bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Adsorption and conformation of lysozyme and α-lactalbumin at water-octane interface

David L. Cheung*

School of Chemistry, National University of Ireland Galway, Galway, Ireland

Abstract

As they contain both hydrophobic and hydrophilic amino acids proteins will readily adsorb onto interfaces between water and hydrophobic fluids such as oil. This adsorption normally causes changes in protein structure, which can lead to a loss of protein function and irreversible adsorption, leading to the formation of protein interfacial films. While these can be advantageous in some applications (e.g. food technology) in most cases it limits our ability to exploit protein functionality at interfaces. To understand and control protein interfacial adsorption and function it is necessary to understand the microscopic conformation of proteins at liquid interfaces. In this paper molecular dynamics simulations are used to investigate the adsorption and conformation of two similar proteins, lysozyme and α-lactalbumin, at a water-octane interface. While they both adsorb onto the interface α-lactalbumin does so in a specific orientation, mediated by two amphipathic helices, while lysozyme adsorbs in a non-specific manner. Using replica exchange simulations both proteins are found to possess a number of distinct interfacial conformations, with compact states similar to the solution conformation being most common for both proteins. Decomposing the different contributions to the protein energy at oil-water interfaces, suggests that conformational change for α-lactalbumin, unlike lysozyme, is driven by favourable protein-oil interactions. Revealing these differences between the factors that govern conformational change at interfaces in otherwise similar proteins can give insight into the control of protein interfacial adsorption,aggregation, and function.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

I. INTRODUCTION

The adsorption of proteins onto liquid interfaces, such as the air-water or oil-water interfaces, has been harnessed in a number of materials applications. Proteins have long been used as emulsion stabilisers in food technology[1] and recently a number of novel biosurfactant proteins have been identified that have potential for use in functional foods with attractive properties[2]. Protein layers at fluid interfaces have also been exploited as templates for the growth of inorganic materials through biomineralization[3]. Interfacial adsorption of proteins is also important in a number of biological contexts. One diverse group are protein biosurfactants[4,5] which act to reduce surface tension and stabilised foams and emulsions are used by a number of organisms. These include the hydrophobins[6], amphiphilic proteins that are expressed by filamentous fungi during the formation of aerial structures such as spores or hyphae, ranaspumin, that is used by the tropical frog *Engystomops pustulosus* in the production of foam nests[7], and latherin[8] which is found in horse sweat. More sophisticated interfacial function is exhibited by lipases[9], enzymes that break down fats and which become active only when adsorbed onto oil-water interfaces, and lung surfactant proteins[10] SP-A and SP-D that act as part of the immune system.

When proteins adsorb onto liquid interfaces they often undergo conformational change[11,12]. This is driven by the hydrophobic amino acids, which typically form the core of the protein, partitioning into the hydrophobic medium. For biosurfactants or other proteins that have evolved to function in interfacial environments this conformational change is often an intrinsic part of their function, such as the lid-opening seen in lipases or unhinging transition exhibited by ranaspumin[13]. In the more general case of proteins for which interfacial adsorption is not part of their function this is typically associated with a loss of activity, which limits the applications of proteins in interfacial environments. The unfolding of proteins at interfaces is also responsible for strong, often essentially irreversible adsorption, which can lead to loss of protein material during processing (e.g. in the biopharmaceutical industry).

As understanding the factors that drive interfacial adsorption and conformational change and how to control this process is important in a number of biological and technological areas, this has been the subject of considerable investigation with both experimental and theoretical approaches. X-ray[14] or neutron reflectivity[15] can examine the size of proteins, in particular the layer thickness. Neutron reflectivity in particular strongly suggested the

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

unhinging transition of the biosurfactant Rsn2 at the air-water interface[13] and changes from mono- to multilayer adsorption for protein mixtures. Information on the orientation of secondary structure elements (helices or strands) at interfaces can be obtained from infrared spectroscopy. Secondary structure at interfaces has been investigated using refractive index matched emulsion circular dichroism[16] (RIME CD) and synchrotron radiation circular dichroism[17] (SRCD). Formation of larger-scale structures has been investigated using AFM, tensiometry, and rheological measurements (e.g. passive particle probe tracking)[18]. While these methods cannot directly investigate protein structure different aggregation behaviour can be related changes in protein conformation.

Complementing these experimental studies in recent years a number of groups have used molecular dynamics simulations to investigate protein interfacial adsorption and structure. Many of these have focused on the initial stages of interfacial adsorption, aiming to determine the factors that mediate protein adsorption[19–21]. Typically this has been shown to be driven by hydrophobic residues. Molecular dynamics simulations have also been used to investigate the interactions between proteins and inorganic ions at interfaces, giving insight into the process of biomineralization at interfaces[22,23]. Simulations have also been used to determine the adsorption strengths of proteins to liquid interfaces. These have shown that amphiphilic proteins have adsorption free energies $\sim 100\ k_{\mathrm{B}}T$[24,25], essentially showing irreversible adsorption to interfaces. For smaller proteins and peptides this has been measured to be smaller, typically $\sim 10\ k_{\mathrm{B}}T$[26,27], but still significant.

Due to the associated timescales conformational change of proteins at interfaces has been less studied. For small peptides it is possible to examine changes in secondary structure associated with interfacial adsorption within timescales that can be accessed in simulation[28]. To do this for larger proteins it is necessary to use advanced simulation techniques, such as replica exchange[29] or metadynamics[30]. Recently replica exchange simulations on lysozyme at the water-dichloroethane interface showed that this can adopt a number of conformations of similar free energy[31]. Simulations of peptides derived from myoglobin also found that these exhibit a number of different interfacial conformations, with the preferred orientation for the two peptides investigated being consistent with their differing emulsification behaviour[32].

To understand protein adsorption and conformations at interfaces it is necessary to have a microscopic picture of this process, for example identifying which regions of the protein undergo changes in structure. While experimental measurements SRCD or RIME CD can

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determine the overall amounts of secondary structure in a protein they cannot determine which regions of the protein undergo these changes. Understanding this will give detailed information about the mechanisms of conformational change and how this is related to protein structure. SRCD measurements on a number of proteins have shown that while the secondary structure of proteins is different at interfaces compared to bulk solution, they often retain a large amount of ordered secondary structure[33–35]. In most cases the distribution of this secondary structure and whether this is the same as in solution is not clear. Additionally previous simulation work suggests that even globular proteins can exhibit a number of distinct conformations at interfaces, experimental measurements are likely to give information averaged over the ensemble of different interfacial conformations.

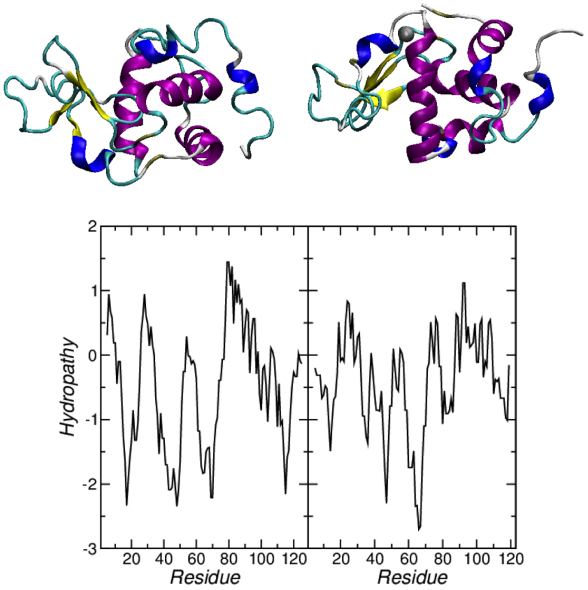

FIG. 1. (a) Structures of LSZ (left) and αLa (right). (b) Kyte-Doolite hydropathy plots for LSZ (left) and αLa (right). Generated using the ExPASy webserver[36] (http://web.expasy.org/cgi-bin/protscale/protscale.pl)

In this paper the conformations of two experimentally studied proteins, lysozyme and α-lactalbumin (Figure 1(a)), at a water-octane interface (which may be considered as an archetypal water-oil interface) are investigated. LSZ and αLa are homologous proteins and are of similar sizes (LSZ has 129 residues, αLa 123). Analysis of their hydropathy plots (Figure 1(b)) show that these have similar patterns of hydrophobic and hydrophilic residues.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Previous work using SRCD on these proteins at oil-water interfaces have shown that these exhibit differing changes in secondary structure at interfaces, with the proportion of α-helix increasing for α-lactalbumin[34] and decreasing for lysozyme[35]. Molecular dynamics simulations are used to investigate how these adsorb at interfaces, and through the use of replica exchange simulations, characterise (at least the initial stages) of conformational change at oil-water interfaces.

