Molecular Dynamics Simulation of Protein Biosurfactants

David L. Cheung * and Suman Samantray

School of Chemistry, National University of Ireland Galway, University Road, H91 TK33 Galway, Ireland; s.samantray1@nuigalway.ie

* Correspondence: david.cheung@nuigalway.ie; Tel.: +353-91-492450

Received: 19 August 2018; Accepted: 6 September 2018; Published: 8 September 2018

Abstract: Surfaces and interfaces are ubiquitous in nature and are involved in many biological processes. Due to this, natural organisms have evolved a number of methods to control interfacial and surface properties. Many of these methods involve the use of specialised protein biosurfactants, which due to the competing demands of high surface activity, biocompatibility, and low solution aggregation may take structures that differ from the traditional head–tail structure of small molecule surfactants. As well as their biological functions, these proteins have also attracted interest for industrial applications, in areas including food technology, surface modification, and drug delivery. To understand the biological functions and technological applications of protein biosurfactants, it is necessary to have a molecular level description of their behaviour, in particular at surfaces and interfaces, for which molecular simulation is well suited to investigate. In this review, we will give an overview of simulation studies of a number of examples of protein biosurfactants (hydrophobins, surfactin, and ranaspumin). We will also outline some of the key challenges and future directions for molecular simulation in the investigation of protein biosurfactants and how this can help guide future developments.

Keywords: biosurfactants; molecular dynamics simulation; protein structure; interfacial science

1. Introduction

In order to control the properties of surfaces and interfaces, many organisms have evolved specialised surface active biomolecules, such as protein or peptide biosurfactants, which fulfill a number of key biological functions [1,2]. Examples of these functions (Figure 1) include the reduction in air–water surface tension, easing the growth of aerial bodies, such as spores or hyphae [3,4], or the spreading of bacterial biofilms [5]. Adsorption of biosurfactant proteins onto surfaces is often used to aid the adhesion of microorganisms [6]. Biosurfactant proteins also commonly interact with membranes, which can lead to antibacterial, antifungal, and antiviral behaviours [7], giving these important defensive roles. Other functions of biosurfactant proteins include inhibition of immune response [8] and stabilization of polyhydroxyalkanoate (PHA) granules [9]. As they fulfill such a diverse range of functions, they have evolved a variety of structures [10,11] that can differ from those of other proteins.

As adsorption onto interfaces is central to the function of biosurfactant proteins, they have evolved a number of strategies to maximise their interfacial adsorption. It should be noted that almost all proteins will adsorb onto interfaces and surfaces, as they contain both hydrophobic and hydrophilic amino acids making them intrinsically amphiphilic. In most cases, this is accompanied by unfolding [12,13] and loss of function, so it is typically avoided. By contrast, biosurfactant proteins typically retain their structure or undergo specific changes in conformation, so they have evolved structural features that favour interfacial adsorption. Alongside this, the structures of biosurfactant proteins are also often influenced by the need to maintain biocompatibility and to minimise their
aggregation in solution. This has caused them to evolve designs that differ from the simple polar head/hydrophobic tail structure found in man-made surfactants and phospholipids [10,11,14]. As these properties would also be advantageous in many material systems, biosurfactant proteins are increasingly being investigated for industrial applications in a number of sectors [15,16] (Table 1).

Figure 1. Schematic illustration of common biological functions of biosurfactants. (a) formation of interfacial layer to assist growth of aerial bodies; (b) formation of surface layer aiding cell attachment and growth; (c) membrane adsorption and poration.

Table 1. Examples of investigated applications for protein biosurfactants.

<table>
<thead>
<tr>
<th>Application</th>
<th>Protein(s)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emulsion/foam stabilisers in food</td>
<td>HFBII (hydrophobin)</td>
<td>[17,18]</td>
</tr>
<tr>
<td></td>
<td>Globulin</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>Casein</td>
<td>[20]</td>
</tr>
<tr>
<td>Drug delivery</td>
<td>HFBI (hydrophobin)</td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>Casein</td>
<td>[22]</td>
</tr>
<tr>
<td>Nanoparticle synthesis</td>
<td>Rsn-2</td>
<td>[23]</td>
</tr>
<tr>
<td>Biomineralization</td>
<td>H⁺ Protein B (hydrophobin)</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>Surfactin</td>
<td>[25]</td>
</tr>
<tr>
<td>Remediation</td>
<td>Surfactin</td>
<td>[26]</td>
</tr>
<tr>
<td>Exfoliation of 2D materials</td>
<td>HFBI (hydrophobin)</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Vmh2 (hydrophobin)</td>
<td>[28]</td>
</tr>
</tbody>
</table>

