The Air-Water Interface Stabilises α-helical Conformations of the Insulin B-Chain

David L. Cheung*

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

Abstract

Adsorption of proteins onto liquid interfaces, such as the air-water interface, often leads to changes in the protein conformation. This can lead to changes in protein assembly behaviour, with aggregation and fibrillation often enhanced. To understand the relationship between protein conformation and aggregation, knowledge of protein structure at interfaces, on the single molecular level, is necessary. Using molecular dynamics simulations the effect of the air-water interface on conformation of the insulin B-chain is investigated. At the air-water interface the protein adopts an α-helical conformation, whereas in bulk solution it adopts disordered structures. The α-helical conformation is templated by the partitioning of hydrophobic side chains into the air, leading to the formation of an amphipathic helix. This structure presents a hydrophobic face which may lead to further aggregation, which helps explain the enhancement of insulin fibrillation at interfaces. This knowledge of the molecular conformation gives new insight into the contribution of protein structural change on the interfacial aggregation of proteins.

I. INTRODUCTION

Hydrophobic-hydrophilic interfaces, such as the air-water interface (AWI), provide useful environments for the investigation and exploitation of protein aggregation and assembly[1]. Due to their intrinsic amphiphilicity proteins naturally adsorb onto such interfaces, often accompanied by conformational changes[2,3]. This has been exploited in a number of contexts, such as the use of interfacial protein layers in food emulsions[4] or in the preparation of ordered protein arrays[5]. Adsorption of proteins onto interfaces also plays a role in some biological processes[6,7]. Notably this includes the growth of biofilms on the air-water interface[8] and modulating surface tension in pulmonary surfactant[9].

It has long been observed that protein aggregation is enhanced at interfaces. This has been extensively studied in context of protein fibrillation[10–13]. Other supramolecular structures, such as interfacial gels and networks, have also been observed for proteins[14,15]. The enhancement of protein aggregation at interfaces has been rationalised as arising due to two complementary effects[16,17]. The first is that the increased concentration of proteins at interfaces, due to their natural amphiphilicity, will lead to an increase in the aggregation rate compared to bulk solution. The second is that adsorption at interfaces causes proteins to adopt conformations favourable for aggregation. Knowledge of the conformation of individual protein molecules at interfaces is necessary to disentangle these two contributions.

There has been particular interest in the interfacial aggregation and fibrillation of insulin[18,19]. Due to its role in diabetes, insulin was the first biopharmaceutrical and there has been considerable interest in understanding its aggregation behaviour, as this influences its processing, storage and release[20]. In its active form insulin is monomeric but it is often found in higher order oligomers in solution[21]. Because of its small size (51-residues) it has been commonly used as a model protein for the investigation of protein aggregation and fibrillation. Formation of insulin fibrils is also the cause of injection amyloidosis[22]. Insulin consists of two chains, a 21-residue A-chain and 30-residue B-chain, held together by a pair of disulfide bonds, and there has been some investigation of the individual insulin A and B-chains. The B-chain forms the interior of insulin dimers and a significant number of close contacts are found between B-chains on different molecules in the insulin crystal structure[23], suggesting that this plays a key role in insulin aggregation.

While in native insulin the two disulfide bonds joining the A- and B-chains restricts

their conformational freedom, NMR studies of the isolated B-chain shows that this can adopt structures similar to the crystal structure, as well as adopting molten globule like states[24,25]. Both the isolated A- and B-chains have been shown to aggregate and form fibrils independently of each other, although the structures of these assemblies can differ from those of the full protein[26–28]. Molecular dynamics (MD) simulations similarly show a greater degree of conformational freedom for the isolated B-chain[29,30], including molten globule states stabilised by electrostatic and hydrophobic interactions.

Investigation of protein conformational change at interfaces requires knowledge of the microscopic protein structure. Methods, such as NMR, which can be applied to protein structure in solution are typically unsuitable due to the weak signal from molecules at the interface. Surface sensitive techniques, such as sum frequency generation[31] (SFG), refractive index matched emulsion circular dichroism[32], and synchrotron radiation circular dichroism[33] spectroscopy can be used. SFG[34] and FTIR[18] measurements of human insulin at the air-water interface also suggest that the B-chain is α-helical, similar to the crystal structure. This structure, which contains two hydrophobic patches, is considered to be more prone to aggregation than larger oligomers.

Due to its ability to access the microscopic level MD simulation has been used to investigate protein conformational change at liquid interfaces. In some cases it is possible to study this using standard molecular dynamics simulations. In particular for small peptides, such as the LK-peptides[35,36] or small amyloidogenic fragments[37], microsecond length MD simulations can examine conformational changes at liquid interfaces. MD simulation has also been used to investigate the aggregation of small peptides at interfaces[38,39], showing in some cases changes in conformation associated with aggregation. For larger proteins, while large scale changes in conformation are outside the reach of standard MD simulations, smaller structural changes or those associated with protein function at interfaces may be studied. In particular lid opening transitions in lipases[40,41] or the hinge opening of the biosurfactant Rsn-2[42] have been investigated using simulation.

To investigate more complex conformational change at interfaces enhanced sampling techniques can be used. Replica exchange molecular dynamics[43] (REMD), both in its original form and variations thereof, have been applied to a number of proteins at interfaces, including the hydrophobin EAS[44], peptides derived from myoglobin[45], and globular proteins, such as lysozyme[46,47]. These have given microscopic insight into how the conformations of

proteins at interfaces affect their behaviour, such as the role of the EAS_{19-45} loop in the oligomerization of EAS and the differing emulsification behaviour of myloglobin peptides. Biased simulation methods, such as metadynamics[48], are also commonly used to enhance conformational sampling and to calculate free energy landscapes. Metadynamics and its variations have been particularly useful for investigating intrinsically disordered proteins, such as amyloid beta[49] and IAPP[50]. While they have been extensively used for the investigation of protein structure in bulk solution[30,51] and at surfaces[52], the use of metadynamics to investigate protein conformation at liquid interfaces has been limited. In a recent study bias-exchange metadynamics was used to investigate changes in conformational preference for some small proteins, including Trp-cage and polyalanine, at the air-water interface[53], demonstrating the utility of advanced simulation methods to investigate protein conformation at interfaces. Metadynamics simulations have also been used to predict VSFG spectra for a number of small peptides at the air-water interface[54].

In this paper I apply replica exchange simulations and metadynamics to study the effect of the AWI on the conformational preferences of the insulin B-chain. As outlined above insulin is a commonly studied protein for aggregation and fibrillation, so provides an ideal model system for investigating the molecular basis of the enhancement of aggregation at interfaces. Specifically I investigate the conformation of individual protein molecules, which of course limits the ability of these simulations to investigate the effect of higher protein concentration at the interface. However, this allows for the study of the initial stages of interfacial aggregation and to investigate interfacial conformational change independent of the increased protein concentration at interfaces. It is shown that the interface causes the formation of ordered protein structure, compared to the disordered conformations found in solution. The driving forces that induce the formation of order at the air-water interface are also investigated.

II. SIMULATION DETAILS

A. Simulated system

For both air-water interface and bulk solution the simulated system contains a single insulin B-chain (Figure 1). The initial structure for the B-chain is taken from the solution

NMR structure of human insulin[55] (1HLS). This contains a central α-helical region, encompassing residues S9-C19, with two flexible tails. The experimental structure contains a H16Y mutation, which was reversed using the psfgen molecule of VMD[56]. Protonation states for the termini and ionisable residues were set appropriate for pH 7 (N-terminus, LYS, and ARG residues positively charged, C-terminus,GLU, and ASP residues negatively charged and HIS residues neutral). Using standard Gromacs utilities this was solvated in a water box of 57.5 Å^3 containing 5640 water molecules for the bulk simulations and a water slab of 60 Å×60 Å×180 Å (containing 6942 water molecules), with water occupying the central third of the simulation box, for the AWI. The AWI simulations used a larger area to allow for protein expansion along the air-water interface. For both AWI and bulk simulations the box sizes are at least 20 Å larger than the protein size in each direction. This is significantly larger than the van der Waals cut-off suggesting minimal interaction between periodic images and that the differences between the system sizes for the AWI and bulk simulations are unlikely to be due to the different box sizes used. As the protein is overall neutral no counter-ions are added to the simulated system.

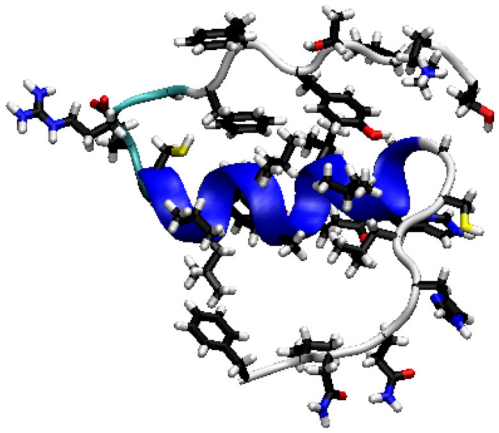

FIG. 1. Structure (top) and sequence (bottom) of insulin B-chain. Structure taken from solution NMR structure of insulin (1HLS[55]).