II. SIMULATION DETAILS

A. Simulated systems

The simulated systems consisted of a single protein, either lysozyme or α-lactalbumin, and a preequilibrated water-octane interface. Initial protein structures were taken from the protein database, 1E8L for LSZ[37] and 1HFZ for α-lactalbumin[38]. For α-lactalbumin the experimental structure contained a V90M mutation; this was reversed in VMD[39] using the psfgen plugin. Charges on polarizable residues were set appropriately for pH=7. Chloride or sodium counter ions were added as needed to neutralise the system. The proteins were placed in the water slab, approximately 40 Åfrom the interface. For each protein three starting orientations were used (shown in Fig. 2), with the long axis aligned along the x, y, and z axes respectively. Simulations from each of the different starting orientations will denoted as Rotx, Roty, and Rotz in the rest of this paper.

B. Methodology

The simulations were performed using the Gromacs molecular dynamics package[40,41]. For the REST simulations version 4.6.3 using a modified version of PLUMED[42,43] was employed, with Gromacs 5.0 used for the other simulations. Standard gromacs tools were used to create the simulation input files. The Gromos54a7 force field[44] was used for modelling the proteins and along with SPC water[45]. The Gromos family of force fields have been parameterised against solvation free energies for amino acid sidechains so may be expected to give a good description of the partitioning of these between water and hydrophobic media and have been used in a number of previous studies of proteins at liquid interfaces.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

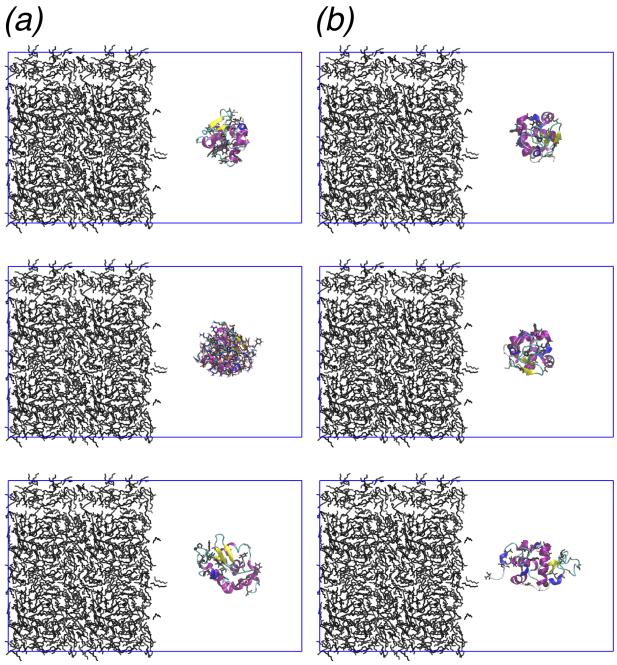

FIG. 2. Initial configurations for (a) LSZ and (b) α-lactalbumin simulations. Long axis orientated along x, y, and z directions (top to bottom). Exposed hydrophobic residues (with solvent accessible surface areas greater than 30 Å^2 in crystal structure) highlighted in black. Water omitted for clarity.

Van der Waals and real-space electrostatic interactions were truncated at 10 Å. Long-range electrostatic interactions were accounted for using PME[46] with a reciprocal space grid of $56 \times 56 \times 96$

The simulations were performed in the NpT-ensemble with the temperature controlled using a velocity rescaling algorithm[47] (relaxation time 0.1 ps) and pressure was controlled using the Parrinello-Rahman barostat, with relaxation time 2 ps. Bond lengths were constrained using the LINCS algorithm[48]. All systems were energy minimized using the steepest descents algorithm, followed by a short NVT simulations (20 ps) with the heavy atoms in the proteins constrained to their initial positions. Following this short NVT and NpT simulations (both 20 ps), without constraints on heavy atoms, were performed.

Simulations were performed in two stages for each of the proteins. For the three different starting conformations a standard NpT MD simulation, with the protein starting in the water region, were performed. These simulations were used to investigate the initial interfacial adsorption and its dependence on the protein orientation. Following this replica

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

exchange with solute tempering (REST) simulations[43,49] were performed to improve sampling over different protein conformations. This is a variation on replica exchange molecular dynamics[50] where the temperature of only a subset of the system, in this case the protein, varies between replicas. This allows for the use of a smaller number of replicas compared to standard REMD. The temperature scaling is effected by scaling the protein-protein and protein-system interactions; the potential energy is given by

$$E_i = \beta_i E_{pp} + \beta_i^{1/2} E_{ps} + E_{ss} \tag{1}$$

where E_{pp}, E_{ps}, and E_{ss} are the protein-protein, protein-solvent (both water and octane), and solvent-solvent interaction and the scaling factor $\beta_i = T_0/T_i$. In these simulations 8 replicas were used with temperatures (scaling factors) of 298 K (1.0), 307.71 K (0.968), 317.74 K (0.938), 328.09 K (0.908), 338.79 K (0.880), 349.83 K (0.852), 361.23 K (0.825), and 373 K (0.80). NPT simulations performed at the highest temperature showed that this was sufficient for both proteins to unfold. For both LSZ and αLa two REST simulations, starting from the Rotx and Rotz NpT simulations, were performed.

The standard NpT and REST simulations were run for 500 ns and 400 ns respectively, with coordinates were saved every 10 ps for later analysis. Prior to the REST simulations NpT simulations of up to 500 ns were performed at each of these temperatures, starting from the end of the 298 K NpT simulation.

C. Analysis

The simulations were analysed using a combination of custom written python scripts (using the MDAnalysis library[51]), VMD[39], and standard Gromacs utilities. Simulation snapshots were generated using VMD, with the protein secondary structure determined using the STRIDE algorithm[52]. Protein conformations were clustered using the Gromacs *g_cluster* utility according to the Gromos criteria[53], with $R_{cut} = 3$ Å. The potential energy contributions were calculated using the rerun option of *mdrun*.

The position of the oil-water interface was determined from the Gibbs dividing surfaces for the water ($z_{interface}^{water}$) and octane ($z_{interface}^{octane}$) as $z_{interface} = (z_{interface}^{water} + z_{interface}^{octane})/2$. The location of the dividing surfaces for the water and octane were found following Vink *et al*[54]. To estimate the surface area, a grid of points is placed on the water-octane interface, defined

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by the Gibbs surface with the area being estimated from the number of grid points that were within a cutoff distance from a protein atom, given by the van der Waals radius+1.4 Å(van der Waals radius of water oxygen). The partition free energy was estimated by summing the water-oil transfer free energies for each residue whose side-chain centre of mass is in the octane ($z < z_{interface}$), using the water-cyclohexane transfer free energies of the amino acid sidechains[55].