Understanding the biological functions and technological applications of biosurfactant proteins requires knowledge of their structure and behaviour at interfaces. As changes in these environments can occur over nanometre length scales, understanding this requires microscopic detail that is challenging to obtain experimentally. Molecular simulation operates directly on the molecular level and thus can provide such microscopic detail. Since the first molecular dynamics simulation of proteins, these have increasingly been recognised as a powerful tool for the investigation of protein structure and function [29]. This has been driven by improvements in computer power and simulation methodologies; the first protein simulation investigated the 58-residue bovine pancreatic trypsin inhibitor [30] (without water), while simulations on current state-of-the-art hardware can handle the entire HIV capsid (64 million atoms) [31] or significant regions of the cell cytoplasm [32]. A number of studies have applied molecular simulation to the investigation of proteins at surfaces and interfaces, giving significant insight into their behaviour in these environments [33]. These have investigated a
range of proteins, including enzymes, such as lipase [34], globular proteins, such as lysozyme [35–37] or β-lactoglobulin [38,39], and intrinsically disordered proteins, such as amyloid-β [40].

In this review, we will give an overview of molecular simulation work investigating the structure and properties of biosurfactant proteins and peptides. This will focus on the well-studied cases of hydrophobins and surfactin, along with the novel biosurfactant protein Rsn-2. As we focus on simulation studies, key examples of protein biosurfactants, such as caseins, that have not been the subject of significant MD work will not be covered. Similarly, lipids and other surface-active biomolecules lie outside the scope of this review. Also excluded is lung surfactant, which, while containing a number of key proteins, the main surface active component is the phospholipid component.

2. Molecular Simulation

In this section, we will briefly describe the molecular dynamics simulation method and common simulation models. This is not intended to be an exhaustive and detailed discussion of these, rather a short overview necessary for understanding the following sections. There are a number of excellent textbooks [41,42] that the interested reader may consult for a fuller description of molecular simulation methods and the underlying algorithms and statistical mechanics.

For simulation of biomolecular systems, the most common technique is molecular dynamics (MD) simulation. This corresponds to numerically solving the classical Newtonian equations of motion in a step-wise, iterative manner. If the positions ($r_i(t)$) and velocities ($v_i(t)$) of the atoms are known at a given time $t$, then the forces ($F_i(t)$) at that time can be calculated. These forces can then be used to determine the velocities and positions a short time $t + \delta t$ later, from which a new set of forces can be calculated, and so on. The time interval $\delta t$ is commonly referred to as the time step; for a biomolecular system, the time step is typically on the order of 1 fs, so a typical MD simulation spanning 10 to 100 nanoseconds will entail 10–100 millions of time steps.

A key factor in molecular simulation is the model used, which determines the overall accuracy and degree of resolution of the study. Most simulation studies in biophysics employ atomistic models, in which the interactions are described using a classical energy function referred to as a force field. This gives the energy of the system as a sum over intra- and intermolecular interactions, with a typical form being [43]

$$E = \sum_{\text{bonds}} \frac{1}{2} k_b (\ell - \ell_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \sum_n \frac{1}{2} V_n (1 + \cos(n\phi + \delta_n))$$

$$+ \sum_{i<j} 4\varepsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}},$$

(1)

where the terms on the right-hand side are bond stretching, bond angle bending, dihedral rotation (together referred to as the bonded interactions), van der Waals, and electrostatic terms, respectively. In the bonded interactions $k_b$, $k_a$, and $V_n$ are force constants that describe the energetic costs of bond stretching, bond angle bending, and dihedral rotation, $\ell_0$ and $\theta_0$ are equilibrium bond lengths, and bond angles. $e$, $\sigma$, and $q_i$ are the van der Waals well depths, van der Waals diameters, and atomic partial charges, respectively. The functional form in Equation 1 is the simplest form using harmonic terms for bond stretching and angle bending and the simple Lennard–Jones potential for van der Waals interactions. More sophisticated forms and potentials that describe other interactions, such as coupling between different intramolecular distortions (e.g., stretch-bend), can be used. Alongside the potential function, force fields also define the parameters (e.g., force constants, equilibrium bond lengths) that appear in Equation (1). These parameters have to be determined either from experimental data or other theoretical methods (such as quantum chemical calculations). A number of different force fields have been developed for modelling biomolecular systems, which differ in their functional forms.
and parameterization procedures. Common models that have been applied to proteins include the Amber [44], Charmm [45], Gromos [46], and OPLS [47] families of force fields. These have been refined over the past 30 years as new experimental data have been gained, with the latest iterations of these force fields being capable of describing the structures and folding of proteins.

The starting point for a MD simulation is typically an experimentally determined protein structure, typically from X-ray crystallography or solution NMR. As structures for a limited number of protein biosurfactants are known, most simulation studies have focused on this small group. Computational methods for protein structure prediction, such as homology modelling, may be used to create conformations for proteins without experimentally determined structures. These generally rely on knowledge of structure for similar proteins (in terms of sequence), which due to the diverse sequences for many biosurfactant proteins can limit their applicability.