Both systems were initially energy minimised using the steepest descent algorithm followed by short (20 ps) NVT simulations (at 300 K), first with the positions of the heavy

atoms in the protein restrained to their initial positions by harmonic potentials (with force constant 2.4 kcal mol^{-1} Å^{-2}), then without the position restraints. A short (20 ps) NpT-simulation was then performed for the bulk solution. For both the air-water interface and bulk solution 200 ns simulations were then performed, which for the AWI simulations was sufficient for the protein to diffuse to the interface. Simulations at higher temperatures (400 K and 440 K) were then used to prepare starting conformations for the replica exchange simulations.

The system was modelled using the Charmm36m force field[57] with TIP3P water[58]. All simulations were run using the Gromacs MD package[59,60] (version 4.6.7). The PLUMED[61,62] plugin was used for the replica exchange and metadynamics simulations (see below). Temperature was controlled using the velocity-rescaling algorithm of Bussi *et al*[63] with a relaxation time of 0.1 ps. For the bulk solution the pressure was controlled using the Parrinello-Rahman barostat[64] with a relaxation time of 2 ps. A cutoff of 10 Å was used for the van der Waals and short-range electrostatic interactions. Long-range electrostatic interactions were evaluated using a Particle Mesh Ewald[65] sum with a Fourier spacing of 0.16 nm. Reciprocal space grids of 36×36×36 (bulk solution) and 40×40×160 (AWI) were used. The equations of motion were integrated using a timestep of 2 fs, with the LINCS algorithm use to constrain bond lengths[66].

B. Simulation Methods

The simulations were performed in three stages. The first stage aimed to sample a diverse ensemble of protein structures. To accomplish this replica exchange with solute tempering[67] (REST) simulations were used. This is a variant of REMD where only part of the system, in this case the protein, is simulated at different temperatures in the different replicas, with the remainder simulated at the same temperature. Changing the temperature for only a small part of the system allows for the use of fewer replicas to span a given temperature range[67]. Selectively changing the temperature is achieved by scaling the protein-protein and protein-solvent interactions by a factor β_i that depends on the temperature. Specifically the potential energy of the ith replica is given by[68]

$$E = \beta_i E_{pp} + \sqrt{\beta_i} E_{ps} + E_{ss} \tag{1}$$

where $\beta_i = T_0/T_i$, and E_{pp}, E_{ps}, and E_{ss} are the protein-protein, protein-solvent, and solvent-solvent interaction energies respectively. The minimum and maximum temperatures were 300 K and 440 K. Eight and twelve replicas were used for bulk solution and the air-water interface respectively, with the scaling factors and effective temperatures for each replica given in Table I. Exchanges between neighbouring replicas were attempted every 500 timesteps (1 ps). The REST simulations were run for 200 ns per replica. Acceptance rates for the different replicas are given in Table IV.

	Scaling factors
AWI	1.0 (300 K), 0.966 (310.6 K), 0.933 (321.6 K), 0.901 (333 K)
	0.870 (344.8 K), 0.840 (357 K), 0.811 (369.7 K), 0.784 (382.8 K)
	0.757 (396.4 K), 0.731 (410.4 K), 0.706 (424.9 K), 0.682 (440 K)
Bulk solution	1.0 (300 K), 0.947 (316.9 K), 0.896 (334.7 K), 0.849 (353.5 K)
	0.803 (373.4 K), 0.761 (394.3 K), 0.720 (416.6 K), 0.682 (440 K)

TABLE I. Scaling factors and temperatures (in parenthesis) for REST simulations at air-water interface and bulk solution

Following the REST simulations well-tempered metadynamics[69] simulations were then used to determine the free energy surface. To accelerate its convergence this was combined with REST simulations, with the metadynamics bias allowed to evolve separately within each replica[70,71]. The WT-metadynamics bias potential is given by

$$V_{bias}\left(\{CV\},t\right) = \sum_{t'<t} \dot{\omega}\tau \ \exp\left[-V_{bias}\left(\{CV\},t'\right)/k_{\mathrm{B}}\Delta T\right] \exp\left[-\sum_i \frac{\left(CV_i(t)-CV_i(t')\right)^2}{2\sigma_i}\right] \tag{2}$$

where $\{CV\}$ are the set of collective variables (CV) used to bias the protein structure, σ_i are the Gaussian widths for each CV, $\dot{\omega} = 0.956$ kcal mol^{-1} ps^{-1} is the initial hill height deposition rate, $\tau = 1$ ps is the time between hill depositions, and ΔT is the virtual temperature difference. ΔT is found from the bias factor $\gamma = (T + \Delta T)/T = 20$. The collective variables used in the metadynamics simulations were found from analysis of the REST simulations. These were chosen to be the number of α-helical hydrogen bonds and

dihedral offset function[52] given by

$$N_{\alpha-HB} = \sum_{i=1}^{N_{HB}} \frac{1-(r_i/r_0)^n}{1-(r_i/r_0)^m} \tag{3a}$$

$$DH = \frac{1}{2}\sum_{i=1}^{N-1} \left(1+\cos\left(\phi_i-\phi_{ref}\right)+\left(1+\cos(\psi_i-\psi_{ref})\right)\right. . \tag{3b}$$

These measure, respectively, how close the protein is to an ideal α-helix and β-strand. In Equation 3a r_0=4.5 Å, n=8, m=12, and the sum runs over all potential α-helical hydrogen bonds, i.e. between backbone carbonyl groups and amine groups separated by 4 residues. Note there is no requirement for these hydrogen bonds to be on consecutive residues, but most commonly a single helix is found containing all residues that participate in α-helix formation. The weight function goes to 1 as $r \to 0$ and 0 at large r. In Equation 3b the sum runs over the ϕ and ψ angles of the protein residues and the reference angles are $\phi_{ref} = -2.36$ rad and $\psi_{ref} = 2.36$ rad, corresponding to an ideal β-strand with alternating residues on opposite sides of the protein backbone. For the 30 residue insulin B-chain $N_{\alpha-HB}$ and DH lie in ranges 0 to 26 and 0 to 58 respectively. The Gaussian widths for these collective variables are set to $\sigma_{\alpha-HB} = 0.4$ and $\sigma_{DH} = 0.1$. The metadynamics simulations were run for 200 ns for bulk solution and 400 ns for the air-water interfaces. Convergence was monitored through the RMSD between the free energy surface calculated in spacings of 20 ns (Figure 9). By the end of the MTDrest simulations this was found to be below 0.01 kcal mol^{-1}.

Once a converged bias potential was found the metadynamcs simulations were then run for a further 100 ns with the bias potential held constant in order to calculate final average quantities. The effect of the bias potential was removed from the constant bias simulations using[72]

$$\langle X \rangle = \frac{\sum_i X_i \exp\left[-\beta F\left(\{CV_i\}\right)\right]}{\sum_i \exp\left[-\beta F\left(\{CV_i\}\right)\right]} \tag{4}$$

where X_i is the value at ith data set, $F = -V_{bias}$ is the free energy, $\{CV_i\}$ are the collective variables used to describe the protein conformation, and $\beta = 1/k_{\mathrm{B}}T$. Uncertainties in these quantities were estimated using the standard deviation $\sigma_X = \sqrt{\langle X^2\rangle - \langle X\rangle^2}$.

C. Analysis

Analysis of the simulations were performed using standard Gromacs utilities, in-house scripts using the MDAnalysis package[73], and VMD[56]. Secondary structure analysis was

performed using the STRIDE algorithm[74]. Protein size was characterised through the radius of gyration

$$R_g^2 = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}^{com})^2 \tag{5}$$

where $\boldsymbol{r}_i$ is the position of the ith atom and $\boldsymbol{r}^{com}$ is the protein centre of mass and the sum runs over atoms in the protein and the eigenvalues of the gyration tensor

$$G_{ij}^2 = \frac{1}{N}\sum_{i=1}^{N}(r_i - r_i^{com})(r_j - r_j^{com}). \tag{6}$$

Formation of compact structures can also be investigated through the number of contacts between Cγ atoms calculated using

$$N_{C\gamma} = \sum_i \sum_{j>i} f_{switch}(r_{ij}) \tag{7}$$

where the double sum runs over Cγ atoms on different residues and the switching function is given by

$$f_{switch}(r) = \frac{1-(r/r_0)^n}{1-(r/r_0)^m} \tag{8}$$

where $m = 12$, $n = 8$ and $r_0 = 4.5$ Å. The same switching function with $r_0 = 4.5$ Å was used for calculation of the number of salt-bridges, where the contacts were taken between Cζ and Nζ atoms in ARG and LYS and Cγ and Cδ atoms in ASP and GLU residues. Hydrogen bonds were identified using the hydrogen bond analysis module of MDAnalysis, using a donor-acceptor cut off distance of 3 Å and an angular cut-off of 120°.