Following previous work the number of hydrophobic contacts and the number of protein-octane contacts were determined using[21]

$$N_{contacts} = \sum_{i<j} \frac{1 - \left(\frac{r}{r_0}\right)^6}{1 - \left(\frac{r}{r_0}\right)^{12}} \tag{2}$$

where $r_0 = 5$ Å. For hydrophobic contacts within the protein the sum runs over the C_γ atoms of the hydrophobic residues. For protein-octane contacts the sum runs over C_γ atoms of the hydrophobic residues and all the united carbon atoms in the octane molecules, with the results for each octane molecule divided by eight. The switching function in Eqn 2 is defined so that it goes to 1 at small r and 0 at large r.

III. LYSOZYME AND α-LACTALBUMIN ADSORB TO WATER-OCTANE INTERFACE IN DIFFERENT MANNERS

Shown in Fig. 3(a) is the protein centre-of-mass-interface separation for LSZ. Independent of starting orientation the protein adsorbs onto the interface within $\sim$ 20 ns. Examination of the separations of individual residues and the interface shows that these differ markedly for the three simulation runs (Figure 3(b), suggesting the adsorption is essentially non-specific. These typically include clusters of residues around surface exposed hydrophobic residues. This leads to a number of hydrophilic amino acids being drawn into the oil phase. Simulations started from the Rotx and Roty orientations show similar numbers of residues in contact with the interface; these residues include both the N and C-termini of the protein. Between these the residues in contact with the interface differ between these two simulations (Figure 3(c)). For the Rotx simulation a cluster of residues around the hydrophobic F34 and F38 residues moves into the oil. This region is more distant from the interface in the Roty simulation. Instead the section G102-N106 enters the octane. The Rotz simulation

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

has fewer residues in contact with the interface. These are also quite distinct from those for Rotx and Roty with contact made with the interface between residues N77 and S85. In all cases the residues in contact with the interface correspond to hydrophobic regions of the protein, shown by peaks in the hydropathy plot for LSZ (Figure 1(b)).

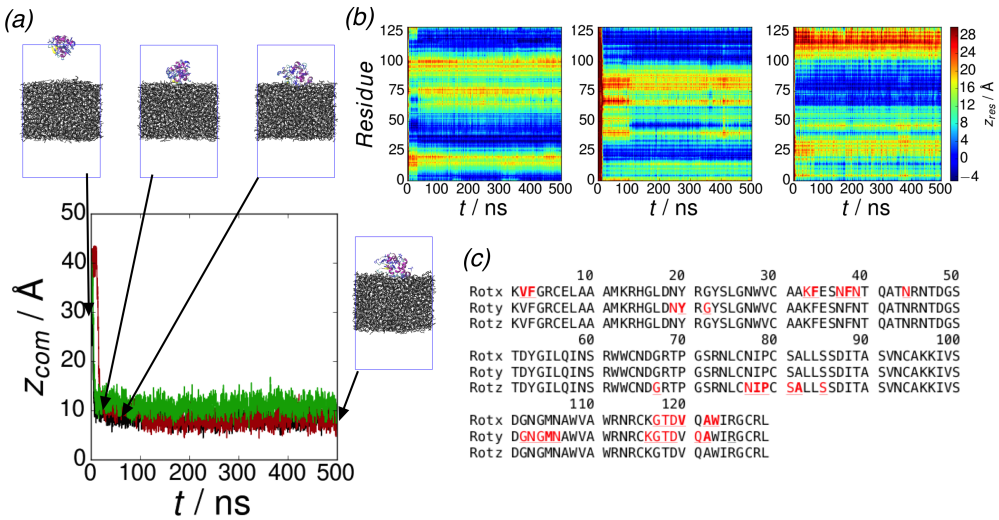

FIG. 3. Approach of LSZ to water-octane interface. (a) Protein centre-of-mass-interface separation. Rotx, Roty, and Rotz starting configurations denoted by black, red, and green lines respectively. Insets show simulation snapshots for Rotx at t =0, t = 10 ns, t = 50 ns, and t = 500 ns (b) Residue centre-of-mass-interface separations for (left to right) Rotx, Roty, and Rotz starting configurations. (c) Annotated sequence with residues in the octane slab ($z_{res} < 0$ for the final 100 ns) underlined. Hydrophobic residues in octane slab shown in bold.

Unlike LSZ αLa can show a significant lag time for interfacial adsorption (Fig. 4(a)). Depending on the starting orientation attachment to the interface occurs after $\sim$ 40 ns. Examination of the individual residue separations (Figure 4(b)) suggests that this is associated with reorientation of the protein from its initial orientation to one which favours interfacial adsorption. For all the starting orientations similar sets of residues are in contact with the interface (Figure 4(c)). This similarity between the residues involved in interfacial attachment suggests the this is more strongly dependent on the protein than the non-specific adsorption seen for lysozyme. αLa contains two amphiphilic helices, helix A (E1-L15) and helix C (F80-G100) that mediate its binding onto membranes[56]; from analysis of the simulations it can be seen that helix C is also involved in the adsorption of αLa onto the water-octane interface,

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with a region just after helix A also involved. The initial reorientation of the protein can be seen in the simulation snapshots (Fig. 4(a)). At short times the protein reorients near the centre of the water slab until a possible binding orientation is found. Once it is oriented correctly attachment to the interface occurs over similar timescales as for LSZ. The final adsorbed orientation and structure is similar for all the starting orientations. Neither LSZ or αLa exhibit transient interfacial attachment observed in molecular dynamics simulations of other proteins[19,21].

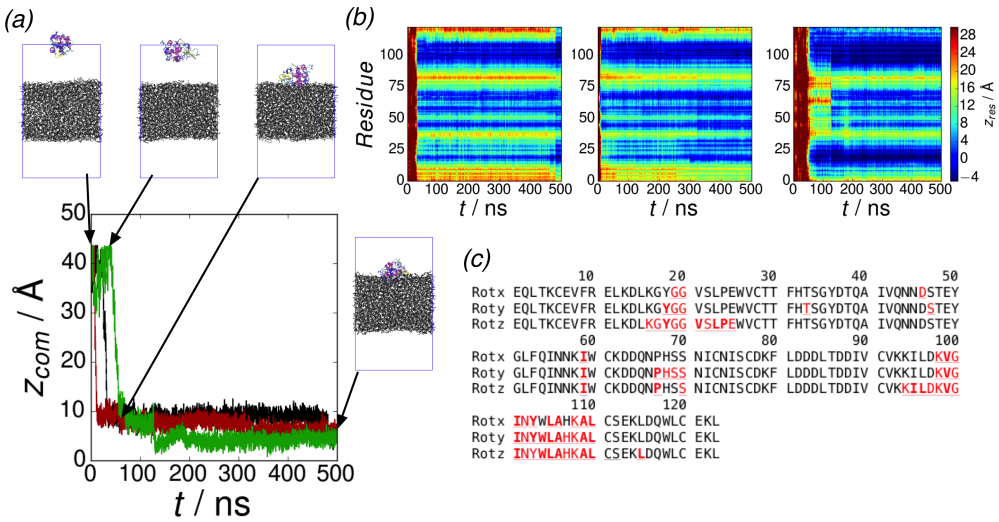

FIG. 4. Approach of αLa to water-octane interface. (a) Protein centre-of-mass-interface separation. Rotx, Roty, and Rotz starting configurations denoted by black, red, and green lines respectively. Insets show simulation snapshots for Rotx at t =0, t = 25 ns, t = 50 ns, and t = 500 ns (b) Residue centre-of-mass-interface separations for (left to right) Rotx, Roty, and Rotz starting configurations. (c) Annotated sequence with residues in the octane slab ($z_{res} < 0$ for the final 100 ns) underlined. Hydrophobic residues in octane slab shown in bold.