While they provide detailed molecular level information, atomistic simulation is limited in the system sizes it can investigate. To study larger systems, it is possible to reduce the level of detail through the use of coarse-grain (CG) models. In these, a number of heavy (non-hydrogen) atoms are grouped together into a single interaction site. One common CG model for biomolecules is the Martini model [48,49], developed by the Marrink group. Originally, this was developed for simulations of membranes and was later extended to include proteins [50]. In the Martini model, each CG bead in a protein (or lipid) typically represents four heavy atoms, so each protein residue will be composed of between one and five beads. This mapping also applies to the solvent so a water bead corresponds to four water molecules. Due to its ability to model larger systems and its inclusion in the commonly used Gromacs package, the Martini model has been extensively applied to proteins, in particular to membrane–protein systems [51]. Direct comparison between CG and atomistic models shows that these can give qualitatively similar results for proteins at interfaces and surfaces, but, due to the loss of resolution, they can differ in fine detail, such as the position of proteins relative to an oil–water interface [52]. A similar model, the Shinoda–DeVane–Klein model [53,54], has also been developed and extended to investigate proteins.

Atomistic molecular dynamics simulation has proven to be a powerful tool for the investigation of the behaviour of biosurfactant proteins, but it has some limitations. One challenge for atomistic simulation is in sampling conformations for complex molecules such as proteins. This arises as different conformations can often be separated by energy barriers significantly higher than the thermal energy ($k_B T$). To overcome this, one common method is replica exchange molecular dynamics [55,56]. Here, a number of replicas of the system are simulated concurrently at different temperatures; replicas at high temperatures are able to get over the energy barriers separating different conformations, while replicas at lower temperatures can explore the regions around the minima in detail. By periodically exchanging the temperatures, each replica is simulated at efficient sampling of different protein conformations can be achieved. This has been widely used in the simulation of biomolecules, with a number of recent applications to proteins at interfaces [36,37,57] and surfaces [58].

Another challenge is in calculating free energies, which can be used to determine the relative stability of different protein conformations or adsorption strengths onto surfaces or interfaces. This is traditionally challenging in MD simulation as it requires sampling of all the possible states a system, including rarely found high energy states. One commonly used method for determining free energy differences is metadynamics [59,60] (MetaD). The MetaD algorithm assumes that the state of the system can be a few collective variables, such as measures of protein structure, molecular positions, or energies. MetaD then introduces a history bias potential that is a function of these collective variables, which is updated throughout the simulation and will eventually converge to the negative of the free energy surface. A number of variants of metadynamics have been proposed [61,62], which differ in the how the bias function is updated, to enhance convergence. MetaD has been applied to investigated the adsorption of short peptides onto surfaces [63–66] and interfaces [67], and with increasing computer power its use for large proteins is becoming possible. When the free energy as a function of molecular
position is needed, such as calculating adsorption free energies, umbrella sampling (US) \[68\] or steered molecular dynamics (SMD) \[69\] simulations have been used. In US, typically harmonic potential, is applied to constrain the group of atoms is constrained around an equilibrium position \(r_i\). Performing a number of simulations with closely spaced equilibrium positions allows us to build up a free energy profile. In SMD, a similar harmonic potential is applied to the molecule of interest, but, in this equilibrium, position moves over time. Monitoring the force applied to the molecule allows for the calculation of the free energy profile. Both of these methods have been used to calculate the surface \[63,70\] and interfacial \[71–73\] adsorption strengths for peptides and proteins.

3. Hydrophobins

Probably the most commonly investigated classes of protein biosurfactants are the hydrophobins \[74\], which are expressed by some strains of filamentous fungi. These are relatively small proteins that are characterised by a conserved pattern of disulfide bonds \[75\]. Commonly, they also have a highly amphiphilic surface structure, which typically includes a large hydrophobic patch (Figure 2), which makes them highly surface active. They fulfill a number of functions typically associated with adsorption onto interfaces and surfaces. Some hydrophobins are used as surfactants to reduce the air–water surface tension, making it easier for fungi to release sprouting bodies such as spores or hyphae. Other hydrophobins are used to mediate adsorption onto hydrophobic surfaces \[76,77\]. Hydrophobins have traditionally been divided into two-classes depending on whether they form fibrillar structures (rodlets) at interfaces (class-I) or not (class-II) \[74\]. Commonly studied examples of hydrophobins include the class-II hydrophobins HFBI and HFBII and the class-I hydrophobins SC3 and EAS. As well as fungal hydrophobins similar proteins have been identified in bacteria, where they play a role in the formation of biofilms \[78\].

