayer regions form (highlighted in

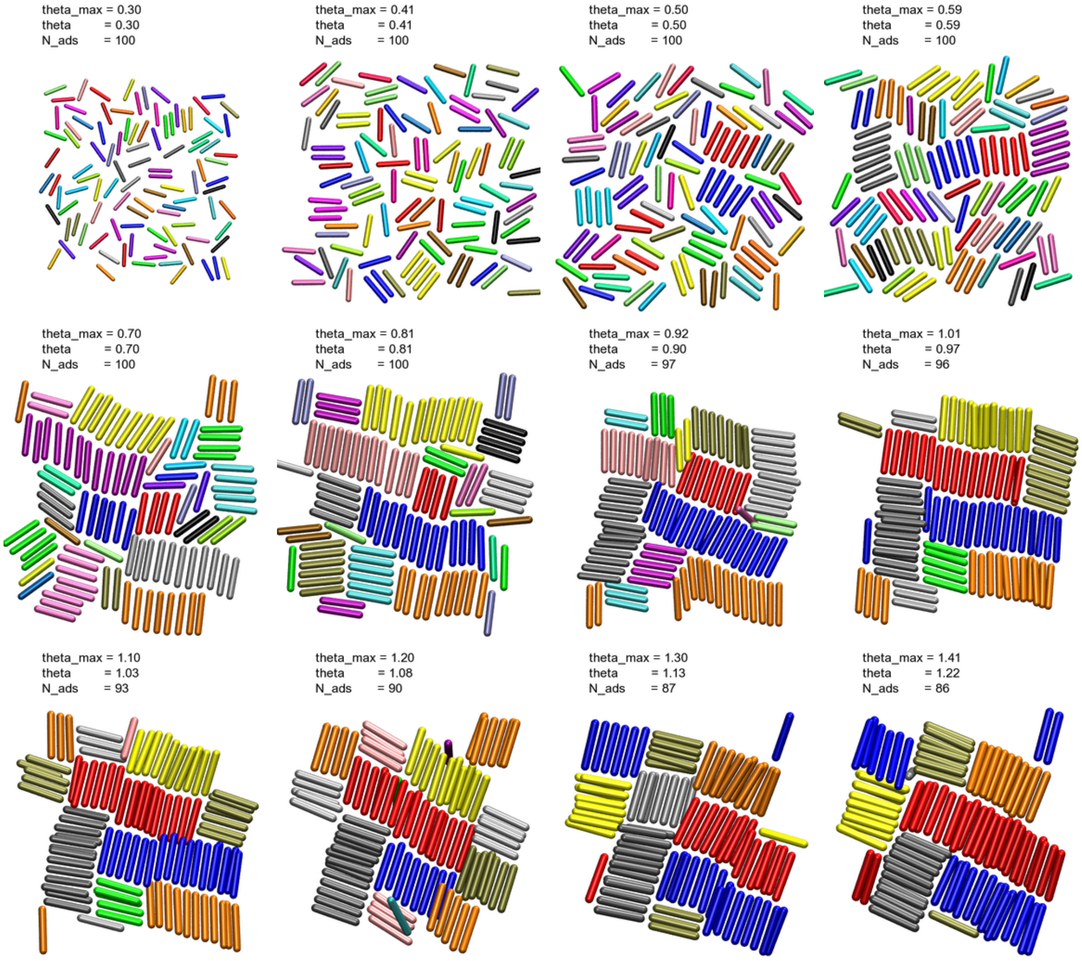


Figure 5.7 *Snapshots of the rods at the interface for rods of aspect ratio AR=9. Time increases from left to right and from top to bottom. For each frame I indicate: maximum surface coverage* $\theta_{max} = N_{start}A_{rod}/(L_xL_y)$*, surface coverage* $\theta = N_{ads}A_{rod}/(L_xL_y)$ *and number of adsorbed rods* N_{ads}*. Rods that belong to the same cluster are represented using the same colour.*

Fig. 5.11C). Overall, these observations are in good agreement with the trend found from MD simulations.

5.5 Conclusions

The simulations reveal two qualitatively different phenomena in compressed monolayers of rod-like particles with different aspect ratios. Firstly, the compression induces the formation of large clusters of ordered domains only for aspect ratios 6, 9 and 15, but not for aspect ratio 3. This behaviour may be

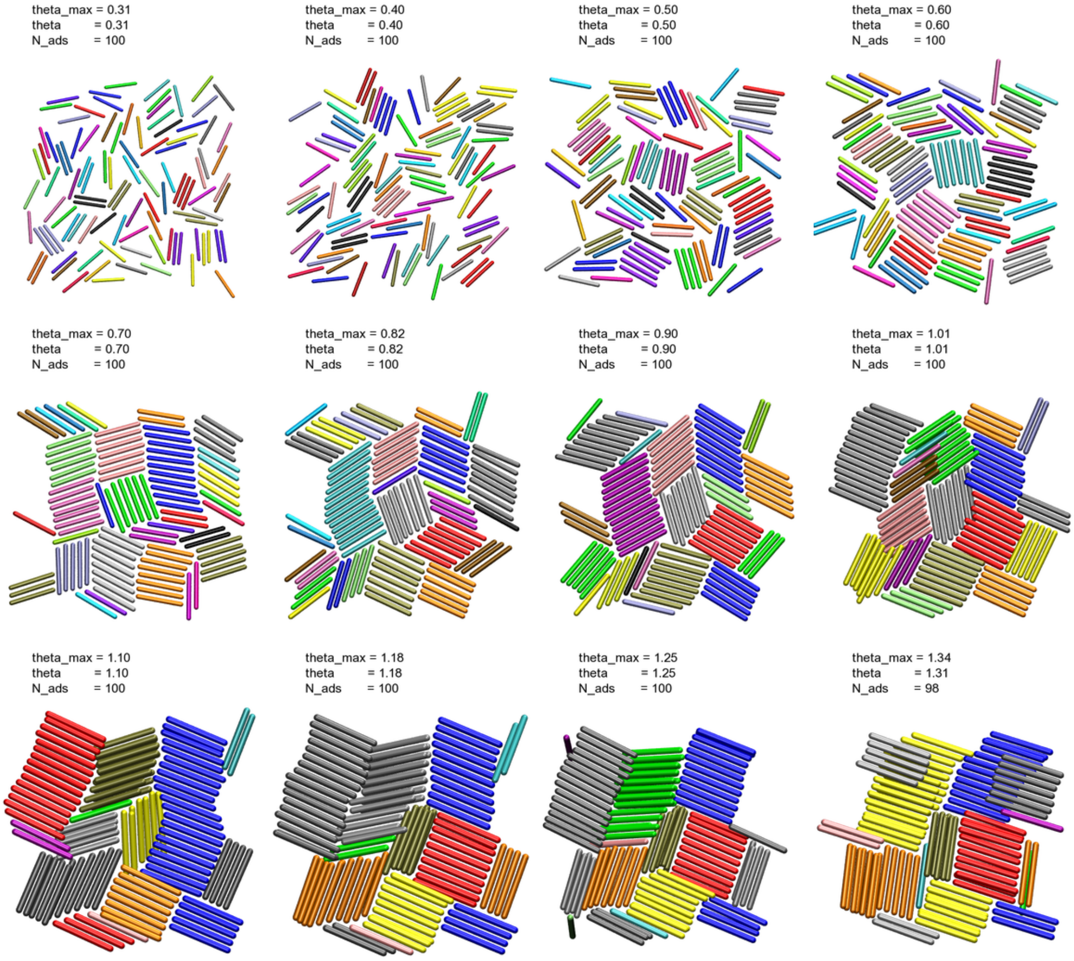


Figure 5.8 *Snapshots of the rods at the interface for rods of aspect ratio AR=15. Time increases from left to right and from top to bottom. For each frame I indicate: maximum surface coverage* $\theta_{max} = N_{start}A_{rod}/(L_xL_y)$*, surface coverage* $\theta = N_{ads}A_{rod}/(L_xL_y)$ *and number of adsorbed rods* N_{ads}*. Rods that belong to the same cluster are represented using the same colour.*