During the adsorption only slight changes in the secondary structure are seen for LSZ (Figure 5(a)). In some of the simulation runs (Rotx and Rotz) there is some loss of α-helix content towards C-terminal end of the protein (residues120-125). From the residue centre-of-mass positions (Figure 3(b)) there is no relationship between this and the location of these regions relative to the interface.

Similarly little change is seen in the radius of gyration (Figure 5(b)). For all simulations the overall radius of gyration remains essentially unchanged suggesting that protein shape

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

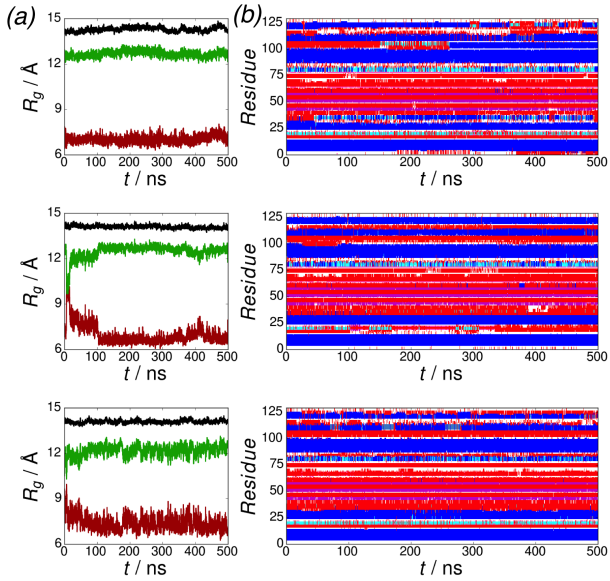

FIG. 5. Protein structure during interfacial adsorption for LSZ at 298 K. (a) Radius of gyration for (top to bottom) Rotx, Roty, and Rotz simulations. R_g, R_g^z, and R_g^{xy} denoted by black, red, and green lines respectively. (b) Secondary structure for (top to bottom) Rotx, Roty, and Rotz simulations.Blue denotes α-helix, cyan 3_{10}-helix, red turn, pink bend, magenta extended, and white random coil.

doesn't change during adsorption. The perpendicular and in plane components of the radius of gyration show some change, consistent with protein reorientation during adsorption.

For αLa more variation is seen in the secondary structure (Figure 6(a)). While the membrane binding helices A and C are largely unchanged during this initial adsorption the other helices change size during this. Helix D shows the largest change, with this shrinking or disappearing altogether. This helix lies in a region that is closer to the interface, with a substantial portion of this penetrating into the oil-phase. Compared to helices A and C it is less amphiphililic so this penetration into the oil phase distorts this helix.

Similar to LSZ the tertiary structure shows little change during interfacial adsorption (Figure 6(b)). The overall R_g is largely constant during adsorption, while R_g^{xy} and R_g^z change consistent with the protein reorienting during adsorption.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

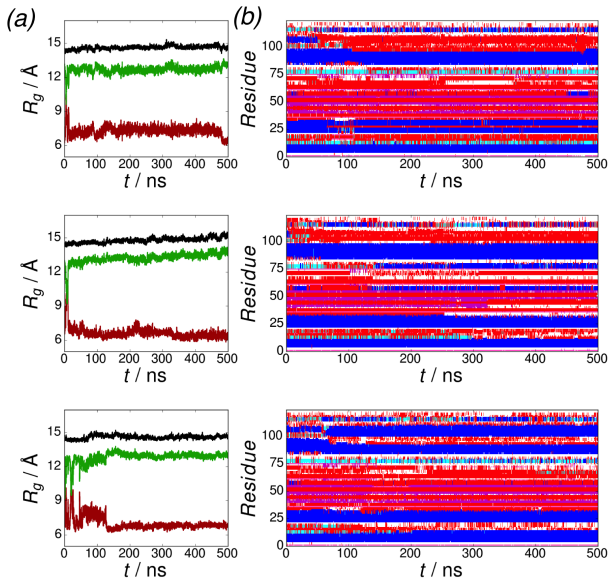

FIG. 6. Protein structure during interfacial adsorption for αLa at 298 K. (a) Radius of gyration for (top to bottom) Rotx, Roty, and Rotz simulations. R_g, R_g^z, and R_g^{xy} denoted by black, red, and green lines respectively. (b) Secondary structure for (top to bottom) Rotx, Roty, and Rotz simulations.Blue denotes α-helix, cyan 3_{10}-helix, red turn, pink bend, magenta extended, and white random coil.

IV. PROTEIN CONFORMATIONAL CHANGE AT WATER-OCTANE INTERFACE

The REST simulations reveal that both LSZ and αLa exhibit changes in secondary and tertiary structure at the octane-water interface. For LSZ the radius of gyration, both overall and the components resolved perpendicular and in the plane of the interface, shows rapid jumps between two sets of values (Figure 7(a)). This suggests that LSZ can adopt (at least) two different structures at the interface. For one of these the radius of gyration is similar to the value for bulk solution suggesting that this corresponds to a compact, globular structure similar to the solution state. The second has significantly larger R_g^{xy} and smaller R_g^z showing that the protein has expanded in the plane of the interface. For both simulations the compact state is significantly more common. This is consistent with simulations of LSZ at the water-DCE interface[31], which also found that LSZ preferentially adopts a compact state but more expanded conformations were observed during the simulations. It is also similar to recent

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

simulations of myoglobin-derived peptides, where both compact and globular states where found[32]. The lack of significant tertiary structure change is also seen in the locations of the residues (Figure 7(b)) which are essentially the same as for the NpT simulations.

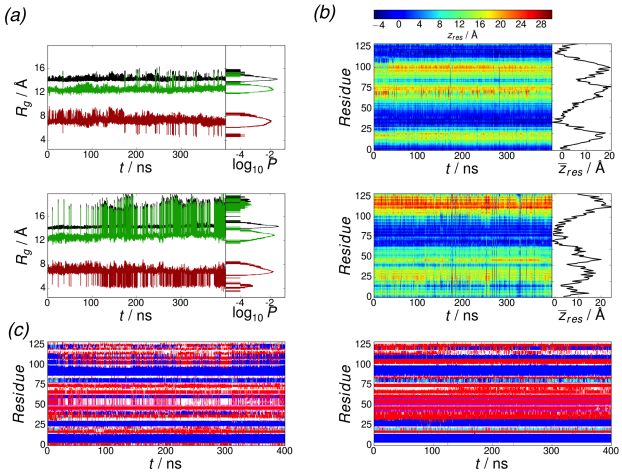

FIG. 7. Structure of LSZ from REST simulations. (a) Radius of gyration for Rotx (top) and Rotz (bottom) simulations. Black, red, and green lines denote R_g, R_g^z, and R_g^{xy} respectively. Right hand panels show probability histograms for R_g, R_g^z, and R_g^{xy} (symbols as in main panel) calculated over final 100 ns of each simulation. (b) Residue-interface separations for Rotx (top) and Rotz (bottom) simulations. Right hand panels show average residue-interface separations calculated over final 100 ns of simulations. (c) Secondary structure distributions for Rotx (left) and Rotz (right) simulations. Blue denotes α-helix, cyan 3_{10}-helix, red turn, pink bend, magenta extended, and white random coil.

These different conformations also exhibit different secondary structure compositions. While the compact state has similar secondary structure composition to the solution state, the α-helix content of the expanded state is lower, which leads to a decrease in the α-helix content (Table I).. This is particularly noticeable for the C-terminal end of the protein, where these helical regions typically become turns. Despite the similarity between the secondary and tertiary structures between the two different REST simulations different sections of the protein are in contact with the interface. These regions are the same as in the NpT simulations, indicating that we do not see reorientation of the protein. Again this suggests that for LSZ adsorption to the water-octane interface is non-specific.