The surface activity of hydrophobins is also being exploited in man-made materials and systems \[15\]. Their high surface activity, combined with their biocompatibility, has led to their use as emulsifiers in foods foams and emulsions \[17,18\]. Hydrophobins have also been used as biocompatible coatings on drug particles and as epitaxial agents for the preparation of graphene and other two-dimensional materials \[27,28\].

Understanding the biological function and materials’ applications of hydrophobins requires knowledge of their structure, in particular at surfaces and interfaces. The relatively small size and simple structure of the hydrophobins (in comparison to other biosurfactants) has made them attractive targets for molecular simulation studies. As many of the biological functions and applications of hydrophobins revolve around their behaviour at interfaces and surfaces, most simulation studies have focused on this.
3.1. Hydrophobin Behaviour at Fluid (Air–Water or Oil–Water) Interfaces

The earliest MD study of a hydrophobin investigated the adsorption of SC3 onto liquid interfaces [80]; however, this used a disulfide bond pattern that was later found to be incorrect. Subsequent studies have used structures derived from protein crystallography or solution NMR, so are more representative of experimental systems. Major focuses of these studies include protein orientation at interfaces, changes in protein conformation, and the interfacial adsorption strength.

The conventional picture of hydrophobins as having a hydrophobic face that drives interfacial adsorption leads to the assumption that this patch points into the hydrophobic media. This has been demonstrated by atomistic MD simulations of the class-II hydrophobins HFBI [81,82] and HFBII [83] at the air–water interface. In both cases, the proteins adsorb onto the air–water interface with the patch pointing into the vacuum (air) region. This orientation of the protein maximises the contact between the hydrophobic patch and the non-aqueous component; an alternative orientation where the patch is oriented in the plane of the interface would maximise the decrease in air–water interfacial area, minimising the interfacial free energy. For HFBI, orientation of the patch towards the hydrophobic component is also seen for the water–decane interface [81] but on a phospholipid (DPPC) bilayer the patch is oriented in the plane of the bilayer. This is due to interactions between the hydrophilic head groups of the lipids with the larger polar regions of the protein surface and demonstrates the ability of hydrophobins to adsorb to both hydrophobic and hydrophilic interfaces. Similar behaviour was also seen for the class-I hydrophobin SC3, where adsorption to both air–water and water–dodecane interface was mediated through a hydrophobic loop [84]. In these simulations, two different protein conformations (taken from simulations of the protein in bulk solution), with the orientation of the hydrophobic loop, relative to the interface, differing by 90 degrees. Independent of the starting orientation and conformation of the protein this loop adsorbed to the interface, showing that it binds onto interfaces through this specific region of the protein.

Due to their rigidity, class-II hydrophobins typically undergo only small changes in conformation at interfaces. MD simulations of both HFBI and HFBII found that the protein secondary and tertiary structures are largely unchanged at the air–water interface [81,83]. This is consistent with circular dichroism [85] and neutron reflectivity measurements [86], which show limited changes in at the air–water interface. A greater change in the structure of HFBI was found at the water–decane interface [81]. This may be due to penetration of oil molecules into the hydrophobic patch, similar to changes observed for α-lactalbumin at the water–octane interface [37]. The change in structure at the water–decane interface was larger than that seen on a DPPC bilayer due to the more hydrophilic environment of the lipid head groups.

Class-I hydrophobins show greater changes in structure at interfaces, which is implicated in their assembly into rodlets [87]. Early studies of the hydrophobin SC3 [84] showed an increase in α-helix content in regions exposed to hydrophobic media. Similar induction of order was found for EAS, where a flexible 26-residue loop adopts a structured (α-helix) conformation at the air–water interface [88]. The flexibility of loop in solution is expected to inhibit self-assembly, while the ordering of this allows for aggregation at interfaces. Changes in protein conformation upon interfacial adsorption has been proposed as a contributing factor in the formation of fibrillar structures for intrinsically disordered proteins [89,90], such as amyloid-beta, so it is logical that similar behaviour would be seen for the rodlet forming class-I hydrophobins. The differences between class-I and class-II hydrophobins were also demonstrated in a direct comparison of HFBII and EAS [91]. Again, these showed only small changes in structure for HFBI while larger conformational changes for EAS. This was used to explain differences in the binding of ions of these two proteins, which rationalised differences in the bimmineralisation behaviour of HFBII and the class-I hydrophobin H*ProteinB [92].