The conformation of the protein at the AWI is additionally examined through the residue-interface separation and residue side chain orientation. The location of the the AWI (z_{inter}) was determined using the Gibbs dividing surface following Vink *et al*[75], with the residue-interface separation given by $\bar{z} = z_{inter} - z_{res}$, where z_{res} is the centre-of-mass of the residue side chain. Positive and negative values of $\bar{z}$ correspond to the residue being on the water and air side of the dividing surface respectively. The side chain orientation (θ) for each residue was determine using the angle between the z-axis (taken to be normal to the AWI) and the vector joining the Cα atom to the terminal heavy atom in the side chain (not calculated for glycine or proline residues).

III. RESULTS

A. Conformations from Unbiased Simulations

Shown in Figure 2(a) are the secondary structure distributions for insulin B-chain at the air-water interface and in bulk solution found from the REST simulations. At the interface the secondary structure at the air-water interface is similar to the initial structure[55]. It has an α-helical core (L11-C19), while outside this it is typically either random coil or turn. In bulk solution the secondary structure shows more variation. While structures similar to the initial conformation are found conformations lacking significant α-helix content are more common, with β-strands being present in these states. In bulk solution only 38.6% of conformations have at least eight α-helical residues, compared to 92.4 % for the AWI simulations. This demonstrates, that even in the absence of metadynamics bias, the differences in conformation at the air-water interface and in bulk solution.

The difference in conformation preferences are also apparent in the Ramachandran plots (Figure 2(b)). This has a single peak in the α-helix region at the air-water interface, while in bulk solution it has peaks in both the α-helix and β-strand regions. To examine this at a single residue level the populations of each quadrant of the Ramachandran plot[76] (lower left $-180^\circ < \phi \leq 0$, $-180^\circ < \psi \leq 0$; upper left $-180^\circ < \phi \leq 0$, $0 < \psi \leq 180^\circ$; lower right $0 < \phi \leq 180^\circ$, $-180^\circ < \psi \leq 0$; upper right $0 < \phi \leq 180^\circ$, $0 < \psi \leq 180^\circ$) are shown in Figure 2(c). Consistent with the secondary structure distribution (Figure 2(a)) the probability of finding residues in the lower left quadrant, which includes the α-helix region, is significantly higher at the air-water interface than in bulk solution. This is clearest for the central region of the protein with the proportion in the top left quadrant (containing β-strand regions) is higher outside of this. The population in the right hand side of the Ramachandran plot is higher in bulk solution. As this falls outside the regions of the Ramachandran plot that are characteristic of ordered structures, with neither protein exhibiting significant population in the α_L region (Figure 2(b)), this suggests a higher degree of disordered structures in bulk solution.

The differences between the secondary structures in the different environments suggest that CV that describe the proportion of α-helix and β-strand ($N_{\alpha-HB}$ and DH, Equation 3) would be suitable for describing the protein structure. The distributions of these two collec-

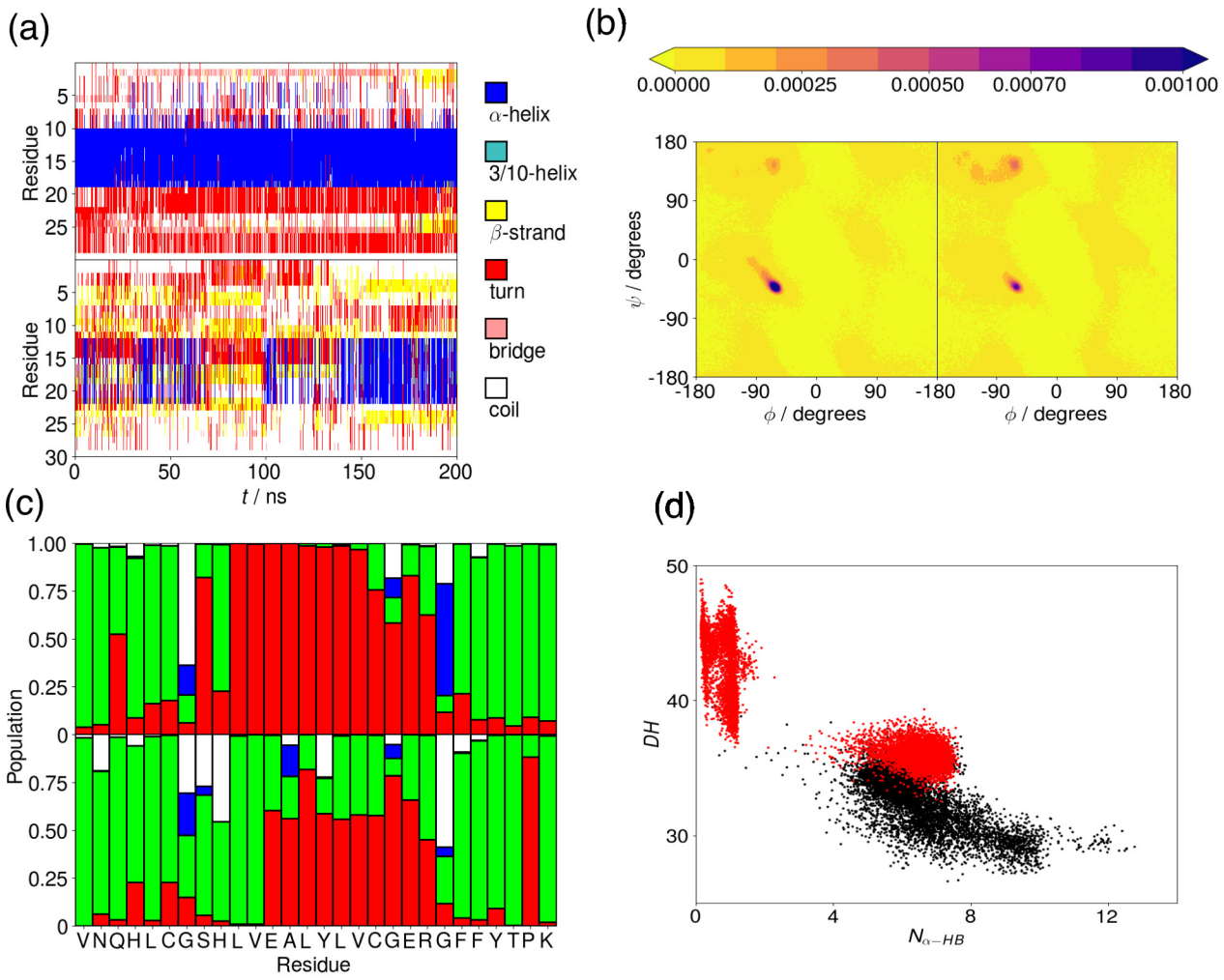

FIG. 2. Comparison between insulin B-chain secondary structure at air-water interface and in bulk solution. (a) Secondary structure distribution (AWI top, bulk solution bottom). (b) Ramachandran plots (AWI left, bulk solution right). (c) Population of each quadrant of the Ramachandran plots for each residue omitting first and last residues in chain) (AWI top, bulk solution bottom). Red, green, blue, and white denote lower left, upper left, lower right, upper right respectively. (d) Distribution of $N_{\alpha-HB}$ and DH from REST simulations at AWI (black) and in bulk solution (red).

tive variables (Figure 2(d)) for the air-water interface and in bulk solution are significantly different, justifying their use as the CVs to be used in the metadynamics simulations.

B. Free energy surfaces

The free energy surfaces for insulin B-chain, both at the air-water interface and in bulk solution (determined from the MTD-REST simulations), are shown in Figure 3(a). The differences in the conformational behaviour in these environments can be clearly seen. At the air-water interface there are a number of low energy regions in the free energy surface while in bulk solution the low energy states are all found at low values of $N_{\alpha-HB}$. This supports the observation of the shift in the conformation towards more α-helical structures at the air-water interface seen in the REST simulations.