While both simulation runs for αLa are started from similar conformations, the protein

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

structures found in these simulations are noticibly different (Figure 8). Starting from the Rotx conformation only a single compact state is found, while both compact and extended states are found from the Rotz conformation (Figure 8(a)). Compared to LSZ the extended state for αLa is smaller ($R_G \sim 17$ Å compared to ~ 20 Å for LSZ). For the Rotz simulation these changes in structure can be seen in the residue centre-of-mass positions (Figure 8(b)). For the majority of the protein the secondary is conserved throughout the simulation and is similar for both runs (Figure 8(c)). In the Rotz simulation helix-B becomes destabilised and some changes to the α-helix content towards the C-terminus are seen in both simulations. The residue-interface separations for these regions differ between the two simulation runs with loss or destabilisation of α-helical regions being more common when these are closer to the interface. Helices B and D are less amphiphilic than A and C that are largely unchanged at the interface, so when these regions become close to the interface the helices will become distorted to allow hydrophobic sidechains to partition into the oil.

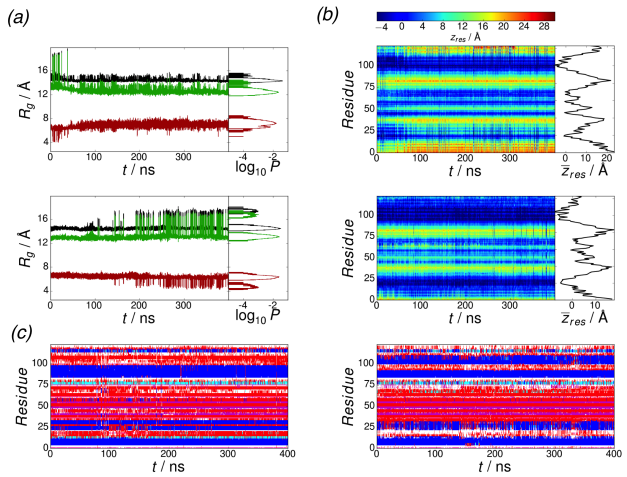

FIG. 8. Structure of αLa from REST simulations. (a) Radius of gyration for Rotx (top) and Rotz (bottom) simulations. Black, red, and green lines denote R_g, R_g^z, and R_g^{xy} respectively. Right hand panels show probability histograms for R_g, R_g^z, and R_g^{xy} (symbols as in main panel) calculated over final 100 ns of each simulation. (b) Residue-interface separations for Rotx (top) and Rotz (bottom) simulations. Right hand panels show average residue-interface separations calculated over final 100 ns of simulations. (c) Secondary structure distributions for Rotx (left) and Rotz (right) simulations. Blue denotes α-helix, cyan 3_{10}-helix, red turn, pink bend, magenta extended, and white random coil.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Protein	Run	α-helix	3_{10}-helix	turn	extended	bridge	coil
	Rotx	36.1 %	4.0 %	34.8 %	6.3 %	2.2 %	16.5 %
LSZ	Rotz	31.3 %	3.7 %	41.7 %	8.6 %	1.3 %	13.4 %
	Average	33.7 %	3.8 %	38.3 %	7.5 %	1.8 %	15.0 %
	Rotx	31.0 %	4.2 %	39.7 %	6.7 %	3.3 %	15.1 %
αLa	Rotz	23.4 %	4.8 %	39.4 %	6.2 %	2.8 %	23.4 %
	Average	27.2%	4.5 %	39.5 %	6.5 %	3.1 %	19.2 %

TABLE I. Secondary structure composition from REST simulations. LSZ Rotx simulation also has 0.1 % π-helix.

V. LYSOZYME AND α-LACTALBUMIN EXHIBIT MULTIPLE INTERFACIAL CONFORMATIONS

Analysis of the radius of gyration suggests that both LSZ and αLa exhibit multiple conformations at the oil-water interface. To gain insight into these two-dimensional probability distributions $P(R_g^z, R_g^{xy})$ were generated (Figure 9). For both proteins compact, globular states are more common, with more extended states being rarer. Both LSZ and αLa possess a network of disulfide bonds helping them retain their structure. αLa is additionally stabilised by its Ca^{2+} ion. These strong interactions make them resistant to changes in tertiary structure at interfaces, unlike the recently studied case of myoglobin-derived peptide which lack these. From the probability maps most of the conformations for LSZ fall into a single compact state and a scattered set of extended states ($R_g^z = 4.2$ Å and $R_g^{xy} = 18$ Å). There is also a small cluster of conformations (around $R_g^z = 4.5$ Å and $R_g^{xy} = 15$ Å) that may be considered intermediate between the compact and extended states. αLa has two compact states and single set of extended states, clustered around $R_g^z = 4.5$ Å and $R_g^{xy} = 17$ Å.

To get more insight into the different protein conformations we performed a cluster analysis on the last 100 ns of each of the REST simulations for both proteins (20000 protein conformations in total). For LSZ 17 distinct clusters were found but only one of these is significantly occupied (Figure 10). This corresponds to the centre of the cluster of compact states ($R_g^z = 6.16$ Å and $R_g^{xy} = 12.86$ Å). From the radius of gyration of each cluster the first six clusters are all relatively compact. A number of distinct extended states were found,

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

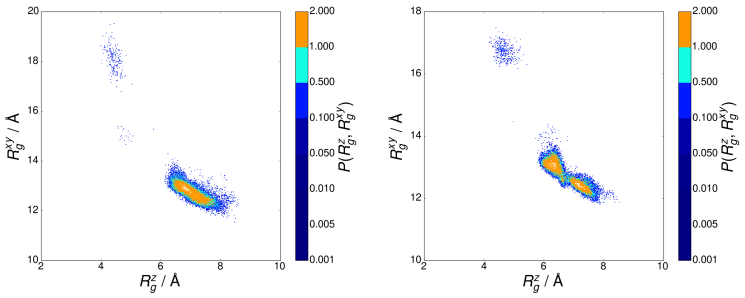

FIG. 9. Probability maps for LSZ (left) and αLa (right).

however, the probability of these are significant;y lower than the compact states. We can estimate the free energy of each cluster from $\Delta F_i = -k_\mathrm{B}T \ln P_i/P_1$ where P_i is the probability of the protein being in the ith cluster. There is a large jump in free energy difference the first and second clusters, illustrating the dominance of the first cluster. The extended states have $\Delta F_i \gtrsim 3$ kcal mol^{-1}.

Despite the differences in tertiary structure the secondary structure is largely the same for the different conformations. With the exception of one cluster, most variation is seen towards the C-terminus, from residue G102 onwards. Most commonly there is a short α-helical region (G102-A107) followed by regions of turn and 3_{10}-helix, as in the first cluster. In the conformations this helix is absent and other α-helix regions are found (e.g. A110-C115 and V120-I124 in the second cluster). Interestingly this holds for most of the extended conformations, suggesting that for the most part these consist of different arrangements of the helical regions. The only cluster that deviates significantly from this is the 9th which has only two α-helical regions (E7-A11 and V92-V99). This corresponds to the small cluster of states between the compact and extended states (Fig. 9).

The cluster analysis for αLa only eight distinct conformations were found (Figure 11). The first two are both compact structures and are approximately equally occupied, consistent with the probability map. There is a smaller free energy difference between compact and extended states for αLa ($\Delta F_i \sim 1.9$ kcal mol^{-1}).