To quantify the effect of the amphiphilic structure of hydrophobins on their interfacial behaviour, a few MD studies have determined the interfacial adsorption strengths. Using a CG model, the adsorption free energies for the class-II hydrophobins HFBI and HFBII at a water–octane interface were calculated [93] using steered molecular dynamics simulations. These were both found
to have adsorption free energies of the order of 60–100 kcal mol$^{-1}$, values similar to those of synthetic nanoparticles [94] and polymers [95]. The two proteins were found to have different affinities for the two solvent components, with HFBII slightly favouring the water and HFBI the oil, which can be rationalised by differences in the sizes of the hydrophobic patches. To explore this, further simulations of HFBII pseudoproteins with different protein solvent interactions were undertaken. Most interestingly, when the interaction with the solvents for each bead in the protein was set to the average value for each bead in the ‘real’ protein, essentially smearing the hydrophobic patch out across the protein, the protein was still attracted to the surface, but it becomes overall slightly hydrophobic. This demonstrates the importance of the amphiphilic surface structure of the protein. Atomistic simulations of the bacterial hydrophobin BslA [96] also found a similar adsorption strength at the air–water interface. Mutation of a single leucine residue in the hydrophobic region (L77K mutation) significantly decreased the adsorption strength. The high adsorption energies and the dependence of these on the size and structure of the hydrophobic region has drawn comparisons between these proteins and amphiphilic Janus particles [97,98].

3.2. Adsorption onto Surfaces

The adsorption of hydrophobins onto diverse solid surfaces has been investigated using MD simulations. Due to their hydrophobic patch and the potential use for surface modification, a number of studies, including some of the earliest hydrophobin simulations [84], investigated their adsorption onto hydrophobic surfaces. Adsorption of HFBII onto silicon was investigated using atomistic MD simulations [99], while Brownian dynamics has been used to investigate the adsorption and oligomerization of HFBI [100]. In both of these studies, the key residues involved in binding the protein to the surface were identified. These were primarily hydrophobic residues, with a number of these being common to both proteins. In the presence of a graphite surface, the proteins would accumulate at the surface, demonstrating that this adsorption is stronger than the solution oligomerization [101].

Self-assembled monolayers (SAM) are commonly used experimentally as a route to control surface chemistry and properties [102]. SAMs can be constructed from a wide range of molecules allowing for the creation of surfaces with different characteristics. Simulation of HFBI on SAMs formed on graphene was used to investigate the adsorption onto hydrophobic (methyl) and hydrophilic (amine) surfaces [103]. Adsorption onto both SAMs was found, demonstrating the dual nature of hydrophobins. While relatively short, these simulations examined the different driving forces for adsorption, suggesting that the adsorption onto hydrophilic surfaces is primarily driven by electrostatic interactions. Later work, over longer timescales and using replica exchange to enhance conformational sampling, demonstrated similar behaviour for HFBII on various SAM surfaces [104]. This also demonstrated the use of the hydrophobic and electric dipoles as simple parameters for describing the surface orientation of the protein; on a methyl-terminated SAM the hydrophobic dipole points towards the surface, while the orientation of the electric dipole depends on the surface charge. Such simple descriptors may prove useful in predicting protein surface behaviour alongside more expensive MD simulations. The ability of hydrophobins to change the character of a surface is not restricted to self-assembled monolayers, as demonstrated by MD simulation of HFBI on poly (dimethyl siloxane) surfaces [105]. As for other hydrophobic surfaces [99,100], the most stable adsorption occurred through the hydrophobic patch, but other adsorption modes were identified. Using MM-PBSA calculations, the relative free energies of these different modes were evaluated, with adsorption free energies for $\sim-200$ kcal mol$^{-1}$ found for states where the patch is in contact with the surface. The adsorption free energy is significantly larger than the adsorption enthalpies (typically $\sim-60$ kcal mol$^{-1}$), suggestive of the importance of solvent and entropic effects (although these were only implicitly accounted for in the MM-PBSA calculations).

As for liquid interfaces, class-II hydrophobins, in both simulation [104] and experimental [106] studies, show only slight changes in conformation on surfaces. Due to their more flexible nature,
class-I hydrophobins show greater changes in conformation. Simulation of EAS at surfaces showed that this binds to silica surfaces in two main modes [107]. The first of these involves a loop between the 3rd and 4th CYS residues, which leads to changes in structure around the amyloidogenic region of the protein, potentially leading to surface fibrillation. As less of the protein is typically in contact with the surface and the contact regions are more flexible, the surface bound water plays a larger role. Surface hydration and water structuring is a key factor in determining the interaction between proteins and surfaces [108,109]. The heterogenous surface structure of EAS was also found to lead to more complex adsorption processes on nanostructured surfaces [110], illustrating the importance of understanding the interplay between protein and surface structure.