related to the isotropic-to-nematic transition observed at aspect ratio ∼7 for rod-like particles in two dimensions [152], even though the non-equilibrium character of my system prevents from making a direct comparison with the equilibrium results.

Secondly, the aspect ratio has a strong effect on the precise mechanism employed by the system to release the compression stress. At aspect ratio 6, flipping seems to be the main strategy used to reduce the area occupied by the rods at the interface, similar to what observed in monolayers of ellipsoidal particles [156]. However, at the highest aspect ratio considered (AR=15), there are no flippers; instead, the area of the adsorbed rods is reduced via the formation of bilayer

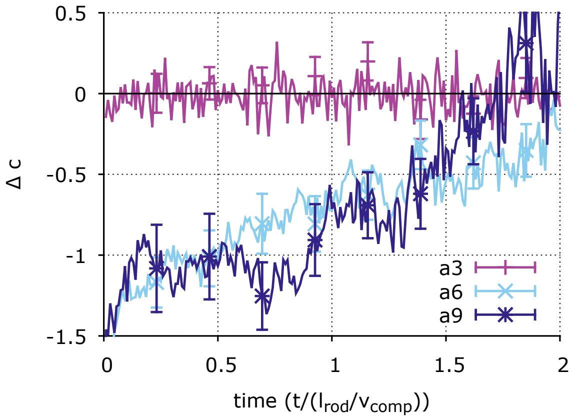


Figure 5.9 *The quantity* Δc *defined in Eq. 5.4 evaluated during compression for rods of aspect ratio AR=3, 6 and 9. The low number of desorption events prevents the evaluation for AR=15. Times have been rescaled by a characteristic timescale* l_{rod}/v_{comp}*, where* l_{rod} *is the length of the considered rod and* v_{comp} *is the velocity of the compression. For AR=6 and AR=9, desorption events are highly correlated with disorder.*

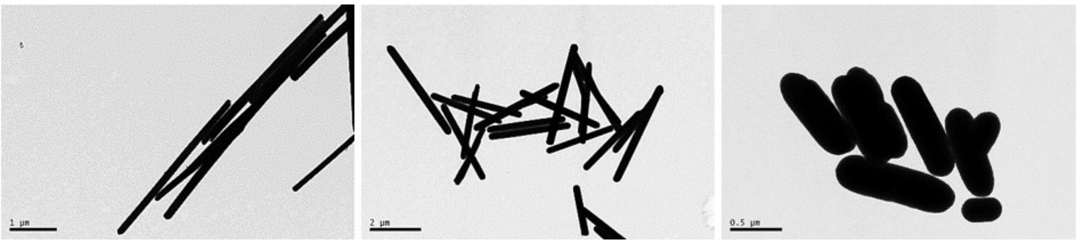


Figure 5.10 *TEM images of the rod-like silica particles used in the experiments, having aspect ratios 15 (left), 9 (centre) and 3 (right). The scales on the bottom-right corner are respectively 1, 2 and 0.5* μm*.*

domains. Aspect ratio 9 shows an intermediate behaviour characterised by flippers in the first stages of compression followed by the formation of bilayer domains. This transition could be possibly explained by the increased free energy of adsorption of the long rods, which makes the flipping kinetically unfavourable. It is also possible that the ability of rods to form bilayers is unique to their geometry, which allows them to easily roll over each other. Furthermore, I showed that the process of rod flipping is highly correlated with the order of the film, with the rods located within a disordered region much more likely to undergo a flipping transition and then desorb from the interface. Since the shortest rods do not display highly ordered domains, the intriguing behaviour that characterises the rods of higher aspect ratio is not observed.

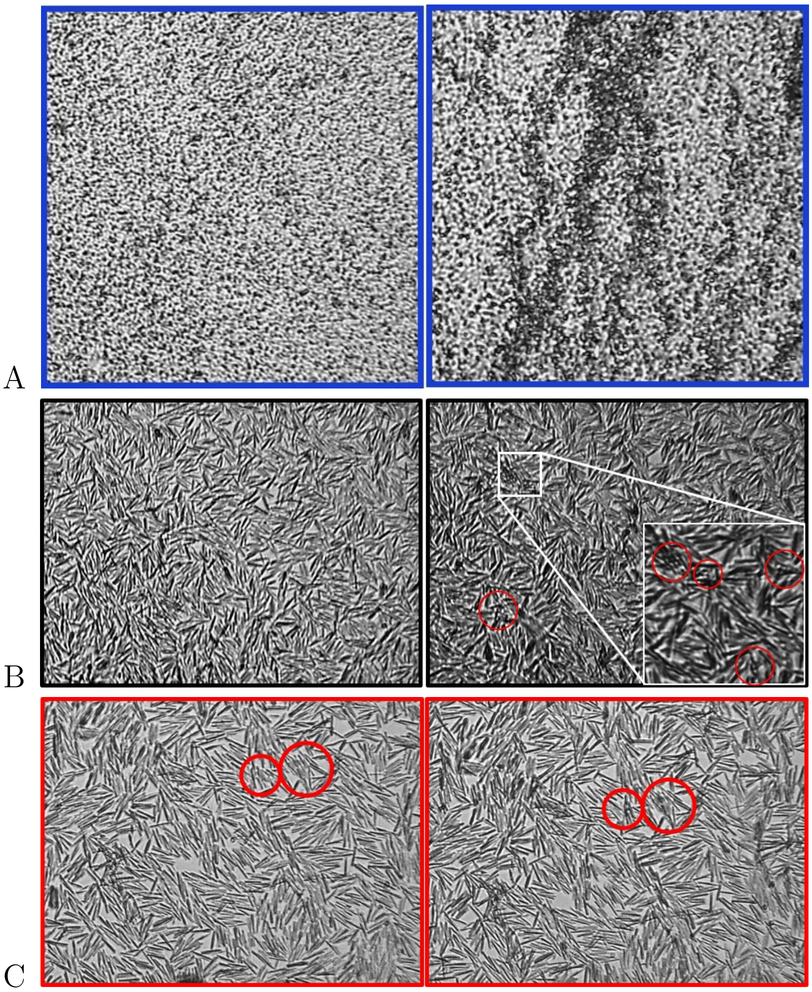


Figure 5.11 *Optical microscope images of monolayers of rods under compression for different aspect ratios: (A) AR=3, (B) AR=9 and (C) AR=15. Compression proceeds from left to right. For aspect ratio 9 I highlighted the flippers, whereas for aspect ratio 15 I highlighted the formation of a bilayer.*

During the formation of Pickering emulsions [39] and bijels [40], the coarsening of the interface exerts a force on the adsorbed colloids that is analogous to the one experienced during the compression of a monolayer. Therefore, I hope that these results will be helpful for the optimisation of the geometry of colloids in the development of better soft-matter materials.