Compared to LSZ more variation in seen the secondary structures for the different clusters. As is seen in the time series of the secondary structure the membrane-binding helices (A and C) are present in almost all the clusters. The other helices are more variable, with one or other of these missing in some of the clusters. Only in the extended conformations

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

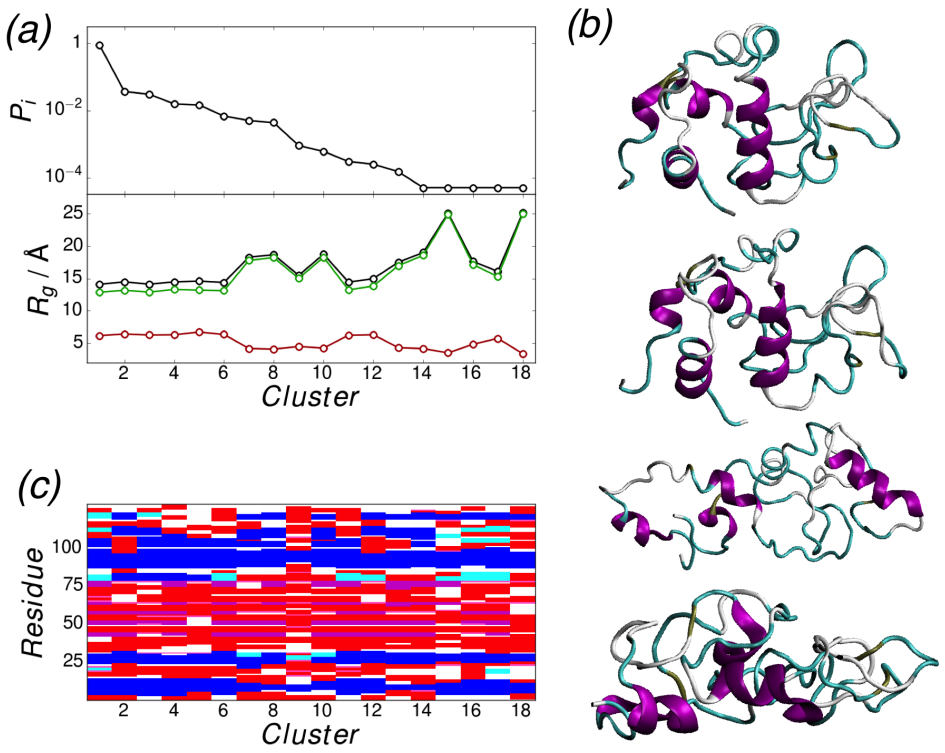

FIG. 10. (a) Cluster probability (top) and radius of gyration (bottom) for each of the cluster average conformations for LSZ. In bottom panel R_g, R_g^z, and R_g^{xy} denoted by black, red, and green lines respectively. (b) Structures of first, second, seventh, and nineth (top to bottom) cluster average structures. (c) Secondary structure for the cluster average structures.

(clusters 4 and 8) are major differences in these areas seen. For cluster 4 neither of the membrane-binding helices are present. Instead helix-D is largely present (Y103-A109) and an additional short α-helix (D38-N44) segment is formed.

Snapshots of these different clusters show clearly how the compact and extended structures differ form each other. While the compact structures are quite similar to each other, the first extended structure is significantly different. It is highly extended in the plane of the interface extending with both α-helices lying flat against the water-octane interface. The second, less common, extended structure has more similarity to the compact conformations in particular towards the N-terminus.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

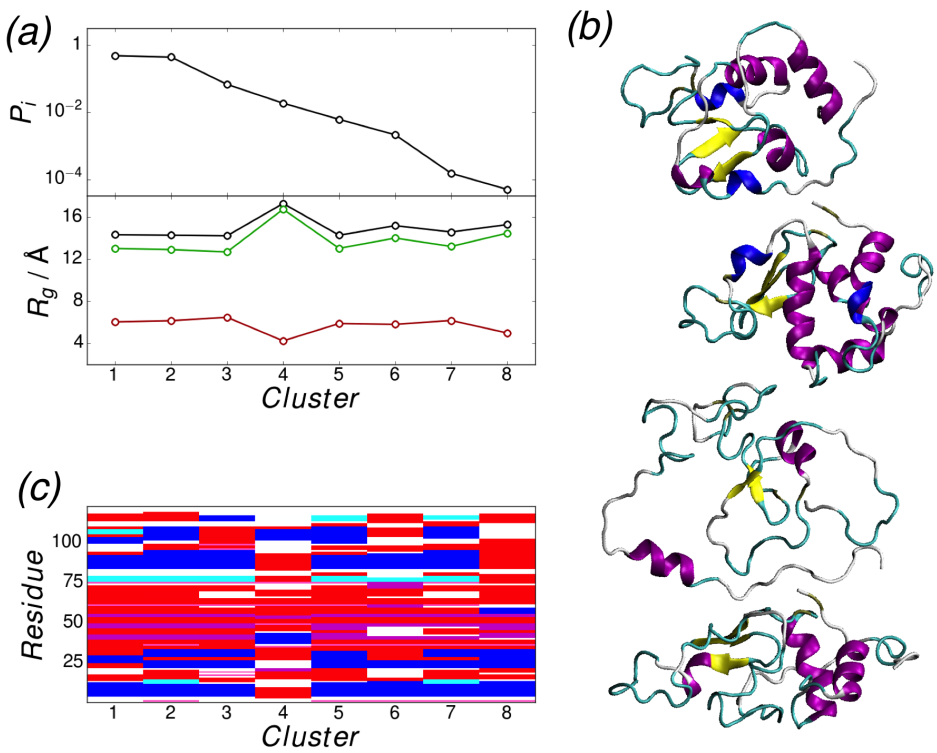

FIG. 11. (a) Cluster probability (top) and radius of gyration (bottom) for each of the cluster average conformations for αLa. In bottom panel R_g, R_g^z, and R_g^{xy} denoted by black, red, and green lines respectively. (b) Structures of first, second and fourth (top to bottom) cluster average structures. (c) Secondary structure for the cluster average structures.

VI. WHAT FACTORS THAT CONTROL CONFORMATIONAL CHANGE?

Changes in protein conformation at water-oil interfaces are driven by a number of effects, including interfacial free energies, partitioning of hydrophobic residues, and differing protein-solvent interactions between the two different solvent components. To gain insight into these different effects we can estimate the differences between these for the extended and compact states, which reveals some differences between the interfacial behaviour of LSZ and αLa. The free energy difference between different interfacial conformations can be estimated from

$$\Delta F = \Delta F_{interface} + \Delta F_{partition} \approx \Delta E_{intra} + \Delta E_{po} + \Delta E_{pw} \tag{3}$$

where the terms on the right hand side are the changes in interfacial free energy, partition free energy, protein intramolecular energy, protein-octane interaction, and protein-water interaction energies respectively. The interfacial free energy is given by $-\gamma A_o$, where γ is

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the water-octane interfacial tension and A_o is the area occupied by protein.

Estimates for these different contributions are presented in Table II. For both proteins the interfacial and partition free energies favour extended states at the interface, as may be expected. The extended states occupy a larger area at the interface, giving rise to a lower interfacial free energy compared to the compact state. In the extended states the hydrophobic residues are able to partition more easily into the oil phase decreasing the partition free energy. The change in both $F_{interface}$ and $F_{partition}$ on going from compact to extended states is larger for LSZ than αLa. For both LSZ and αLa the intramolecular interaction favours the compact states.

Protein	Conformation	N	A_o	$-\gamma A_o$	$\Delta F_{partition}$	E_{intra}	E_{po}
LSZ	Compact	1981	269±2	-20.1±0.15	24.7±0.6	-3118±1.5	-114.7±0.8
	Expanded	21	740±12	-54.8±0.9	-34±4	-3100±20	-120±15
αLa	Compact	1952	366±2	-27.3±0.15	-3.0±0.3	-3737±3	-166.3±0.9
	Expanded	50	676±8	-50.4±0.6	-32±2.4	-3590±12	-290±1.6

TABLE II. Contributions to energies of compact and extended states for LSZ and αLa. For both proteins conformations with $R_g \leq 16$ Åwere considered compact and $R_g > 16$ Åwere considered extended. Averages calculated over final 100 ns of each REST simulation. Areas in Å^2, energies in kcal mol^{-1}. Uncertainties estimated as standard errors.