4. Surfactin and Other Lipopeptides

A number of organisms, in particular bacteria, express a number of lipopeptides that have surface active behaviour. Probably the most studied of these is the lipopeptide surfactin, expressed by *Bacillus subtilis*. This consists of a cyclic heptapeptide head, with a sequence EL\(^{D}\)LAD\(^{D}\)LL, and a 12–17 carbon atom alkyl tail (Figure 3). Its amphiphilic structure makes it highly surface active, causing a reduction in the air–water interfacial tension from 72 mN m\(^{-1}\) to 27 mN m\(^{-1}\) at a bulk concentration 0.0005% [111]. Surfactin also has a very low critical micelle concentration (\(\sim 0.2 \times 10^{-3}\) mol L\(^{-1}\)), due to its high proportion of hydrophobic components. It plays a key role in the spreading of *Bacillus subtilis*, with significant differences between the biofilms formed by wild type *Bacillus subtilis* and *Bacillus subtilis* with the surfactin gene deleted, which disappear when surfactin is added to the mutant bacterium [112]. Surfactin has also been shown to have antibacterial, antiviral, and antifungal properties [113].

![Figure 3. Surfactin structure. The left hand panel shows chemical structure. The R group is a 12 to 17 carbon atom alkyl chain. The right hand panel shows 3D structure, with Glu and Asp residues highlighted, generated using VMD [79] (from PDB structure 2NPV).](image)

NMR studies [111] suggest that the head group adopts a saddle-like structure, with the two polarizable residues (Glu and Asp) sitting on the same side of the head group (Figure 3). This work also suggested that the flexible tail folds back to interact with the peptide ring which was also found in molecular dynamics simulations [114]. While the Glu and Asp residues being on the same size of the ring suggests that these residues would anchor the molecule at an interface, molecular dynamics simulations show that compact structure caused by the folding of the tail leads to significant tumbling motion with the hydrophilic side chains spending significant time pointing away from the water [115]. This suggests, at low densities, surfactin may act more as (roughly) spherical nanoparticles than traditional surfactants. The mechanism for the decrease in interfacial tension and stabilisation of interfaces by surfactin may then be similar to that in emulsions stabilised by nanoparticles or colloids (Pickering emulsions) [116]. As the density increases, surfactin adopts a more traditional surfactant-like structure at interfaces [117,118].
The plane of the peptide ring lies in the plane of the interface. The interfacial behaviour of surfactin depends on the pH of the aqueous phase, with greater conformational freedom being found for low pH due to the decreased hydrophilicity of the Glu and Asp residues when these become protonated [118]. This can also be seen when surfactin derivatives with methyl ester groups added to the Glu and Asp residues [119].

Due to its amphiphilic nature, surfactin self-assembles into a number of different structures in solution. Experimental investigations have shown that this depends on solution pH and ionic strength. Early cryo-EM work showed the formation of spherical micelles with diameters 5–9 nm and ellipsoidal micelles with typical lengths and widths of 19 nm and 11 nm [120] at pH = 7. At higher pH (9.5), cylindrical micelles were found, with which the addition of electrolytes turned back into spherical micelles. For very high pH (~12), the lactone bond closing the peptide ring was broken and the head becomes linear. Later, EM [121] and SANS [122] work found at neutral pH smaller spherical micelles, with diameters ~5 nm. Molecular dynamics simulations found micelles of around 4 nm diameter for neutral surfactin molecules [118,123]; the discrepancy between this and experimental measurements may be attributed to differences in the definition of micelle sizes between simulation and experiment. For the simulations, it should be noted that the sizes of micelles found were limited by the numbers of molecules in the simulated systems. A more rigorous investigation of micellization of surfactin may require systems containing more molecules, ideally enough to form several micelles [124]. Nonetheless, these simulations were able to investigate the internal structure of the micelles, finding stratified structure with the hydrophobic tails, leucine side chains, peptide ring, and charged side chains forming distinct layers. This differs from SANS results [122], which placed the hydrophobic tails and leucine side chains together in the micelle core. When the Glu and Asp side chains are deprotonated, the sizes of the aggregates decrease due to the increase in electrostatic repulsion between the head groups [118]. Experimental investigations of surfactin micelles also indicate the formation of β-sheets, which has not been observed in simulations. The timescales associated with secondary structure change in proteins can be significant compared to simulation times (the longest simulations above consisted of 250 ns [118]). This may be especially true for cyclic peptides, such as surfactin for which the conformations may be more restricted.

While most attention has been paid to surfactin, other lipopeptides have attracted attention. In recent work, molecular simulation, using both atomistic and coarse-grained models, have been used to investigate the behaviour of a similar lipopeptide from mycosubtilin [125]. Due to the structural similarity to surfactin, its conformational behaviour, both in bulk and at the air–water interface, from atomistic simulations is similar. Using the Martini model, the aggregation of this lipopeptide at the air–water interface was investigated. It was found that this aggregates rapidly, forming half micelles at the interface, suggesting that it is unable to form a stable Langmuir monolayer. As many lipopeptides have cyclic peptide heads, a few studies have investigated the structure [126] and self-assembly [127] of cyclic peptides at interfaces.