Chapter 6

Conclusions

In this thesis I discussed the interfacial behaviour of proteins (and colloids) as observed from molecular dynamics simulations. For the considered systems, experiments are already very informative; however, they only provide a qualitative understanding of the process. In all cases, MD simulations complemented the results of the experiments by providing molecular level details that would not be otherwise accessible.

In chapter 3 I described the results obtained from coarse-grained (section 3.2) and all-atom (section 3.3) MD simulations of the protein BslA. The unusual character of this amphiphilic protein is already suggested from its crystal structure: BslA is able to tune the hydrophobicity of its cap region via a conformational change. Firstly, I showed that the behaviour of a BslA monomer at the interface can be fully captured by coarse-graining the protein into a much simpler Janus ellipsoidal colloid. Similarly to what observed for its colloidal counterpart, at low cap hydrophobicity the adsorbed protein adopts a side-on configuration, maximising the coverage of the unfavourable interface between water and oil; on the contrary, at high hydrophobicity the protein adopts an end-on configuration, where the apolar cap and the polar domain are respectively partitioned into the oil and water phases (and therefore optimising protein-solvent interactions). Secondly, all-atom simulations revealed the details of the BslA conformational change and self-assembly upon adsorption. In solution, the cap of BslA is highly disordered and many hydrophobic residues are buried, but when the protein reaches the interface, the cap folds into an ordered and hydrophobic β-sheet. The self-assembly was modelled starting from dimers found in the crystal structures of the proteins BslA and YweA, showing that the two-dimensional interfacial crystal is stabilised by a set of symmetric lateral interactions based on salt bridges. All-atom steered MD simulations allowed to compute the binding free energies of these BslA dimers and they showed that the proposed model compares successfully against experiments.

In chapter 4 I analysed the surface activity of the protein Ranaspumin-2. Similarly to what we have seen in the previous chapter, the combination of all-atom and coarse-grained simulations can provide a complete picture of the adsorption of the protein. All-atom simulations of Rsn-2 suggest that the protein adsorbs via a two-step mechanism. In the first step, the protein grabs the interface via its disordered and hydrophobic N-terminal tail. In the second step, the protein undergoes an unhinging conformational change where the α-helix and the β-sheet expose their hydrophobic sides towards the oil phase, while preserving the secondary structure content. However, all-atom simulations are expensive and the limited investigated time window may prevent the observation of all possible protein-interface configurations. For this reason, I developed a new structure-based coarse-grained model for the study of protein unfolding at interfaces. This model is highly efficient and enables the analysis of the full free energy landscape. Coarse-grained simulations with this model reveal the same two-step adsorption mechanism highlighted by all-atom simulations, including the subtle balance between a parallel and a perpendicular orientation of the folded protein relative to the interface.

Here I would like to make a comment on a particular feature shared by the two considered proteins: the importance of intrinsic disorder. In both cases, an intrinsically disordered region also plays an important role in the adsorption. However, these disordered regions are functionally distinct and they provide different advantages to each protein. For BslA, the presence of structural disorder resulted in a decrease of the hydrophobicity of the cap. This allows the protein to remain in a monomeric state in water, as opposed to hydrophobins, which are also water-soluble but form small oligomers to protect their hydrophobic patch [168]. In contrast, the hydrophobic residues of the disordered N-terminal tail of Rsn-2 are easily accessible to the solvent, allowing the protein to increase its interaction range and to efficiently capture the interface in the first adsorption step. This strategy is reminiscent of the so-called fly-casting mechanism employed by many intrinsically disordered proteins to bind their target site [139]. The different role of disorder in the two proteins is also reflected in their different kinetics of adsorption, which is diffusion-limited for Rsn-2 and slowed down by a free energy barrier for BslA.

Intrinsic disorder was also found to play an important role in the self-assembly of the hydrophobin EAS [17] at interfaces. In that case, the level of intrinsic disorder of the EAS loop region is high in solution and low at the interface. Since protein-protein association constrains the conformations of the loop, there is an entropic cost that inhibits self-assembly in solution. On the other hand, at the interface the loop is already ordered, the entropic penalty disappears and the protein is able to self-assemble. This consideration is also relevant for the two proteins considered in this thesis: in solution, the BslA cap and the Rsn-2 tail are both disordered, inhibiting protein association in this environment. At the interface, the Rsn-2 tail is expected to be still disordered, possibly explaining why the protein does not form an elastic film at the interface (from rheological measurements performed by Dr. Ryan Morris at the University of Edinburgh). On the other hand, the disorder-to-order transition of the BslA cap decreases its entropy and the orientational freedom of the protein (see Fig. 3.14), which should in turn contribute to lower the entropic cost of the observed self-assembly.

In chapter 5 I studied the compression of rod monolayers using dissipative particle dynamics simulations. I showed an intriguing behaviour characterised by rod flipping, ordering and bilayer formation that is highly dependent on the aspect ratio of the rods. For the shortest aspect ratio, AR=3, film compression simply causes the desorption of the rods. For AR=6, compression induces two-dimensional nematic ordering and rods desorb via a flipping mechanism at the boundary between different ordered domains (or equivalently within disordered regions). As the aspect ratio is increased up to AR=15, flipping disappears; instead, rods roll over each other forming bilayers. These observations are in good agreement with experiments.

These studies also offer an opportunity to discuss the application of MD simulations to protein adsorption. Here I will make some general comments based on my experience. Physics-based coarse-grained models such as MARTINI [58] can be used to analyse protein orientation relative to the interface and type of interactions that stabilise the adsorption. Furthermore, their low computational cost allows the calculation of the free energy of adsorption. However, these models are not designed to study protein conformational changes, in which case there are two possibilities: 1. all-atom simulations, or 2. a coarse-grained model that can capture conformational changes. All-atom simulations can successfully describe the behaviour of most systems, but their high computational cost and the fact that conformational changes are rare events often represent obstacles. For Rsn-2, it was possible to observe the unhinging transition during unbiased simulations, but this was not the case for the disorder-to-order transition in the BslA cap. In these situations one may employ enhanced sampling techniques such as replica exchange to speed-up the exploration of the phase space. Alternatively, the structure-based coarse-grained model introduced in section 4.3 can be used to study conformational changes upon adsorption and is computationally very efficient. I suggest that an essential feature of reliable coarse-grained models for adsorption is the inclusion of side-chains, because the amphiphilicity of secondary structure elements controls protein-interface interactions. In its current implementation, the main limitation of this structure-based model is that it can only be used to study unfolding, but not the formation of new secondary structure elements, because the Hamiltonian is based on a single native structure of the protein. However, with the inclusion of multiple basins [26], the model may be used to study the conformational change between the known solution and interfacial structures.