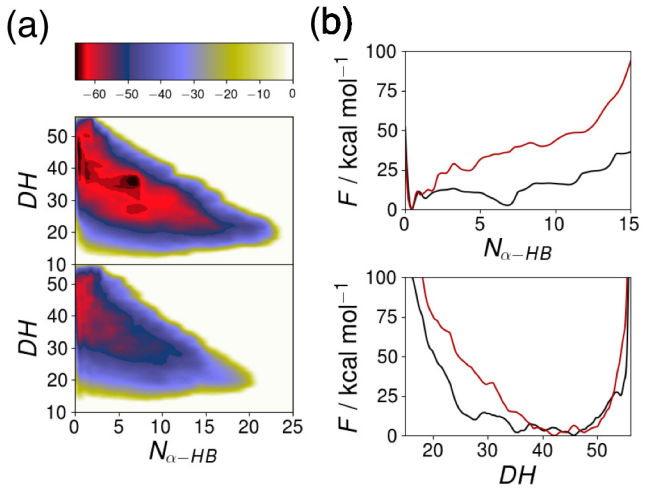

FIG. 3. (a) 2D free energy surfaces for insulin B-chain at air-water interface (top) and in bulk solution (bottom). (b) 1D free energy profiles. Black and red curves denote air-water interface and bulk solution respectively.

One-dimensional free energy profiles (Figure 3(b)) also shows these changes in the conformational preferences. These show the shift towards higher $N_{\alpha-HB}$ and lower DH at the air-water interface. In particular, while in bulk solution there is only a single minimum in $F(N_{\alpha-HB})$ at $N_{\alpha-HB} \sim 1$, at the air-water interface a second minimum is found at higher numbers of α-helical hydrogen bonds. The minimum at low $N_{\alpha-HB}$ corresponds to states similar to those found in bulk solution while the minimum at higher $N_{\alpha-HB}$ corresponds

to a more ordered, α-helical state. The structures corresponding to these states and the driving forces for their formation will be discussed in the following sections. For DH the free energy profiles both at the air-water interface and in bulk solution are relatively broad. The low energy region of $F(DH)$ extends to lower values of DH at the air-water interface, suggesting a lower preference for β-strands compared to bulk solution.

C. Differences between interfacial and solution conformations

Qualitative differences between the air-water and solution conformations can be seen in simulation snapshots (Figure 4). At the air-water interface the protein tends to form more extended conformations, typically with an ordered core and disordered regions at the termini. As in the REST simulations this is typically an α-helix (e.g. the $N_{\alpha-HB}$ $\sim$4.5 and $N_{\alpha-HB}$ $\sim$6.8), which correspond to states in the higher minimum in $F(N_{\alpha-HB})$ (Figure 3(b)). These structures are similar to the crystallographic structure of the insulin B-chain[23]. Figure 4 also shows representative structures at the air-water interface with low $N_{\alpha-HB}$, corresponding to states in the lower minimum in $F(N_{\alpha-HB})$. These are similar to those conformations in bulk solution, where the protein is typically found in more compact and less ordered structures[30].

Despite these significant differences in structure, the size of the protein is similar in both environments, with the radius of gyration and gyration tensor eigenvalues being similar (Table II). The differences in structure have more effect of quantities that describe the intraprotein interactions. The more compact structure in bulk solution is shown by the higher number of C_γ contacts. This is consistent with previous simulation studies of insulin B-chain in solution which found molten globule-like behaviour with conformations stabilized by hydrophobic interactions[30]. At the air-water interface the number of intra-protein hydrogen bonds (both overall and between the protein backbone) are higher than in bulk solution, due in part to the lower number of water molecules around the protein at the interface.

Examination of the secondary structure propensity (Figure 5(a)) of each residue shows the formation of a large α-helical region (L11-C19) in the centre of the protein at the air-water interface. This structure is consistent with SFG studies of human insulin at the air-water interface[34]. The remainder of protein is largely random coil, although it has some tendency to form turns and a small region (L6-S9) that can also form a short 3/10-helix segment. In bulk solution there are a number of short β-strand regions. The structure

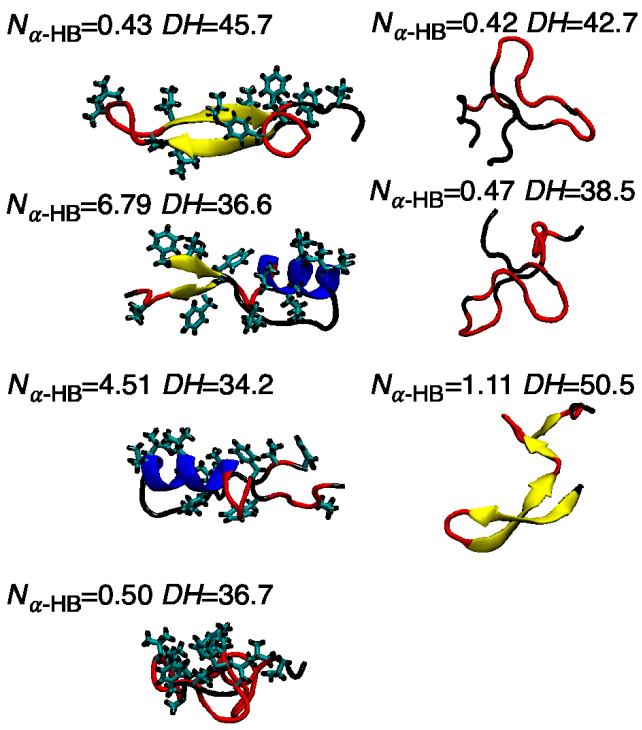

FIG. 4. Simulation snapshots showing low free energy structures for insulin B-chain at air-water interface (left) and in bulk solution (right). For air-water interface structures hydrophobic side chains are highlighted.

of insulin B-chain in solution is different from experimental structures determined using solution NMR[25,77]. These, however, were found in water/trifluoroethanol solutions and at low-pH to avoid protein aggregation, which is different to the conditions in this work (pure water, pH=7). Previous simulations under the same conditions as this work gave similar structures[30]. Simulations using the same models at low-pH reproduce the experimental structures[29], giving confidence that the present simulations provide an adequate description of the protein structure in bulk solution.

The Cα contact map (Figure 5(b)) shows that for the air-water interface only residues close (in sequence) to each other are contact. Typically residues only form contacts within 2-3 positions of each other, with slightly more distant contacts being found in the α-helical region. More distant contacts are found for the protein in bulk solution, with two sets of residues forming longer ranged contacts. These correspond to the formation of β-sheets from

	AWI	Solution
R_g / Å	10.5±0.9	10.5±0.8
G_{max} / Å	9±1	8.6±0.8
G_{mid} / Å	4.9±0.8	4.9±0.5
G_{min}/ Å	3.5±0.2	3.0±0.3
$N_{C\gamma}$	29±4	33±3
N_{sb}	0.8±0.4	0.8±0.3
N_{hbond}	19±4	13±2
$N_{hbond}^{backbone}$	14±4	9±1
$N_{water-hbond}$	87±10	98±9
$N_{water-hbond}^{backbone}$	41±7	48±5

TABLE II. Radius of gyration, gyration tensor eigenvalues, number of Cγ contacts, and number of salt bridges and number of hydrogen bonds for insulin B-chain at air-water interface and in bulk solution.

the strands (Figure 4). The second and third β-strands contain the L11-L17 amyloidogenic fragment[78].

For both the air-water interface and bulk solution these contacts are driven by hydrogen bonds. Shown in Figure 5(c) are the hydrogen bonding patterns at the air-water interface and in bulk solution, defined using pairs of residues that have hydrogen bonds with a probability of over 50 %. At the air-water interface most of the common hydrogen bonds are formed in the α-helical region (L11-C19). This analysis also shows that E13 plays an important role in determining the protein structure: As well as having a hydrogen bond between its backbone carbonyl oxygen and the amine group in L17, it hydrogen bonds with S9 stabilising the intermittent turn and the Oϵ atoms in its side chain hydrogen bond with the R22 sidechain. The turn at the N-terminus is additionally stabilised by the L6-S9 hydrogen bond. In bulk solution hydrogen bonds are formed between residues in the first two and second two β-strands consistent with the formation of β-sheets. Similar intramolecular interations are responsible for stabilising structures with low α-helix content at the air-water interface.