While most contributions behave similarly for LSZ and αLa the protein-octane interaction behaves differently for these proteins. For LSZ this is essentially unchanged between the compact and extended states while it is significantly lower for αLa. The protein-octane interaction in the compact state for αLa is more favourable than LSZ, consistent with the lower partition free energy.

To understand the relationship between the intramolecular and protein-octane interaction energies and the protein conformation we can examine the probability distributions $P(R_g, E_{intra})$ and $P(R_g, E_{po})$ (Fig. 12). For both LSZ and αLa $P(R_g, E_{intra})$ is similar for the compact states, with a narrow distribution of R_g values and the E_{intra} values spanning a range of a few 100 kcal mol^{-1}. For LSZ these take on a wider range of R_g values compared to αLa; in both cases they span a similar range of energies as in the compact states but are skewed towards higher energies, reflecting the weaker intramolecular interactions for the

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extended states.

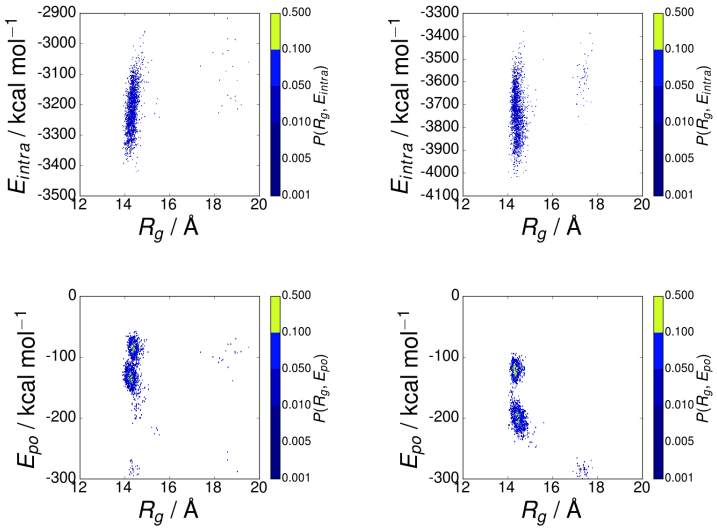

FIG. 12. Probability maps for radius of gyration and protein intramolecular interaction (top) and protein-octane interaction (bottom) for LSZ (left) and αLa (right)

More variation is seen in $P(R_g, E_{po})$. For αLa almost all states are concentrated into three distinct groups, two at lower R_g, corresponding to compact states, and a third extended group. The most extended group has a lower E_{po} than the compact ones showing, as would be intuitively expected, in the extended state allows for more favourable interactions between protein and oil. LSZ shows more groups, consistent with the higher number of conformations found in the cluster analysis. It possesses a number of distinct compact conformations with different protein-octane interaction energy. The extended states typically do not have a lower E_{po} than the compact states. This shows that, unlike αLa, conformation change of LSZ at the water-octane interface is not, at least in its initial stages, driven by favourable interactions between the protein and oil.

When the proteins go from compact to extended states this changes the number of contacts between hydrophobic residues and introduce contacts between octane molecules and hydrophobic residues that normally form the compact protein core (Figure 13). Both of these show bimodal behaviour with frequent transitions between values appropriate for compact and extended states. The number of contacts between hydrophobic residues in the compact states is similar for both LSZ and αLa, while this is lower for LSZ in the extended states.

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

αLa typically has more contacts between hydrophobic residues and octane, reflecting the more favourable E_{po} compared to LSZ.

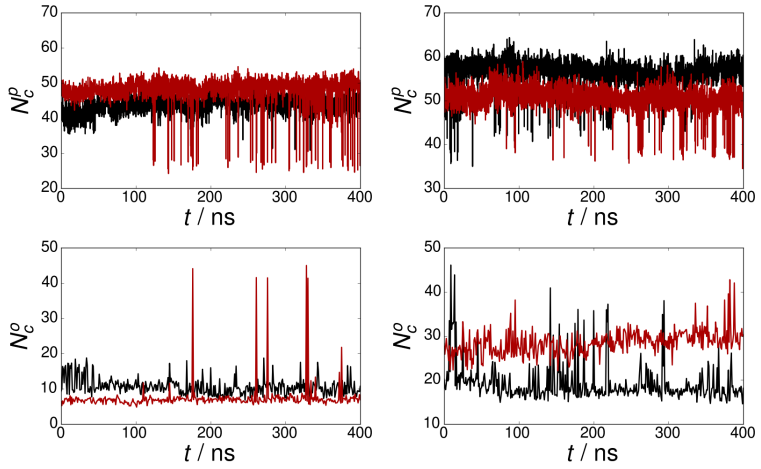

FIG. 13. Number of contacts between hydrophobic residues (top) and hydrophobic residues and octane (bottom) for LSZ (left) and αLa (right).

VII. CONCLUSIONS

Using atomistic molecular dynamics simulations the interfacial adsorption and conformation of two proteins, lysozyme and α-lactalbumin, was investigated. Due to their intrinsic amphiphilicity both proteins adsorb onto the interface. Running simulations with different starting protein orientation suggests that αLa has a preferred orientation for adsorption to the interface, with adsorption occurring through regions near the membrane binding helices. By contrast LSZ adsorbs in a non-specific manner.

Due to their strong intramolecular interactions, including a number of disulfide bonds neither protein undergoes significant changes in structure during this initial adsorption. Using REST simulations we showed that both proteins exhibit compact and extended states. The extended states are expanded in the plane of the interface, with sizes perpendicular to the interface being comparable to a single α-helix. For both proteins the compact states are more common than the extended states, in contrast to recent work investigating myoglobin derived peptides; this reflects the rigidity of these proteins.

Examining the different interfacial conformations reveals some differences between LSZ and αLa. While compact states are more common for both proteins, for LSZ the interfacial

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

conformations are dominated by a single conformation, with a variety of different extended states are also found. Using a cluster analysis we found 17 distinct interfacial conformations for LSZ and estimated free energy difference between the compact and extended states to be $\gtrsim 3$ kcal mol^{-1}. αLa has fewer conformations but these are populated more evenly. Again it is most commonly found in compact but has two highly populated compact states. The extended states are more highly populated than for LSZ with the free energy difference between the compact and extended states being smaller than for LSZ (~ 1.9 kcal mol^{-1}).

As would be expected the extended states are stabilised by the decrease in interfacial free energy, due to the increase in occupied interfacial area, and the partitioning of hydrophobic residues into the oil phase. The compact states are stabilised by the protein intramolecular interactions. For αLa the extended states are additionally stabilised by contacts between hydrophobic residues and octane molecules, leading to a decrease in the protein-octane interaction energy. LSZ in contrast shows no significant difference in protein-octane interaction energy between the compact and extended states. Indeed examining the joint probability distribution $P(R_g, E_{po})$ showed that the lowest protein-octane interaction energy is found for a cluster of compact states.

This insight into the conformations adopted by proteins at interfaces can give new insight into protein interfacial adsorption and aggregation. The contrast between the behaviour of these proteins and myoglobin-derived peptides studied in recent work demonstrates the key role that protein intramolecular interactions play in determining interfacial conformations. While LSZ and αLa are largely found in compact states, the weaker intermolecular interactions in the myoglobin peptides leads to the formation of extended states[32]. Weakening the interactions in LSZ and αLa, for example through reducing some the disulfide bonds, may be used to promote the formation of extended states.

One noticeable difference in behaviour between LSZ and αLa is the interaction between the protein and oil molecules. For αLa the protein-oil interaction is more favourable in the extended state while no significant difference was found for LSZ. This suggests that the effect of the oil phase on the interfacial conformation is dependent on the protein. The destabilisation of the compact state by protein-oil interactions is consistent with recent work on the protein biosurfactant Rsn-2, where the clamshell opening was found at the cyclohexane-water interface but not at the air-water interface due to penetration of oil molecules into the hydrophobic protein core[21].