Appendix A

Adsorption to an interface in the presence of an energy barrier

In this section I describe a theoretical model of protein (or colloid) adsorption at interfaces in the presence of an energetic barrier. In particular I aimed at rationalising the experimental findings obtained by Bromley et al. in Ref. [1], which show that the time scale of BslA adsorption at the oil/water interface is much longer than what expected from the purely diffusive model by Ward and Tordai [13].

The bulk concentration of the protein, c, obeys the diffusion equation,

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \tag{A.1}$$

where D is the protein's diffusion coefficient and $x \geq 0$ is the direction perpendicular to the interface. The boundary conditions are

$$c(x, t = 0) = c(x = \infty, t) = c_0 \tag{A.2}$$

$$D\frac{\partial c(x,t)}{\partial x}\bigg|_{x=0} = kc(0,t) \tag{A.3}$$

where k is a kinetic constant, with the dimensions of a velocity, which models the barrier in a simple way ($k \to \infty$ corresponds to diffusion-limited adsorption). If we rescale the time by D/k^2, the space by D/k and the concentration by c_0, the equations reduce to:

$$\frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial x^2} \quad , \quad c(x,0) = c(\infty,t) = 1 \quad , \quad \left.\frac{\partial c}{\partial x}\right|_{x=0} = c(0,t) \tag{A.4}$$

To find the analytical form of $c(x,t)$, we can solve Eqs.(A.1-A.3) in Laplace space, and then compute the inverse Laplace transform of the solution. I define the Laplace transform in time of the concentration $W(x,s) = \mathcal{L}(c) = \int_0^{+\infty} c(x,t)e^{-st}dt$. Taking the laplace transform of the diffusion equation and the boundary conditions we obtain:

$$\frac{\partial^2 W}{\partial x^2} - sW = -1 \tag{A.5}$$

whose solution is:

$$\begin{aligned} W(x,s) &= Ae^{\sqrt{s}x} + Be^{-\sqrt{s}x} + \frac{1}{s} \\ &= -\frac{1}{s + s\sqrt{s}} e^{-\sqrt{s}x} + \frac{1}{s} \end{aligned} \tag{A.6}$$

where the coefficients A and B have been found from the boundary conditions $W(\infty,s) = 1/s$ and $\partial_x W(0,s) = W(0,s)$. The concentration is then obtained by taking the inverse Laplace transform of the expression above:

$$c(x,t) = \left[1 - \operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right)\right] + e^{x+t}\operatorname{erfc}\left(\frac{x}{2\sqrt{t}} + \sqrt{t}\right), \tag{A.7}$$

This solution is found by separating the first term of Eq. A.6 into $1/(s + s\sqrt{s}) = 1/\sqrt{s} - 1/(s+\sqrt{s})$, and then using $\mathcal{L}\operatorname{erfc}(a/2\sqrt{t}) = e^{-a\sqrt{s}}/s$ and $\mathcal{L}e^{a+t}\operatorname{erfc}(a/2\sqrt{t} + \sqrt{t}) = e^{-a\sqrt{s}}/(s + \sqrt{s})$ [169]. In real units the concentration is:

$$c(x,t) = c_0\left[1 - \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)\right] + c_0 e^{\frac{kx}{D} + \frac{k^2 t}{D}}\operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}} + k\sqrt{\frac{t}{D}}\right), \tag{A.8}$$

where the term in square brackets on the right hand side is the diffusion-limited results ($k \to \infty$).

We can then compute the surface density of the protein as follows:

$$\Gamma(t) = \int_0^t -kc(0,t')dt' = 2c_0\sqrt{\frac{Dt}{\pi}} + \frac{c_0 D}{k}\left(e^{\frac{k^2 t}{D}}\operatorname{erfc}\left(k\sqrt{\frac{t}{D}}\right) - 1\right) \tag{A.9}$$

As expected, taking k=$+\infty$ recovers the diffusion-limited result by Ward and

Tordai [13]:

$$\Gamma_{WT}(t) = 2c_0\sqrt{\frac{Dt}{\pi}} \tag{A.10}$$

Dimensional analysis suggests that k is proportional to D, and inversely to the size of the protein under consideration, σ; furthermore the presence of an energy barrier ΔE should slow down the dynamics a-la-Kramers [170], so that we expect

$$k \sim \frac{D}{\sigma} e^{-\frac{\Delta E}{k_B T}} \tag{A.11}$$

where k_B is the Boltzmann constant and T is the temperature.

Experimentally, the kinetics of diffusion is obtained from the Regime I time as a function of the bulk concentration c_0, which is the time at which the surface tension starts to decrease sharply and it signals a surface coverage θ=$\Gamma/\Gamma_{\text{full}}$ in the range ~50-100% [9]. The experimental data for the protein BslA can be explained with the theory which I have just outlined. Fig. A.1 shows the data together with the predictions of a model with no barrier (which can explain the L77K mutant data), and with a barrier of $\simeq 10k_BT$ hindering protein adsorption (which can explain the wild type data). Fig. A.1 also shows that the data for L77K-BslA suggests that in Regime I the coverage is not full (allowing for this also for the WT case would lead to very little change in the estimated energy barrier). The theoretical curves have been obtained using: σ=4.3 nm, Γ_{full}=1.57 mg/m^2 and D=9.87·10^{-7} cm^2/s (computed from the Stokes-Einstein relation).

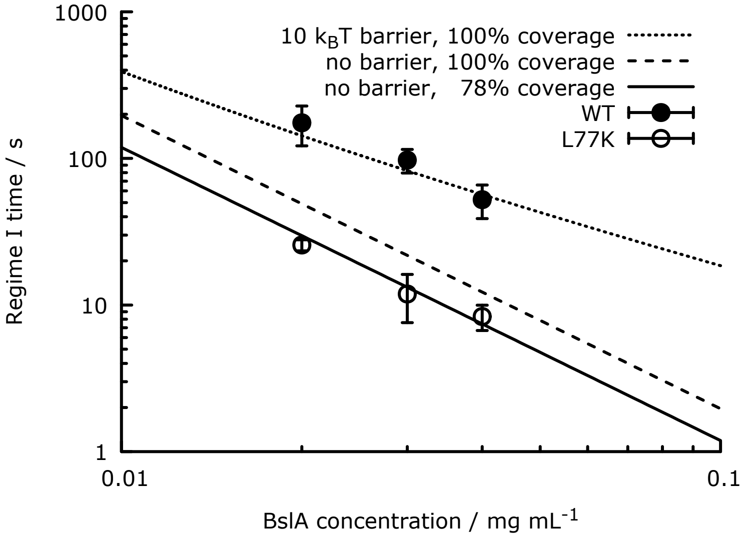


Figure A.1 *Regime I times versus BslA bulk concentrations. The dashed line shows the prediction assuming a diffusive dynamics of the protein in the solvent with no barrier to the adsorption and full coverage of the interface at the end of regime I. The times for L77K BslA are faster than this prediction; experimental data can be fitted by assuming that the surface coverage at the end of regime I is around 78%. On the other hand regime I times for WT BslA are slower than the prediction; in order to explain the data I introduce a barrier of 10* $k_B T$ *to the adsorption of the protein at the interface.*

Appendix B

Quaternions and rotations

In section 3.2.2, in order to represent the orientation of the Janus ellipsoidal colloid and to perform rotations during Monte-Carlo simulations, we employed the quaternion formalism [91]. Because of the direct correspondence between unit quaternions and rigid-body rotations, the former can be used to represent orientations and to perform rotations in computer simulations of rigid bodies. In principle, Euler angles could also be employed instead, but quaternions are significantly simpler to handle.