Differences in the protein structure are also reflected in the average number of intraprotein hydrogen bonds for each residue (Figure 5(d)). This is typically higher at the air-water

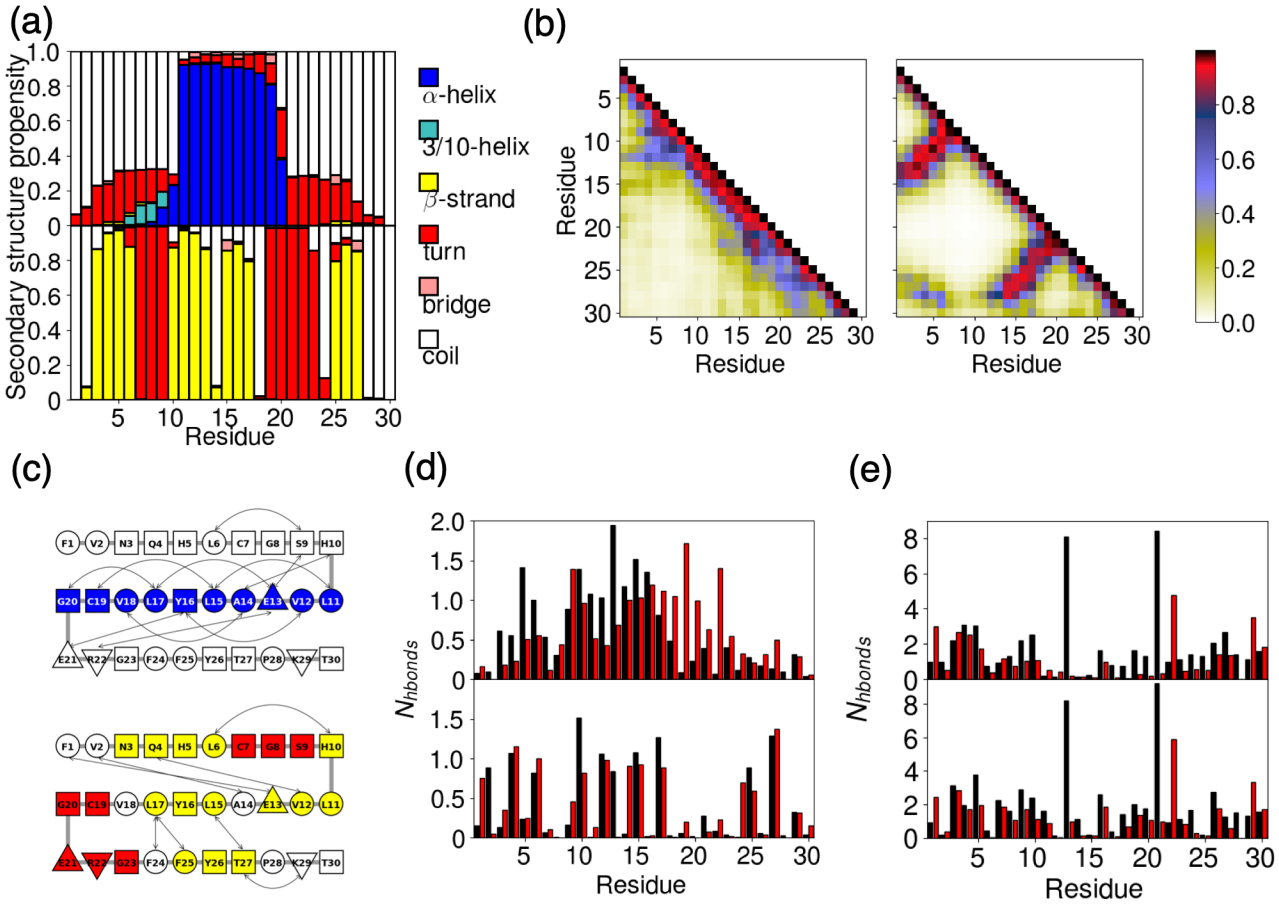

FIG. 5. Comparison between insulin B-chain structure at air-water interface and bulk solution. (a) Secondary structure propensities for air-water interface (top) and bulk solution (bottom). (b) Cα contact map for insulin B-chain. Air-water interface (left) and bulk solution (right). (c) Hydrogen bonding pattern for air-water interface (top) and bulk solution (bottom). Residues colour-coded by most common secondary structure (as in part (a)). (d) Average number of intraprotein hydrogen bonds per residue accepted (black) and donated (red) for air-water interface (top) and bulk solution (bottom). (e) Average number of residue-water hydrogen bonds accepted (black) and donated (red) for air-water interface (top) and bulk solution (bottom).

interface (Table II). At the air-water interface residues in the central α-helical region have a higher number of hydrogen-bonds than those in the less ordered tails. In bulk solution the highest number of hydrogens are found for residues in the β-strands. The number of protein-water hydrogen bonds (Figure 5(e)) is lower at the AWI, which reflects the smaller number of water molecules around the protein in this environment. At both the air-water interface

and in bulk solution the two glutamic acid residues and the arginine make a significant number of hydrogen bonds. This suggests that these residues lie in the water region so the environment around these side chains will be similar for both simulations.

D. Driving forces for conformational change and aggregation at air-water interface

To understand the reasons for the different conformational preferences between the air-water interface and solution it is useful to consider the driving force for α-helix formation at the air-water interface. Shown in Figure 6(a) are simulation snapshots (same conformations as in Figure 4). These show that the hydrophobic residues preferentially partition into the vacuum (air) layer. This is also shown by the average residue-interface separations and their sidechain orientation (Figure 6(b)). Hydrophobic residues typically reside closer to the interface than hydrophilic ones and have their side chains orientated towards the interface.

This preferential partitioning of hydrophobic side chains leads to the formation of ordered structures, in particular the central α-helix, stabilising states that correspond to the higher $N_{\alpha-HB}$ minimum in $F(N_{\alpha-HB})$ (Figure 3(b)). From the helical wheel projection for this region (Figure 6(c)) it can be seen that this helix has hydrophobic residues concentrated onto one side giving it an amphipathic character. The formation of α-helical structures at the air-water interface is consistent with SFG measurements of insulin at the bare air-water interface[34]. In comparison at lipid monolayers the SFG signal for human insulin is reduced[79], suggesting either an increased tendency for the formation of SFG inactive dimers or that the protein adopts more disordered conformations. This suggests that the hydrophobicity of the interface plays a key role in determining protein interfacial conformation and aggregation.

As shown in the simulation snapshots (Figure 6(a)) the α-helix lies largely in the plane of the air-water interface. The orientation of the helix relative to the interface can be examined through the angle between the helix long axis and the z-axis ($\cos\phi = \hat{u}_{helix}.\hat{z}$). Shown in Figure 6(d)) is the probability histogram of helix orientation from the MTDrest simulations (only conformations where helical segments of five or more residues were considered). This is largely peaked about $\cos\phi \approx 0$, with the average $\cos\phi \approx -0.014$, indicating that the helix does lie largely in the plane of the air-water interface. For $|\cos\phi| > 0.5$ the probability is essentially 0 suggesting that the largest angle between the helix and air-water interface is approximately 30°. The orienting effect of the interface is also likely to enhance the

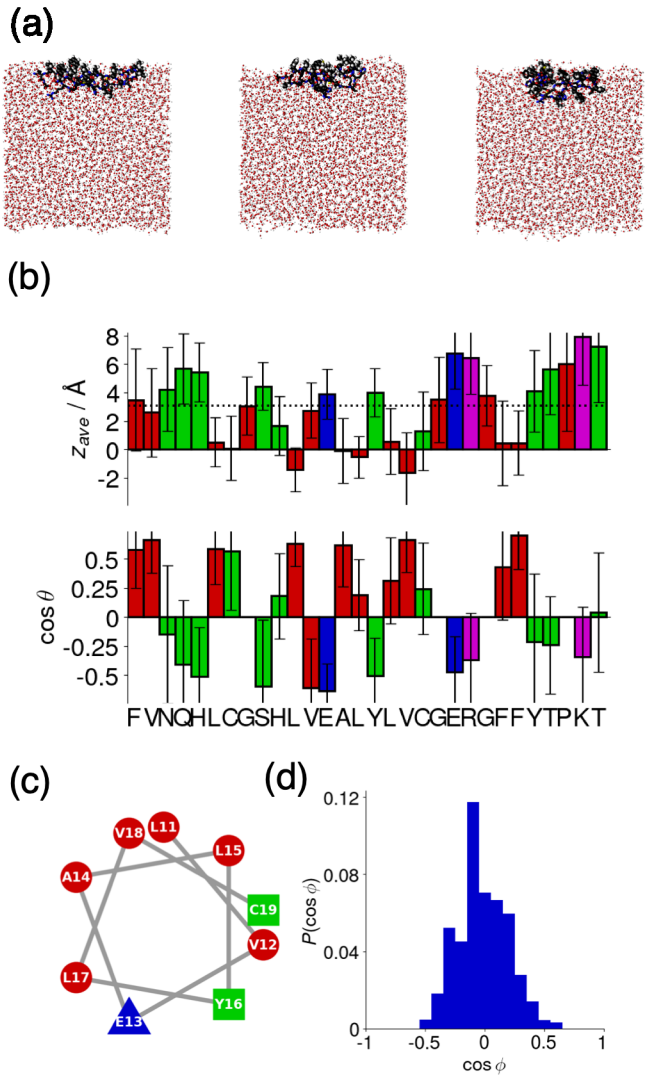

FIG. 6. Driving force for α-helix formation at air-water interface. (a) Simulation snapshots for insulin B-chain. Hydrophobic sidechains highlighted as VDW spheres. (b) Average residue centre-of-mass-interface separations (top) and average residue sidechain orientation (bottom). Red, green, blue, and magenta denote hydrophobic, polar, negatively-charged, and positively-charged residues respectively. Dotted line in top panel shows average position. (c) Helical wheel plot for residues L11-C19. (d) Probability histogram of helix orientation relative to z-axis.

aggregation and fibrillation of insulin. Formation of fibrils, through mechanisms such as the steric zipper model, requires proteins to be favourably oriented relative to each other. As proteins tend to lie in the plane of the interface this increases the probability of proteins being oriented, potentially easing the formation of fibrils.