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

While SRCD studies showed that the proportion of α-helix for αLa increased at the water-alkane interface this was not seen in our simulations. This could be a timescale issue, as the continuous simulation time was 900 ns, which may still not be long enough to observe secondary structure change for these larger proteins. The SRCD experiments were performed at a water-hexadecane interface while we used a water-octane interface in this work; however, as an alkane was used in both cases this is unlikely to have a significant effect. Alternatively interactions between proteins at the interface could also cause changes in the protein secondary structure. The protein interfacial density in this work is about a factor of 5 lower than the lowest density examined in experiment. This could also explain the differences in secondary structure for LSZ; while we see a decrease in the α-helix proportion we do not see an increase in β-sheet proportion. Aggregation of LSZ at the interface could lead to the formation of intermolecular β-sheets that we were unable to observe in this work.

ACKNOWLEDGEMENTS

I wish to acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. The modified version of PLUMED for performing REST simulations was provided by Giovanni Bussi (SISSA).

---

* david.cheung@nuigalway.ie

[1] E. Dickinson, Soft Matter **2**, 642 (2006).

[2] N. R. Stanley-Wall and C. E. MacPhee, Phys. Biol. **12**, 063001 (2015).

[3] A. Schulz, B. M. Liebeck, D. John, A. Heiss, T. Subkowski, and A. Böker, J. Mater. Chem. **21**, 9731 (2011).

[4] A. Cooper and M. W. Kennedy, Biophys. Chem. **151**, 96 (2010).

[5] M. Schor, J. L. Reid, C. E. MacPhee, and N. R. Stanley-Wall, Trends Biochem. Sci. **41**, 610 (2016).

[6] M. B. Linder, Curr. Opin. Colloid Interface Sci. **14**, 356 (2009).

[7] A. Cooper, M. W. Kennedy, R. I. Fleming, E. H. Wilson, H. Videler, D. L. Wokosin, T.-J. Su, R. J. Green, and J. R. Lu, Biophys. J. **88**, 2114 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[8] S. J. Vance, R. E. Mcdonald, A. Cooper, B. O. Smith, and M. W. Kennedy, J. R. Soc. Interface **10**, 20130453 (2013).

[9] P. Reis, K. Holmberg, H. Watzke, M. E. Leser, and R. Miller, Adv. Colloid Interface Sci. **147-148**, 237 (2008).

[10] J. A. Zasadzinski, J. Ding, H. E. Warriner, F. Bringezu, and A. J. Waring, Curr. Opin. Colloid Interface Sci. **6**, 506 (2001).

[11] Y. F. Yano, J. Phys. Condens. Matter **24**, 503101 (2012).

[12] J. L. Zhai, L. Day, M.-I. Aguilar, and T. J. Wooster, Curr. Opin. Colloid Interface Sci. **18**, 257 (2013).

[13] C. D. Mackenzie, B. O. Smith, A. Meister, A. Blume, X. Zhao, J. R. Lu, M. W. Kennedy, and A. Cooper, Biophys. J. **96**, 4984 (2009).

[14] D. Gidalevitz, Z. Huang, and S. A. Rice, Proc. Natl. Acad. Sci. U. S. A. **96**, 2608 (1999).

[15] X. Zhao, F. Pan, and J. R. Lu, J. R. Soc. Interface **6**, S659 (2009).

[16] F. A. Husband, M. J. Garrood, A. R. Mackie, G. R. Burnett, and P. J. Wilde, J. Agric. Food Chem. **49**, 859 (2001).

[17] A. J. Miles and B. A. Wallace, Chem. Soc. Rev. **35**, 39 (2006).

[18] S. Jordens, P. A. Rühs, C. Sieber, L. Isa, P. Fischer, and R. Mezzenga, Langmuir **30**, 10090 (2014).

[19] D. Zare, K. M. McGrath, and J. R. Allison, Biomacromolecules **16**, 1855 (2015).

[20] D. Zare, J. R. Allison, and K. M. McGrath, Biomacromolecules **17**, 1572 (2016).

[21] G. B. Brandani, S. J. Vance, M. Schor, A. Cooper, M. W. Kennedy, B. O. Smith, C. E. MacPhee, and D. L. Cheung, Phys. Chem. Chem. Phys. **19**, 8584 (2017).

[22] A. Schulz, M. Fioroni, M. B. Linder, A. Nessel, M. Bocola, T. Subkowski, U. Schwaneberg, A. Böker, and F. Rodríguez-Ropero, Soft Matter **8**, 11343 (2012).

[23] A. Jain, M. Jochum, and C. Peter, Langmuir **30**, 15486 (2014).

[24] D. L. Cheung, Langmuir **28**, 8730 (2012).

[25] 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, Proc. Natl. Acad. Sci. **112**, 5419 (2015).

[26] C. A. Miller, N. L. Abbott, J. J. D. Pablo, C. A. Miller, N. L. Abbott, and J. J. D. Pablo, Langmuir **25**, 2811 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[27] O. Engin, A. Villa, M. Sayar, and B. Hess, J. Phys. Chem. B **114**, 11093 (2010).

[28] C. Dalgicdir, C. Globisch, C. Peter, and M. Sayar, PLOS Comput. Biol. **11**, e1004328 (2015).

[29] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).

[30] A. Laio and F. L. Gervasio, Reports Prog. Phys. **71**, 126601 (2008).

[31] M. Arooj, N. S. Gandhi, C. A. Kreck, D. W. Arrigan, and R. L. Mancera, J. Phys. Chem. B **120**, 3100 (2016).

[32] D. L. Cheung, Langmuir **32**, 4405 (2016).

[33] J. Zhai, A. J. Miles, L. K. Pattenden, T.-H. Lee, M. A. Augustin, B. A. Wallace, M.-I. Aguilar, and T. J. Wooster, Biomacromolecules **11**, 2136 (2010).

[34] J. Zhai, S. V. Hoffmann, L. Day, T.-H. Lee, M. A. Augustin, M.-I. Aguilar, and T. J. Wooster, Langmuir **28**, 2357 (2012).

[35] L. Day, J. Zhai, M. Xu, N. C. Jones, S. V. Hoffmann, and T. J. Wooster, Food Hydrocoll. **34**, 78 (2014).

[36] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, and B. A., in *Proteomics Protoc. Handb.* (2005), pp. 571–601.

[37] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, and L. J. Smith, Protein Sci. **10**, 677 (2001).

[38] A. C. Pike, K. Brew, and K. R. Acharya, Structure **4**, 691 (1996).

[39] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).

[40] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).

[41] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).

[42] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comput. Phys. Commun. **185**, 604 (2014).

[43] G. Bussi, Mol. Phys. **112**, 379 (2013).

[44] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. Van Gunsteren, Eur. Biophys. J. **40**, 843 (2011).

[45] H. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[46] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[47] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/155341; this version posted June 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[48] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[49] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, Proc. Natl. Acad. Sci. U. S. A. **102**, 13749 (2005).

[50] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[51] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, J. Comput. Chem. **32**, 2319 (2011).

[52] D. Frishman and P. Argos, Proteins Struct. Funct. Genet. **23**, 566 (1995).

[53] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. Van Gunsteren, and A. E. Mark, **38**, 236 (1999).

[54] R. L. C. Vink, J. Horbach, and K. Binder, J. Chem. Phys. **122**, 134905 (2005).

[55] A. Radzicka and R. Wolfenden, Biochemistry **27**, 1664 (1988).

[56] Ø. Halskau, N. Å. Frøystein, A. Muga, and A. Martínez, J. Mol. Biol. **321**, 99 (2002).