Quaternions are 4-dimensional vectors with a particular algebra. The four basis elements are denoted as 1, i, j and k, so that a quaternion can be written as:

$$q = q_1 + q_x i + q_y j + q_z k \tag{B.1}$$

The multiplication between two quaternions is called the Hamilton product, and is obtained from the products of the basis elements according to the distributive law. The four basis elements multiply each others according to the following multiplication matrix:

·	1	i	j	k
1	1	i	j	k
i	i	-1	k	-j
j	j	-k	-1	i
k	k	j	-i	-1

A rotation is defined by the rotation axis, represented by a unit vector $\hat{u} =$

(u_x, u_y, u_y), and by the rotation angle θ. This rotation can be represented by the unit quaternion:

$$q = \cos\frac{\theta}{2} + (u_x i + u_y j + u_y k)\sin\frac{\theta}{2} \tag{B.2}$$

The rotation of any vector $p = (p_x, p_y, p_z)$ can be performed by considering the vector as the quaternion with zero real part $p = 0 + p_x i + p_y j + p_z k$, and then using the Hamilton product as following:

$$p_R = qpp^{-1} \quad , \tag{B.3}$$

where p_R is the new vector after the rotation and $q^{-1} = \cos\frac{\theta}{2} - (u_x i + u_y j + u_y k)\sin\frac{\theta}{2}$. Two rotations represented by the quaternions q_1 and q_2 can be simply combined via the Hamilton product: $q_3 = q_2 q_1$. In our Monte-Carlo simulations, the orientation of the Janus colloid is represented by the quaternion q corresponding to the rotation with respect to the initial configuration of the rigid body (with the main axis perpendicular to the interface). At each time step, we attempt to change the current orientation q by performing a rotation Δq around a random axis u with an infinitesimal rotation angle $\Delta\theta$=0.01 rad:

$$\Delta q = \cos\frac{\Delta\theta}{2} + (u_x i + u_y j + u_y k)\sin\frac{\Delta\theta}{2} \tag{B.4}$$

The Monte-Carlo move is accepted according to the Metropolis criterion, where the energy of the system is computed from the orientation q and the distance between colloid and the interface (see methods in section 3.2.2).

Bibliography

[1] K. M. Bromley, R. J. Morris, L. Hobley, G. Brandani, R. M. C. Gillespie, M. McCluskey, U. Zachariae, D. Marenduzzo, N. R. Stanley-Wall, and C. E. MacPhee. Interfacial self-assembly of a bacterial hydrophobin. *PNAS*, 112(17):5419–5424, 2015.

[2] G. B. Brandani, M. Schor, R. Morris, N. Stanley-Wall, C. E. MacPhee, D. Marenduzzo, and U. Zachariae. The Bacterial Hydrophobin BslA is a Switchable Ellipsoidal Janus Nanocolloid. *Langmuir*, 31(42):11558–63, 2015.

[3] R. J. Morris, G. B. Brandani, V. Desai, B. O. Smith, M. Schor, and C. E. MacPhee. The Conformation of Interfacially Adsorbed Ranaspumin-2 is an Arrested State on the Unfolding Pathway. *arXiv:1602.04099*, 2016.

[4] E. Dickinson. Adsorbed protein layers at fluid interfaces: interactions, structure and surface rheology. *Colloids Surf., B*, 15(2):161–176, 1999.

[5] M. B. Linder, G. R. Szilvay, T. Nakari-Setälä, and M. E. Penttilä. Hydrophobins: the protein-amphiphiles of filamentous fungi. *FEMS Microbiol. Rev.*, 29(5):877–96, 2005.

[6] R. E. McDonald, R. I. Fleming, J. G. Beeley, D. L. Bovell, J. R. Lu, X. Zhao, A. Cooper, and M. W. Kennedy. Latherin: a surfactant protein of horse sweat and saliva. *PLOS ONE*, 4(5):e5726, 2009.

[7] L. Hobley, A. Ostrowski, F. V. Rao, K. M. Bromley, M. Porter, A. R. Prescott, C. E. MacPhee, D. M. F. v. Aalten, and N. R. Stanley-Wall. BslA is a self-assembling bacterial hydrophobin that coats the Bacillus subtilis biofilm. *PNAS*, 110(33):13600–13605, 2013.

[8] C. D. Mackenzie, B. O. Smith, A. Meister, A. Blume, X. Zhao, J. R. Lu, M. W. Kennedy, and A. Cooper. Ranaspumin-2: structure and function of a surfactant protein from the foam nests of a tropical frog. *Biophys. J*, 96(12):4984–92, 2009.

[9] C. Beverung, C. Radke, and H. Blanch. Protein adsorption at the oil/water interface: characterization of adsorption kinetics by dynamic interfacial tension measurements. *Biophysical Chemistry*, 81(1):59–80, 1999.

[10] M. E. Aulton and K. M. Taylor. *Aulton's Pharmaceutics: The Design and Manufacture of Medicines*. Elsevier Health Sciences UK, 2013.

[11] M. B. Linder. Hydrophobins: proteins that self assemble at interfaces. *Curr. Opin. Colloid Interface Sci.*, 14(5):356–363, 2009.

[12] D. Graham and M. Phillips. Proteins at liquid interfaces. *J. Colloid Interface Sci.*, 70(3):427–439, 1979.

[13] A. F. H. Ward and L. Tordai. Time-Dependence of Boundary Tensions of Solutions I. The Role of Diffusion in Time-Effects. *J. Chem. Phys*, 14(7):453, 1946.

[14] A. N. Frumkin. Surface tension curves of the higher fatty acids and the equation of condition of the surface layer. *Z. phys. Chem*, 116:466–484, 1925.

[15] S. Damodaran. In situ measurement of conformational changes in proteins at liquid interfaces by circular dichroism spectroscopy. *Anal. Bioanal. Chem.*, 376(2):182–8, 2003.

[16] I. Macindoe, A. H. Kwan, Q. Ren, V. K. Morris, W. Yang, J. P. Mackay, and M. Sunde. Self-assembly of functional, amphipathic amyloid monolayers by the fungal hydrophobin EAS. *PNAS*, 109(14):E804–11, 2012.

[17] A. De Simone, C. Kitchen, A. H. Kwan, M. Sunde, C. M. Dobson, and D. Frenkel. Intrinsic disorder modulates protein self-assembly and aggregation. *PNAS*, 109(18):6951–6, 2012.

[18] D. L. Cheung. Molecular Simulation of Hydrophobin Adsorption at an Oil–Water Interface. *Langmuir*, 28(23):8730–8736, 2012.