This interfacial structure, in particular the α-helical core is similar to the crystal structure of insulin[23]. Both the AWI and the interior of the crystal expose the protein to hydrophobic environments (air for the AWI, hydrophobic interior for the crystal) so the similarity of these structures may not be surprising. To quantify the similarity between the crystal and simulation structures the C_α-C_α DRMSD (Table III) between the insulin crystal (1ZNI[23]) and simulation structures has been calculated (the insulin crystal structure contains two molecules so the DRMSD has been calculated for these separately). As can be seen the structure at the AWI is closer to the experimental crystal structure, in particular in the helix region (taken as residues 8 to 19). The AWI structure is also more similar to the structure in water-TFE mixtures determined using NMR than the solution structure, due to the increased hydrophobicity of water-TFE mixture compared to pure water. Stabilisation of the α-helix at the interface, as seen in SFG experiments, keeps insulin in a conformation with exposed hydrophobic regions, which are more prone to aggregation[21]. Helical intermediates have also been seen for other fibril forming proteins, including amyloid beta[80], amylin[81] and alpha-synuclein[82], at interfaces. This also demonstrates that the isolated B-chain retains some of the conformational behaviour of the B-chain in the full insulin molecule, explaining the similarity in their aggregation[28].

	AWI	Bulk
1ZNI B (helix)	1.3±0.8 Å	6.5±0.6 Å
1ZNI B (full)	5.7±0.9 Å	6±1 Å
1ZNI D (helix)	1.2±0.9 Å	6.4±0.6 Å
1ZNI D(full)	4.8±0.9	6±1 Å

TABLE III. Average C_α-C_α DMRSD between experimental (1ZNI) and simulation structures for insulin B-chain.

The penetration of the C-terminus (Y26-A30) into the water phase is also suggests a

role in interfacial aggregation. In insulin crystals this region forms intermolecular β-sheets. Mutations to this region are also used in a number of insulin analogues to control aggregation and release behaviour, with insulin analogues with these 5 residues deleted[83] or switching the order of these residues[84] showing reduced aggregation compared to the unmodified protein. The partitioning of this region into the water increases the ability of this region to interact with other proteins, making aggregation more likely. Similar disordered regions are involved in binding and recognition for intrinsically disordered proteins[85]. The N-terminus is also deeper into the water phase than the core of the protein, which may also lead to further interactions with other proteins.

IV. CONCLUSIONS

Using atomistic molecular dynamics simulations with enhanced sampling techniques I have investigated the effect that the air-water interface has on the conformational preference of a model fibril forming protein, the B-chain of insulin. Adsorption at the interface causes the protein to adopt a more α-helical structure compared to bulk solution, with a significant decrease in the relative free energy of helical states at the air-water interface. The adoption of α-helical structures at the air-water interface is in agreement with previous SFG studies on human insulin[34]. Compared to larger oligomers the insulin monomer is more prone to aggregation[21] due to the presence of exposed hydrophobic patches. The α-helical structure adopted at the air-water interface does exhibit exposed hydrophobic regions, in particular the α-helix , which may be expected to enhance the interfacial aggregation[86]. This aggregation is also promoted by the relative freedom of the C-terminal region, which extends into the water component. Differences in the C-terminal region may explain some of the differences in behaviour between bovine and human insulin. The B-chains of human and bovine insulin also differ only in this region (with the terminal threonine residue in human insulin replaced by an alanine). The presence of a more hydrophobic amino acid in this region may tend to reduce the tendency for this region to partition into the water, supporting the reduced aggregation at interfaces for bovine insulin[79], although differences with the A-chain may also play a role in this. Many synthetic insulin analogues, in particular those that are designed to be fast acting by reducing aggregation, also feature changes to this region. For example, insulin lispro swaps the order the lysine and proline residues in the C-terminus (residues

28 and 29) and insulin aspart replaces the C-terminus proline with aspartic acid. These changes to the C-terminus, in particular the increased number of charged residues for aspart and changes to flexibility through moving or deleting the proline residues. The differences in behaviour of insulin from different sources and for insulin variants will be investigated in future work. The structure adopted at the air-water interface is similar to the crystal structure of insulin and the similarity between these conformations suggests that the air-water interface presents the protein with a similar environment to the hydrophobic interior of the crystal.

While fluid interfaces are typically thought to have a destabilising effect on proteins, with interface-induced unfolding being common[2], this study demonstrates that such interfaces can alternatively have a stabilising or ordering effect, with a protein that is disordered in bulk takes on an ordered structure at the interface.

Enhancement of protein aggregation at interfaces is a general phenomena, with this being driven by a combination of higher protein concentration at interfaces and the adoption of aggregation prone conformations[16,17]. While fully unravelling the contribution of these two different effects on interfacial aggregation would require simulations of multiple proteins, which is computationally prohibitive the the present work, involving the study of single molecules, has demonstrated that, in the case of the insulin B-chain, the AWI leads to adoption of conformations that are more ordered and more prone to aggregation than in bulk solution. Additionally the interface causes the proteins to lie in the interfacial plane, increasing the probability of two molecules having orientations favourable for aggregation. This suggests the interface induced changes to protein conformation plays a key role in the enrichment of fibrillation, i.e. this is not simply a consequence of increased protein concentration at interfaces. This is likely a general phenomenon but extension of this to other proteins and to other interfaces will be necessary to determine specific driving forces and how this is affected by protein structure and environment. For instance simulation of insulin on lipid monolayers may be used to resolve the differences in insulin behaviour at more complex interfaces compared to the bare AWI[34,79]. This may then be used to give further insight into the mechanisms of protein aggregation, such as the formation of protein fibrils, which may be used to guide and control the formation of protein aggregates.

ACKNOWLEDGEMENTS

Computational resources for this work were provided by the SFI/HEA funded Irish Centre for High End Computing.

CONFLICTS OF INTEREST

There are no conflicts of interest to declare.

Appendix A: Acceptance rates and convergence

Shown in Table IV are the acceptance rates for the different stages of the simulations (REST, MTD-REST, MTD-REST with static bias). In all cases the acceptance rates are above 20 % for all simulations and all pairs of replicas. These are higher for the AWI simulations, due to the smaller differences between the replicas. The acceptance rates for the different stages of the simulations are also approximately the same, suggesting that the addition of the metadynamics bias does not negatively affect exchanges between the replicas.

Simulation	0↔1	1↔2	2↔3	3↔4	4↔5	5↔6	6↔7	7↔8	8↔9	9↔10	10↔11
AWI											
REST	38 %	40 %	39 %	41 %	37 %	38 %	45 %	41 %	45 %	47 %	43 %
MTD-REST	44 %	43 %	42 %	43 %	41 %	41 %	48 %	42 %	47 %	49 %	44 %
MTD-REST (static bias)	40 %	39 %	37 %	41 %	39 %	37 %	45 %	42 %	41 %	44 %	40 %
Bulk											
REST	21 %	23 %	28 %	26 %	31 %	28 %	33 %				
MTD-REST	26 %	23 %	30 %	28 %	31 %	27 %	32 %				
MTD-REST (static bias)	24 %	22 %	26 %	25 %	27 %	24 %	31 %				

TABLE IV. Exchange acceptance rates for AWI (top) and bulk (bottom) simulations.

Motion of trajectories between replicas can be monitored through the variation of the REST scaling parameter (β_i) for different replicas (Figure 7). As can be seen, particularly when the metadynamics bias is applied, the replicas explore different values of β_i. Shown in

Figure 8 are representative examples of the time variation of the secondary structure across the simulations. In most cases changes to the secondary structure are clearly seen, again showing that sampling of different protein conformations is occurring within each trajectory.