5. Ranaspumin

While hydrophobins and lipopeptides, such as surfactin, have structures that, at least on first impression, are amphiphilic, surfactant proteins that lack this surface heterogeneity are possible. One such example is the protein Rsn-2 [128,129], which is found in the foam nests of the tropical frog Engystomops pustulosus. This is the main surfactant component of the foam, which is a mixture of proteins and polysaccharides that is used to hydrate frog eggs. Solution NMR revealed that this protein consists of an α-helix and β-sheet connected by a flexible linker and has a largely polar surface (Figure 4a). Neutron reflectivity and IRRAS measurements show that it retains its secondary structure, but the protein thickness halves at the air–water interface [129]. This suggests a novel interfacial conformational change where the protein opens, with the α-helix and β-strand separating and lying in the plane of the interface (Figure 4b).
The largely polar surface of Rsn-2 inhibits aggregation in solution but provides little driving force for interfacial adsorption. Atomistic molecular dynamics simulations [130] show that Rsn-2 adsorbs onto the air–water interface through its flexible N-terminus, which contains most of the solvent exposed hydrophobic residues. This is similar to the fly-casting mechanism often used in protein recognition [131]. While contact between this region and the interface is necessary, for permanent adsorption, the protein needs to be in specific orientations, possibly to ensure that the subsequent opening transition is possible. The importance of the N-terminus in adsorption was found by studying Rsn-2 mutants with the flexible tails deleted [130]. No adsorption, within the simulation timescales, was found when this region was completely deleted, while deleting only the first three residues of the tail is sufficient to partially inhibit adsorption. This is consistent with dynamic surface tension measurements where the decay in surface tension was significantly slower for the the mutant without the N-terminal tail. Deletion of the hydrophilic C-terminus was also found to affect adsorption, both in simulation and experiment, possibly due to this region playing a role in orienting the protein at the interface. Adsorption onto the water–cyclohexane proceeds in a similar manner, but, unlike at the air–water interface, the opening transition is observed within the simulations. The opening of the protein may be enhanced in this case as oil molecules can partition into the interior of the protein weakening the hydrophobic interactions that hold the solution structure together. Similar enhancement of protein conformational change at interface due to penetration of oil molecules into a protein core has been observed for α-lactalbumin [37]. This is contrary to what are normally observed globular proteins where adsorption onto higher surface tension interfaces [132], in this case the air–water interface, gives rise to greater conformational changes, showing that the specific role of the hydrophobic component should be carefully considered.

![Figure 4](image)

**Figure 4.** (a) structure of Rsn-2 with in cartoon (left) and surface (right) representations. In cartoon representation, the hydrophobic N-terminus loop is highlighted. On the surface representation, blue and red denote hydrophilic and hydrophobic regions, created using PDB code 2WGO; (b) illustration of Rsn-2 opening transition during interfacial adsorption.

Due to the long timescale associated with the unhinging transition, atomistic modelling is unable to rigorously investigate this. To circumvent this CG modelling, using a Go-type model [133] (one bead per residue) was used [134]. Consistent with the atomistic simulations, this showed that the opening transition occurs in two steps, with adsorption onto the interface preceding conformational change. Comparison between the opening transition at the interface and unfolding transition in
solution suggests that the unhinging of the protein corresponds to the initial stages of unfolding, with complete denaturation of the protein inhibited by the interface [135].

While Rsn-2 is the main biosurfactant protein component in the foam, foams prepared only using it have significantly shorter lifetimes than those prepared using the full foam mixture [16]. This suggests that other components, either protein, lipid, or polysaccharide, also play a role in foam stabilisation. Understanding this synergistic effect is key to a complete picture of how the frog foam nest is stabilised, so future work may investigate the interfacial behaviour of the different components and mixtures of these.

6. Conclusions

Over the past few decades, molecular dynamics simulation has emerged as a powerful tool for the investigation of biomolecular systems. Due to its ability to give direct insight into structure and dynamics on a molecular level, it is particularly useful for the investigation of interfacial and surface phenomena, which are otherwise difficult to investigate. As these play a key role in the biological functions and potential applications of protein biosurfactants, MD simulation has, as shown in this review, been of significant utility for investigating these proteins. In particular, MD simulation has revealed molecular features that govern interfacial behaviour of protein biosurfactants, such as how the structure of the hydrophobic patches of hydrophobins affects their interfacial behaviour [93, 96, 99] and how changes in protein conformation relate to interfacial aggregation [88] and function [91, 134].