[19] D. Zare, K. M. McGrath, and J. R. Allison. Deciphering β-Lactoglobulin Interactions at an Oil-Water Interface: A Molecular Dynamics Study. *Biomacromolecules*, 16(6):1855–61, 2015.

[20] M. Arooj, N. S. Gandhi, C. A. Kreck, D. W. M. Arrigan, and R. L. Mancera. Adsorption and Unfolding of Lysozyme at a Polarized Aqueous-Organic Liquid Interface. *J. Phys. Chem. B*, 120(12):3100–3112, 2016.

[21] N. R. Stanley-Wall and C. E. MacPhee. Connecting the dots between bacterial biofilms and ice cream. *Physical biology*, 12(6):063001, 2015.

[22] D. Wendell, J. Todd, and C. Montemagno. Artificial photosynthesis in ranaspumin-2 based foam. *Nano Lett.*, 10(9):3231–6, 2010.

[23] K. Kobayashi and M. Iwano. BslA (YuaB) forms a hydrophobic layer on the surface of Bacillus subtilis biofilms. *Mol. Microbiol.*, 85(1):51–66, 2012.

[24] C. I. Branden and J. *Introduction to Protein Structure*. Garland Science, 1999.

[25] K. A. Dill and J. L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110):1042–6, 2012.

[26] K. Okazaki, N. Koga, S. Takada, J. N. Onuchic, and P. G. Wolynes. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *PNAS*, 103(32):11844–9, 2006.

[27] P. Tompa. Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences*, 37(12):509–16, 2012.

[28] D. Eisenberg, R. M. Weiss, T. C. Terwilliger, and W. Wilcox. Hydrophobic moments and protein structure. In *Faraday Symp. Chem. Soc.*, volume 17, pages 109–120. The Royal Society of Chemistry, Royal Society of Chemistry, 1982.

[29] C. Levinthal. How to fold graciously. *Mossbauer spectroscopy in biological systems*, pages 22–24, 1969.

[30] C. Clementi, H. Nymeyer, and J. N. Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol*, 298(5):937–53, 2000.

[31] The unfoldomics decade: an update on intrinsically disordered proteins. *BMC genomics*, 9 Suppl 2(2):S1, 2008.

[32] T. Terakawa, H. Kenzaki, and S. Takada. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J. Am. Chem. Soc.*, 134(35):14555–62, 2012.

[33] H. Kenzaki and S. Takada. Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLoS Comput. Biol.*, 11(8):e1004443, 2015.

[34] Wilson CK Poon and David Andelman. *Soft condensed matter physics in molecular and cell biology.* CRC Press, 2006.

[35] M. J. Rosen and J. T. Kunjappu. *Surfactants and Interfacial Phenomena.* John Wiley & Sons, 2012.

[36] P. M. Chaikin, T. C. Lubensky, and T. A. Witten. *Principles of condensed matter physics*, volume 1. Cambridge Univ Press, 2000.

[37] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–7, 2005.

[38] H. Lehle and M. Oettel. Stability and interactions of nanocolloids at fluid interfaces: effects of capillary waves and line tensions. *J. Phys. Condens. Matter*, 20(40):404224, 2008.

[39] R. Aveyard, B. P. Binks, and J. H. Clint. Emulsions stabilised solely by colloidal particles. *Adv. Colloid Interface Sci.*, 100-102:503–546, 2003.

[40] M. E. Cates and P. S. Clegg. Bijels: a new class of soft materials. *Soft Matter*, 4(11):2132, 2008.

[41] J. Loudet, a. Yodh, and B. Pouligny. Wetting and Contact Lines of Micrometer-Sized Ellipsoids. *Phys. Rev. Lett.*, 97(1):018304, 2006.

[42] J. C. Loudet and B. Pouligny. How do mosquito eggs self-assemble on the water surface? *Eur. Phys. J. E Soft Matter*, 34(8):76, 2011.

[43] L. Botto, L. Yao, R. L. Leheny, and K. J. Stebe. Capillary bond between rod-like particles and the micromechanics of particle-laden interfaces. *Soft Matter*, 8(18):4971, 2012.

[44] M. Cavallaro, L. Botto, E. P. Lewandowski, M. Wang, and K. J. Stebe. Curvature-driven capillary migration and assembly of rod-like particles. *PNAS*, 108(52):20923–20928, 2011.

[45] A. Walther and A. H. E. Müller. Janus particles. *Soft Matter*, 4(4):663, 2008.

[46] N. Glaser, D. J. Adams, A. Böker, and G. Krausch. Janus particles at liquid-liquid interfaces. *Langmuir*, 22(12):5227–9, 2006.

[47] A. Kumar, B. J. Park, F. Tu, and D. Lee. Amphiphilic Janus particles at fluid interfaces. *Soft Matter*, 9(29):6604–6617, 2013.

[48] M. Oettel and S. Dietrich. Colloidal interactions at fluid interfaces. *Langmuir*, 24(4):1425–41, 2008.

[49] D. C. Rapaport. *The Art of Molecular Dynamics Simulation.* Cambridge University Press, 2004.

[50] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, 2010.

[51] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–20, 2011.

[52] V. Leone, F. Marinelli, P. Carloni, and M. Parrinello. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.*, 20(2):148–54, 2010.

[53] A. Barducci, M. Bonomi, M. K. Prakash, and M. Parrinello. Free-energy landscape of protein oligomerization from atomistic simulations. *PNAS*, 110(49):E4708–13, 2013.

[54] R. B. Best, W. Zheng, and J. Mittal. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.*, 10(11):5113–5124, 2014.

[55] C. Camilloni and M. Vendruscolo. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J. Am. Chem. Soc.*, 136(25):8982–91, 2014.

[56] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–25, 2006.

[57] R. B. Best and G. Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B*, 113(26):9004–15, 2009.

[58] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. Marrink. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.*, 4(5):819–834, 2008.

[59] D. L. Parton, J. W. Klingelhoefer, and M. S. P. Sansom. Aggregation of model membrane proteins, modulated by hydrophobic mismatch, membrane curvature, and protein class. *Biophys. J*, 101(3):691–9, 2011.

[60] X. Periole, M. Cavalli, S. Marrink, and M. a. Ceruso. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.*, 5(9):2531–2543, 2009.

[61] A. Radzicka and R. Wolfenden. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry*, 27(5):1664–1670, 1988.

[62] H. F. Trotter. On the product of semi-groups of operators. *Proc. Amer. Math. Soc.*, 4(10):545–551, 1959.

[63] W. C. Swope. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys*, 76(1):637, 1982.

[64] R. Hockney, S. Goel, and J. Eastwood. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.*, 14(2):148–158, 1974.

[65] H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys*, 81(8):3684, 1984.

[66] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys*, 126(1):14101, 2007.

[67] W. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[68] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[69] P. Das, M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *PNAS*, 103(26):9885–90, 2006.

[70] G. A. Tribello, M. Ceriotti, and M. Parrinello. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *PNAS*, 109(14):5196–201, 2012.

[71] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys*, 3:300, 1935.

[72] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[73] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[74] M. Souaille and B. Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.*, 135(1):40–57, 2001.