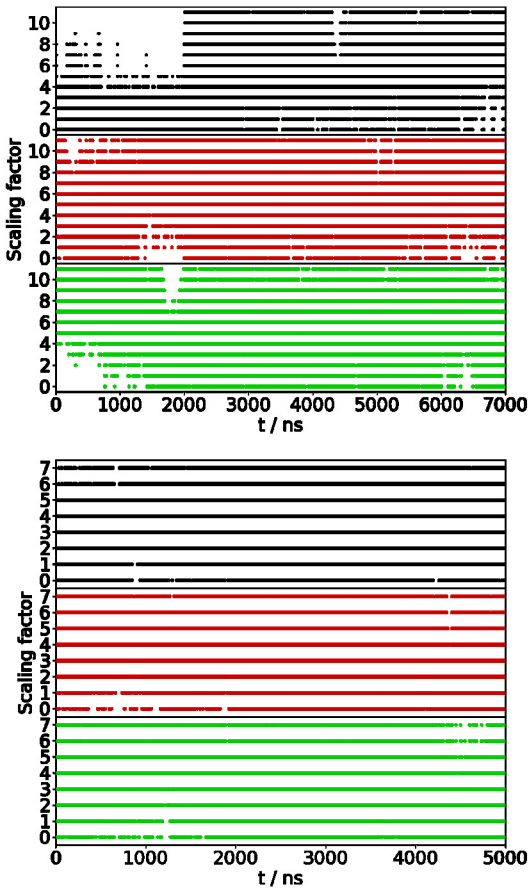

FIG. 7. (top) Plot of scaling factor (i denotes β_i) against time for AWI simulations. Black, red, and green denote simulations starting at replica 0, 6, and 11. (top) Plot of scaling factor (i denotes β_i) against time for bulk solution simulations. Black, red, and green denote simulations starting at replica 0, 4, and 7.

Convergence of the MTD-REST simulations was monitored through comparison of the 1D free energy profiles ($F(N_{\alpha-HB})$ and $F(DH)$) calculated at different times. Shown in Figure 9 are the 1D free energy profiles calculated after 360 ns, 380 ns, and 400 ns (AWI) and 160 ns, 180 ns, and 200 ns. For these different times only slight differences are found in $F(N_{\alpha-HB})$ and $F(DH)$, with those calculated after in the lsat 20 ns of the simulations

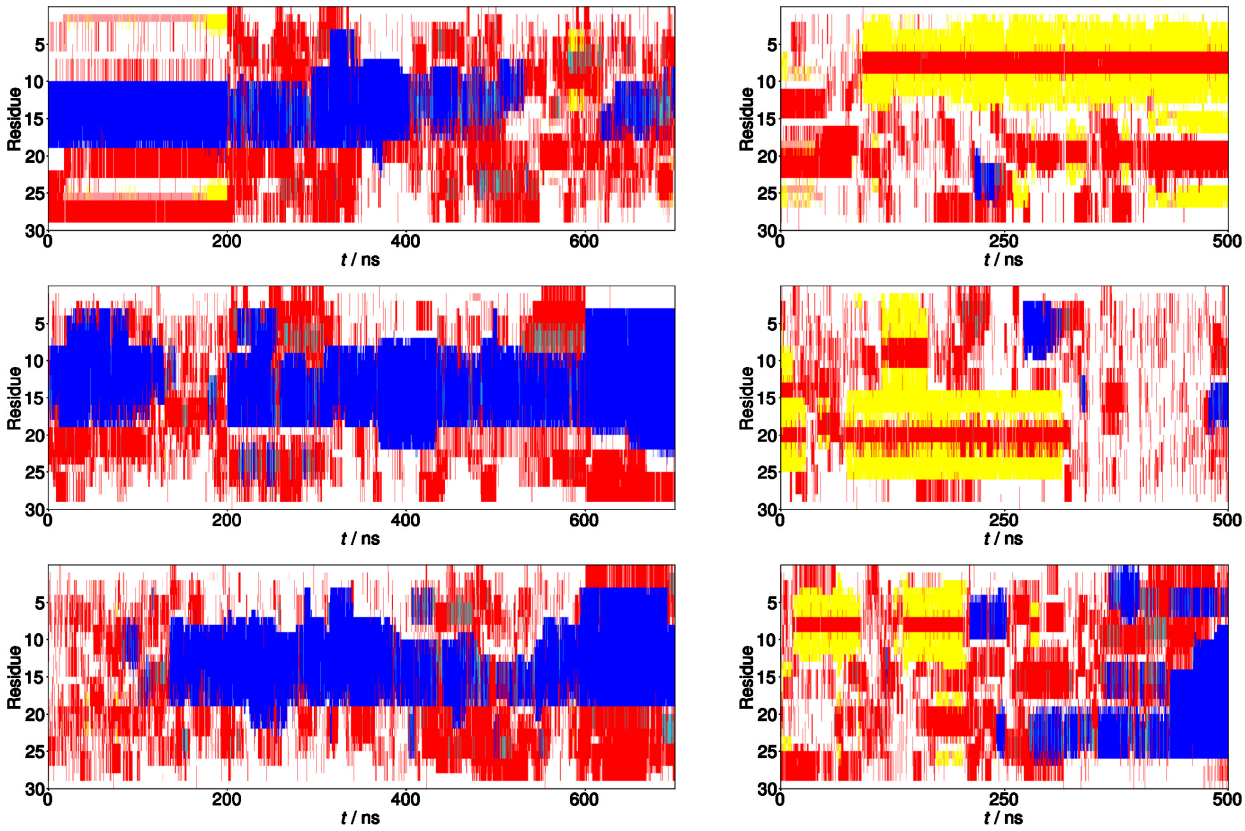

FIG. 8. Secondary structures for demultiplexed replicas. Left hand column shows AWI simulations starting from replica 0 (top), 6 (middle), and 11 (bottom). Right hand column shows bulk solution simulations starting from replica 0 (top), 4 (middle), and 7 (bottom). Colours as in Figure 2(a).

being particularly similar. This suggests that the free energy surface has largely converged at the end of the MTD-REST simulations.

The trajectories of the two main CVs ($N_{\alpha-HB}$ and DH) across the static-bias simulations are shown in Figure 10. Frequent transitions between different values of both of these are observed within the static bias simulations, indicating that the simulations are sampling across different proteins conformations. Across the static-bias simulations the distribution of values remains constant suggesting that the results would not change if the simulation lengths were increased.

* david.cheung@nuigalway.ie

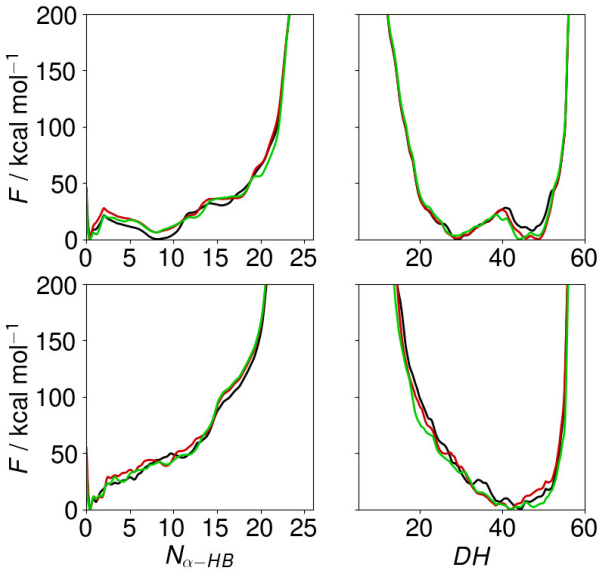

FIG. 9. (top) One-dimensional free energy profiles calculated after 360 ns, 380 ns, and 400 ns for MTDrest simulation at AWI. (bottom) One-dimensional free energy profiles calculated after 160 ns (black),180 ns (red), and 200 ns (green) of MTDrest simulations for bulk simulations.

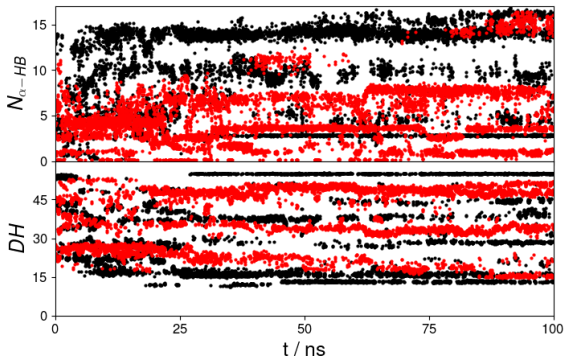

FIG. 10. Time variation of $N_{\alpha-HB}$ (top) and DH (bottom) in MTDrest simulations. Black and red symbols denote AWI and bulk simulations respectively.

[1] L. He, A. F. Dexter, and A. P. J. Middelberg, Chem. Eng. Sci. **61**, 989 (2006).

[2] Y. F. Yano, J. Phys. Conden. Matt. **24**, 503101 (2012).