Despite the success of MD simulation, they face a number of limitations. Possibly the largest outstanding question is the transferability of models used for protein simulations to interfacial and surface environments. These are typically developed for proteins in aqueous solution and so are tuned to reproduce protein structural properties there. Of the commonly used protein force fields, only the Gromos force fields are parameterized against properties (free energies of solvation) that may be related to interfacial behaviour. While transferability to other environments is often assumed, this is generally untested. For fluid interfaces, this is in part due to the lack of experimental information on protein behaviour there. For solid surfaces, adsorption free energies of peptides have been determined and used to test commonly used force fields [136], which has shown that some common protein force can reproduce experimental adsorption free energies. Agreement with experiments can be improved through tuning force fields specifically for interfacial environments [137]. Doing this for other environments is possible but needs accurate experimental information for parameterization and validation. The question of transferability is more acute for CG models. The reduction in resolution of these models means that many effects, such as ordering of water molecules, that are explicit in atomistic models are only accounted for implicitly. This can affect the transferability of CG models to different state points, let alone different environments, so the use of CG models for proteins at surfaces and interfaces should be evaluated carefully.

Other significant limitations include the limited system sizes and time scales that can be investigated. The limitations on system size have largely restricted atomistic MD to the investigation of single proteins at interfaces and surfaces and simulations of aggregation and assembly have only been performed for smaller peptides, such as surfactin. As many protein aspects of protein behaviour at interfaces and surfaces, such as the formation of interfacial structures, involve multiple proteins this has limited the ability of simulation to give insight into these. Restriction to relatively short time scales has limited the investigation of conformational changes of proteins. With increases in computer power, such as the use of general purpose-GPUs [138], and developments in simulation methodologies, the investigation of larger systems over longer times will become possible. Also the coupling of MD simulation with methods, such as finite element analysis, which can access these larger length and time scales should be investigated.

There are a number of avenues in which simulation of protein biosurfactants can proceed in the next few years. As structures of further protein biosurfactants are determined, we are finding more
novel mechanisms for interfacial activity. One interesting example of this is the horse sweat protein latherin [139], which undergoes an unzipping conformational change at interfaces [140]. It may be speculated that similar proteins, such as the PLUNC proteins [141–143], may show similar behavior and that other unforeseen mechanisms may be found in other proteins. These unusual changes in conformation may be exploited in applications or inspire the design of synthetic molecules that also undergo conformational changes at interfaces [144–146].

Closer connection between simulation and experiment and applications is also a key avenue for future research. As mentioned above, most prior work has focused on isolated proteins at interfaces and surfaces, aiming at investigating the underlying biophysics of adsorption. For understanding applications of protein biosurfactants, knowledge of the behavior of protein aggregates is necessary. While simulations of peptide aggregation on interfaces [147–149] and membranes [150,151] have been performed, for larger proteins the times scales involved in their aggregation, particularly taking changes in conformation that often accompany this, are still prohibitive. To investigate larger systems CG models, with caveats described above, or coupling MD with continuum approaches, would provide a feasible route. For proteins, even in bulk solution, such methods are still in their infancy [152] but provide an exciting possibility for extrapolating from the molecular to materials scale.

Author Contributions: The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding: This work was funded through a NUI Galway College of Science Scholarship awarded to S.S.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations
The following abbreviations are used in this manuscript:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>Coarse-Grain</td>
</tr>
<tr>
<td>DPPC</td>
<td>1,2-dipalmitoyl-sn-glycerol-3-phosphocholine</td>
</tr>
<tr>
<td>EM</td>
<td>Electron Microscopy</td>
</tr>
<tr>
<td>GPU</td>
<td>Graphics Processing Unit</td>
</tr>
<tr>
<td>IRRAS</td>
<td>Infra-Red Reflection Adsorption Spectroscopy</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MM-PBSA</td>
<td>Molecular Mechanics Poisson–Boltzmann Solvation Approximation</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>OPLS</td>
<td>Optimised Potentials for Liquid Simulation</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PLUNC</td>
<td>Palate Lung Nasal Epithelial Clone</td>
</tr>
<tr>
<td>SAM</td>
<td>Self-Assembled Monolayer</td>
</tr>
<tr>
<td>SANS</td>
<td>Small Angle Neutron Scattering</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics</td>
</tr>
</tbody>
</table>

References


64. Deighan, M.; Pfauendtner, J. Exhaustively Sampling Peptide Adsorption with Metadynamics. Langmuir 2013, 29, 7999–8009. [CrossRef] [PubMed]


81. Euston, S.R. Molecular simulation of adsorption of hydrophobin HFBI to the air–water, DPPC-water and decane-water interfaces. *Food Hydrocoll.* 2013, 42, 66–74. [CrossRef]


100. Mereghetti, P.; Wade, R.C. Diffusion of hydrophobin proteins in solution and interactions with a graphite surface. BMC Biophys. 2011, 4, 1–11. [CrossRef] [PubMed]


140. Vance, S.J.; Mcdonald, R.E.; Cooper, A.; Smith, B.O.; Kennedy, M.W. The structure of latherin, a surfactant allergen protein from horse sweat and saliva. J. R. Soc. Interface 2013, 10, 20130453. [CrossRef] [PubMed]