[75] A. Barducci, G. Bussi, and M. Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.*, 100(2):020603, 2008.

[76] S. Park and K. Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys*, 120(13):5946–5961, 2004.

[77] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78(14):2690–2693, 1997.

[78] A. Magarkar, N. Mele, N. Abdel-Rahman, S. Butcher, M. Torkkeli, R. Serimaa, A. Paananen, M. Linder, and A. Bunker. Hydrophobin film structure for HFBI and HFBII and mechanism for accelerated film formation. *PLoS Comput. Biol.*, 10(7):e1003745, 2014.

[79] S. R. Euston. Molecular dynamics simulation of protein adsorption at fluid interfaces: a comparison of all-atom and coarse-grained models. *Biomacromolecules*, 11(10):2781–2787, 2010.

[80] M. Deighan and J. Pfaendtner. Exhaustively sampling peptide adsorption with metadynamics. *Langmuir*, 29(25):7999–8009, 2013.

[81] M. Deighan, M. Bonomi, and J. Pfaendtner. Efficient Simulation of Explicitly Solvated Proteins in the Well- Tempered Ensemble. *J. Chem. Theory Comput.*, 2(1):8–11, 2012.

[82] J. N. Wilking, T. E. Angelini, A. Seminara, M. P. Brenner, and D. A. Weitz. Biofilms as complex fluids. *MRS Bull*, 36(5):385–391, 2011.

[83] L. Hobley, C. Harkins, C. E. MacPhee, and N. R. Stanley-Wall. Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes. *FEMS Microbiol. Rev.*, 39(5):649–69, 2015.

[84] D. Romero, C. Aguilar, R. Losick, and R. Kolter. Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *PNAS*, 107(5):2230–2234, 2010.

[85] A. T. Kovács, J. v. Gestel, and O. P. Kuipers. The protective layer of biofilm: a repellent function for a new class of amphiphilic proteins. *Mol. Microbiol.*, 85(1):8–11, 2012.

[86] B. J. Park and D. Lee. Equilibrium orientation of nonspherical Janus particles at fluid–fluid interfaces. *ACS nano*, 6(1):782–790, 2011.

[87] F. Sciortino, A. Giacometti, and G. Pastore. Phase Diagram of Janus Particles. *Phys. Rev. Lett.*, 103(23):237801, 2009.

[88] J. J. Muller and H. Schrauber. The inertia-equivalent ellipsoid: a link between atomic structure and low-resolution models of small globular proteins determined by small-angle X-ray scattering. *J. Appl. Crystallogr.*, 25(2):181–191, 1992.

[89] B. J. Park and D. Lee. Configuration of nonspherical amphiphilic particles at a fluidfluid interface. *Soft Matter*, 8(29):7690, 2012.

[90] Kurt Binder, editor. *Monte Carlo Methods in Statistical Physics*, volume 7 of *Topics in Current Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.

[91] J. B. Kuipers. *Quaternions and rotation sequences*, volume 66. Princeton university press, 1999.

[92] P. P. Klein. On the Ellipsoid and Plane Intersection Equation. *Applied Mathematics*, 03(11):1634–1640, 2012.

[93] A. Goebel and K. Lunkenheimer. Interfacial Tension of the Water/ n - Alkane Interface. *Langmuir*, 13(2):369–372, 1997.

[94] B. Hess, C. Kutzner, D. v. d. Spoel, and E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[95] S. Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Mol. Phys*, 50(5):1055–1076, 1983.

[96] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schafer, X. Periole, D. P. Tieleman, S. J. Marrink, L. V. Schäfer, D. H. D. Jong, L. V. Scha, L. V. Schafer, X. Periole, D. P. Tieleman, and S. J. Marrink. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.*, 9(1):687–697, 2012.

[97] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, and S. J. Marrink. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.*, 6(6):e1000810, 2010.

[98] D. J. Donahue and F. E. Bartell. The Boundary Tension at Water-Organic Liquid Interfaces. *J. Phys. Chem.*, 56(4):480–484, 1952.

[99] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. v. Gunsteren, and A. E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.*, 38(1-2):236–240, 1999.

[100] W. W. L. DeLano. The PyMOL Molecular Graphics System. 2002.

[101] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw. Systematic validation of protein force fields against experimental data. *PLOS ONE*, 7(2):e32131, 2012.

[102] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*, 79(2):926, 1983.

[103] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[104] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992.

[105] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys*, 98(12):10089, 1993.

[106] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys*, 52(2):255–268, 2006.

[107] C. Camilloni, P. Robustelli, A. . Simone, A. Cavalli, and M. Vendruscolo. Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. *J. Am. Chem. Soc.*, 134(9):3968–71, 2012.

[108] E. T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106(4):620–630, 1957.

[109] A. Cavalli, C. Camilloni, and M. Vendruscolo. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys*, 138(9):094112, 2013.

[110] J. M. Krieger, G. Fusco, M. Lewitzky, P. C. Simister, J. Marchant, C. Camilloni, S. M. Feller, and A. . Simone. Conformational recognition of an intrinsically disordered protein. *Biophys. J*, 106(8):1771–9, 2014.

[111] A. Kannan, C. Camilloni, A. B. Sahakyan, A. Cavalli, and M. Vendruscolo. A conformational ensemble derived using NMR methyl chemical shifts reveals a mechanical clamping transition that gates the binding of the HU protein to DNA. *J. Am. Chem. Soc.*, 136(6):2204–7, 2014.

[112] P. Kukic, C. Camilloni, A. Cavalli, and M. Vendruscolo. Determination of the individual roles of the linker residues in the interdomain motions of calmodulin using NMR chemical shifts. *J. Mol. Biol*, 426(8):1826–38, 2014.

[113] J. Henriques, C. Cragnell, and M. Skepö. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.*, 11(7):3420–31, 2015.

[114] K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, and M. Vendruscolo. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.*, 131(39):13894–5, 2009.

[115] C. Camilloni, A. Cavalli, and M. Vendruscolo. Assessment of the Use of NMR Chemical Shifts as Replica-Averaged Structural Restraints in Molecular Dynamics Simulations to Characterize the Dynamics of Proteins. *J. Phys. Chem. B*, 117(6):1838–1843, 2013.

[116] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.*, 185(2):604–613, 2014.

[117] B. Fu, A. B. Sahakyan, C. Camilloni, G. G. Tartaglia, E. Paci, A. Caflisch, M. Vendruscolo, and A. Cavalli. ALMOST: An all atom molecular simulation toolkit for protein structure determination. *J. Comput. Chem.*, 35(14):1101–1105, 2014.

[118] L. Wang, R. A. Friesner, and B. J. Berne. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B*, 115(30):9431–8, 2011.

[119] G. Bussi. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol. Phys*, 112(3-4):379–384, 2014.

[120] B. Efron. *The Jackknife, the Bootstrap, and Other Resampling Plans.* SIAM, 1982.

[121] P. R. t. Wolde and D. Frenkel. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science*, 277(5334):1975–1978, 1997.

[122] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, 125(7):1731–7, 2003.

[123] F. Castellani, B. v. Rossum, A. Diehl, M. Schubert, K. Rehbein, and H. Oschkinat. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*, 420(6911):98–102, 2002.

[124] M. Arai, K. Sugase, H. J. Dyson, and P. E. Wright. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *PNAS*, 112(31):9614–9, 2015.

[125] K. Okazaki and S. Takada. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *PNAS*, 105(32):11182–7, 2008.

[126] E. Bianchi, R. Blaak, and C. N. Likos. Patchy colloids: state of the art and perspectives. *Phys. Chem. Chem. Phys*, 13(14):6397–410, 2011.

[127] A. Cordomi, G. Caltabiano, and L. Pardo. Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters. *J. Chem. Theory Comput.*, 8(3):948–958, 2012.

[128] O. Berger, O. Edholm, and F. Jähnig. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J*, 72(5):2002–13, 1997.

[129] C. Caleman, P. J. v. Maaren, M. Hong, J. S. Hub, L. T. Costa, D. v. d. Spoel, P. J. V. Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. V. D. Spoel. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.*, 8(1):61–74, 2012.

[130] A. Laio and F. L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, 2008.

[131] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[132] I MathWorks. Matlab and statistics toolbox release, 2012.

[133] A. Rodriguez and A. Laio. Machine learning. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–6, 2014.

[134] H. Taketomi, Y. Ueda, and N. Go. Studies on protein folding, unfolding and fluctuations by computer simulation. *International journal of peptide and protein research*, 7(6):445–459, 1975.

[135] C. Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.*, 18(1):10–5, 2008.

[136] R. B. Best, G. Hummer, and W. A. Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *PNAS*, 110(44):17874–9, 2013.

[137] W. Li, W. Wang, and S. Takada. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *PNAS*, 111(29):10550–5, 2014.

[138] J. M. Rogers, V. Oleinikovas, S. L. Shammas, C. T. Wong, D. . Sancho, C. M. Baker, and J. Clarke. Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *PNAS*, 111(43):15420–5, 2014.

[139] Y. Levy, J. N. Onuchic, and P. G. Wolynes. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.*, 129(4):738–9, 2007.

[140] O. Givaty and Y. Levy. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol*, 385(4):1087–97, 2009.

[141] M. Cieplak, D. B. Allan, R. L. Leheny, and D. H. Reich. Proteins at air-water interfaces: a coarse-grained model. *Langmuir*, 30(43):12888–96, 2014.

[142] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. d. Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007.

[143] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic acids research*, 38(Web Server issue):W657–61, 2010.

[144] J. K. Noel, P. C. Whitford, and J. N. Onuchic. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B*, 116(29):8692–702, 2012.

[145] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.*, 117(1):1–19, 1995.

[146] J. F. Dama, M. Parrinello, and G. A. Voth. Well-Tempered Metadynamics Converges Asymptotically. *Phys. Rev. Lett.*, 112(24):240602, 2014.

[147] M. Bonomi, a. Barducci, and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.*, 30(11):1615–21, 2009.

[148] V. Narayanaswami, R. S. Kiss, and P. M. M. Weers. The helix bundle: a reversible lipid binding motif. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*, 155(2):123–33, 2010.

[149] M. Wientzek, C. M. Kay, K. Oikawa, and R. O. Ryan. Binding of insect apolipophorin III to dimyristoylphosphatidylcholine vesicles. Evidence for a conformational change. *J. Biol. Chem.*, 269(6):4605–4612, 1994.

[150] R. A. Staniforth, S. Giannini, L. D. Higgins, M. J. Conroy, A. M. Hounslow, R. Jerala, C. J. Craven, and J. P. Waltho. Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J.*, 20(17):4774–81, 2001.

[151] S. Jordens, L. Isa, I. Usov, and R. Mezzenga. Non-equilibrium nature of two-dimensional isotropic and nematic coexistence in amyloid fibrils at liquid interfaces. *Nat. Commun.*, 4:1917, 2013.

[152] M. A. Bates and D. Frenkel. Phase behavior of two-dimensional hard rod fluids. *J. Chem. Phys*, 112(22):10034, 2000.

[153] W. Zhou, J. Cao, W. Liu, and S. Stoyanov. How Rigid Rods Self-Assemble at Curved Surfaces. *Angew. Chem. Int. Ed.*, 48(2):378–381, 2009.

[154] B. Madivala, S. Vandebril, J. Fransaer, and J. Vermant. Exploiting particle shape in solid stabilized emulsions. *Soft Matter*, 5(8):1717, 2009.

[155] G. B. Davies, T. Krüger, P. V. Coveney, J. Harting, and F. Bresme. Assembling Ellipsoidal Particles at Fluid Interfaces Using Switchable Dipolar Capillary Interactions. *Adv. Mater.*, 26(39):6715–6719, 2014.

[156] M. G. Basavaraj, G. G. Fuller, J. Fransaer, and J. Vermant. Packing, flipping, and buckling transitions in compressed monolayers of ellipsoidal latex particles. *Langmuir*, 22(15):6605–12, 2006.

[157] P. Espanol and P. Warren. Statistical mechanics of dissipative particle dynamics. *EPL*, 30(4):191, 1995.

[158] G. RD and W. PB. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.*, 107(11):4423–4435, 1997.

[159] M. Laradji and M. J. A. Hore. Nanospheres in phase-separating multicomponent fluids: a three-dimensional dissipative particle dynamics simulation. *J. Chem. Phys*, 121(21):10641–7, 2004.

[160] M. J. A. Hore and M. Laradji. Microphase separation induced by interfacial segregation of isotropic, spherical nanoparticles. *J. Chem. Phys*, 126(24):244903, 2007.

[161] M. J. A. Hore and M. Laradji. Prospects of nanorods as an emulsifying agent of immiscible blends. *J. Chem. Phys*, 128(5):054901, 2008.

[162] M. Huang, Z. Li, and H. Guo. The effect of Janus nanospheres on the phase separation of immiscible polymer blends via dissipative particle dynamics simulations. *Soft Matter*, 8(25):6834, 2012.

[163] M. Huang and H. Guo. The intriguing ordering and compatibilizing performance of Janus nanoparticles with various shapes and different dividing surface designs in immiscible polymer blends. *Soft Matter*, 9(30):7356, 2013.

[164] J. Desplat, I. Pagonabarraga, and P. Bladon. LUDWIG: A parallel Lattice-Boltzmann code for complex fluids. *Comput. Phys. Commun.*, 134(3):273–290, 2001.

[165] J. A. Elliott and A. H. Windle. A dissipative particle dynamics method for modeling the geometrical packing of filler particles in polymer composites. *J. Chem. Phys*, 113(22):10367, 2000.

[166] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Clarendon Press, 1989.

[167] J. S. Rowlinson and B. Widom. *Molecular Theory of Capillarity.* Courier Corporation, 2013.

[168] X. Wang, J. F. Graveland-Bikker, C. G. d. Kruif, and G. T. Robillard. Oligomerization of hydrophobin SC3 in solution: from soluble state to self-assembly. *Protein Sci.*, 13(3):810–21, 2004.

[169] M. R. Spiegel and J. Liu. *Mathematical Handbook of Formulas and Tables.* McGraw-Hill, 1999.

[170] H. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.