[3] J. L. Zhai, L. Day, M.-I. Aguilar, and T. J. Wooster, Curr. Opin. Coll. Interf. Sci. **18**, 257

(2013).

[4] J. Peng, J. R. Simon, P. Venema, and E. Van Der Linden, Langmuir **32**, 2164 (2016).

[5] B. R. Silver, V. Fülöp, and P. R. Unwin, New J. Chem. **35**, 602 (2011).

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

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

[8] M. Sunde, C. L. Pham, and A. H. Kwan, Ann. Rev. Biochem. **86**, 585 (2017).

[9] A. G. Serrano and J. Pérez-Gil, Chem. Phys. Lipids **141**, 105 (2006).

[10] A. Morinaga, K. Hasegawa, R. Nomura, T. Ookoshi, D. Ozawa, Y. Goto, M. Yamada, and H. Naiki, Biochim. Biophys. Acta - Proteins Proteomics **1804**, 986 (2010).

[11] M. Hoernke, J. A. Falenski, C. Schwieger, B. Koksch, and G. Brezesinski, Langmuir **27**, 14218 (2011).

[12] V. K. Morris, Q. Ren, I. Macindoe, A. H. Kwan, N. Byrne, and M. Sunde, J. Biol. Chem. **286**, 15955 (2011).

[13] S. Campioni, G. Carret, S. Jordens, L. Nicoud, R. Mezzenga, and R. Riek, J. Am. Chem. Soc. **136**, 2866 (2013).

[14] E. Dickinson, Coll. Surf. B **20**, 197 (2001).

[15] D. B. Jones and A. P. J. Middelberg, AIChE J. **49**, 1533 (2003).

[16] L. Jean, C. F. Lee, C. Lee, M. Shaw, and D. J. Vaux, FASEB J. **24**, 309 (2010).

[17] L. Jean, C. F. Lee, and D. J. Vaux, Biophys. J. **102**, 1154 (2012).

[18] S. Johnson, W. Liu, G. Thakur, A. Dadlani, R. Patel, J. Orbulescu, J. D. Whyte, M. Micic, and R. M. Leblanc, J. Phys. Chem. B **116**, 10205 (2012).

[19] S. Li and R. M. Leblanc, J. Phys. Chem. B **118**, 1181 (2014).

[20] S. Frokjaer and D. E. Otzen, Nat. Rev. Drug Discovery **4**, 298 (2005).

[21] J. L. Whittingham, D. J. Scott, K. Chance, A. Wilson, J. Finch, J. Brange, and G. Guy Dodson, J. Mol. Bio. **318**, 479 (2002).

[22] F. E. Dische, C. Wernstedt, G. T. Westermark, P. Westermark, M. B. Pepys, J. A. Rennie, S. G. Gilbey, and P. J. Watkins, Diabetologia **31**, 158 (1988).

[23] G. Bentley, E. Dodson, G. Dodson, D. Hodgkin, and D. Mercola, Nature **261**, 166 (1976).

[24] Q.-x. Hua, M. Kochoyan, and M. A. Weiss, Proc. Natl. Acad. Sci. **89**, 2379 (1992).

[25] F. Y. Dupradeau, G. Le Flem, T. Richard, J. P. Monti, H. Oulyadi, and Y. Prigent, J. Peptide

Res. **60**, 56 (2002).

[26] D.-P. Hong and A. L. Fink, Biochemistry **44**, 16701 (2005).

[27] G. L. Devlin, T. P. Knowles, A. Squires, M. G. McCammon, S. L. Gras, M. R. Nilsson, C. V. Robinson, C. M. Dobson, and C. E. MacPhee, J. Mol. Biol **360**, 497 (2006).

[28] V. Babenko, M. Piejko, S. Wójcik, P. Mak, and W. Dzwolak, Langmuir **29**, 5271 (2013).

[29] N. Todorova, F. S. Legge, H. Treutlein, and I. Yarovsky, J. Phys. Chem. B **112**, 11137 (2008).

[30] N. Todorova, F. Marinelli, S. Piana, and I. Yarovsky, J. Phys. Chem. B **113**, 3556 (2009).

[31] T. Weidner and D. G. Castner, Phys. Chem. Chem. Phys. **15**, 12516 (2013).

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

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

[34] S. Mauri, T. Weidner, and H. Arnolds, Phys. Chem. Chem. Phys. **16**, 26722 (2014).

[35] C. Dalgicdir, C. Globisch, C. Peter, and M. Sayar, PLOS Comp. Bio. **11**, e1004328 (2015).

[36] C. Dalgicdir and M. Sayar, J. Phys. Chem. B **119**, 15164 (2015).

[37] A. E. Miller, P. B. Petersen, C. W. Hollars, R. J. Saykally, J. Heyda, and P. Jungwirth, J. Phys. Chem. A **115**, 5873 (2011).

[38] Y. Xue, L. He, A. P. J. Middelberg, A. E. Mark, and D. Poger, Langmuir **30**, 10080 (2014).

[39] A. Hung, Molec. Sim. **42**, 580 (2016).

[40] J. J. James, B. S. Lakshmi, A. S. N. Seshasayee, and P. Gautam, FEBS Letters **581**, 4377 (2007).

[41] M. Z. Abdul Rahman, A. B. Salleh, R. N. Z. R. Abdul Rahman, M. B. Abdul Rahman, M. Basri, and T. C. Leow, Protein Sci. **21**, 1210 (2012).

[42] 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).

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

[44] A. De Simone, C. Kitchen, A. H. Kwan, M. Sunde, C. M. Dobson, D. Frenkel, A. D. Simone, C. Kitchen, A. H. Kwan, M. Sunde, et al., Proc. Natl Acad. Sci. **109**, 6951 (2012).

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

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

[47] D. L. Cheung, J. Chem. Phys. **147**, 195101 (2017).

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

[49] D. Granata, F. Baftizadeh, J. Habchi, C. Galvagnion, A. De Simone, C. Camilloni, A. Laio, and M. Vendruscolo, Sci. Rep. **5**, 1 (2015).

[50] G. H. Zerze, C. M. Miller, D. Granata, and J. Mittal, J. Chem. Theory Comput. **11**, 2776 (2015).

[51] S. Piana and A. Laio, J. Phys. Chem. B **111**, 4553 (2007).

[52] M. Deighan and J. Pfaendtner, Langmuir **29**, 7999 (2013).

[53] M. Huai Han and C. Cheng Chiu, J. Taiwan Institute of Chem. Eng. **0**, 1 (2018).

[54] H. Lutz, V. Jaeger, T. Weidner, and B. L. D. Groot, Journal of Chemical Theory and Computation **15**, 698 (2018), ISSN 1549-9618.

[55] S. Ludvigsen, M. Roy, H. Thøgersen, and N. C. Kaarsholm, Biochemistry **33**, 7998 (1994).

[56] W. Humphrey, A. Dalke, and K. Schulten, J. Molec. Graphics **14**, 33 (1996).

[57] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller, and A. D. Mackerell, Nature Methods **14**, 71 (2016).

[58] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).

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

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

[61] G. Bussi, Molec. Phys. **112**, 379 (2013).

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

[63] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101/1 (2007).

[64] M. Parrinello and A. Rahman, J. App. Phys. **52**, 7182 (1981).

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

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

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

[68] L. Wang, R. A. Friesner, and B. J. Berne, J. Phys. Chem. B **115**, 9431 (2011).

[69] A. Barducci, G. Bussi, and M. Parrinello, Physical Review Letters **100**, 020603 (2008).

[70] J. Schneider and L. Colombi Ciacchi, J. Am. Chem. Soc. **134**, 2407 (2012).

[71] L. B. Wright, J. P. Palafox-Hernandez, P. M. Rodger, S. Corni, and T. R. Walsh, Chem. Sci. **6**, 5204 (2015).

[72] G. Torrie and J. Valleau, J. Comput. Phys. **23**, 187 (1977).

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

[74] M. Heinig and D. Frishman, Nuc. Acids Res. **32**, 500 (2004).

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

[76] G. O. Rutter, A. H. Brown, D. Quigley, T. R. Walsh, and M. P. Allen, Phys. Chem. Chem. Phys. **17**, 31741 (2015).

[77] B. Hawkins, K. Cross, and D. Craik, International Journal of Peptide and Protein Research **46**, 424 (1995).

[78] M. I. Ivanova, S. A. Sievers, M. R. Sawaya, J. S. Wall, and D. Eisenberg, Proc. Natl Acad. Sci. **106**, 18990 (2009).

[79] S. Mauri, R. Pandey, I. Rzeźnicka, H. Lu, M. Bonn, and T. Weidner, Frontiers in Physics **3**, 1 (2015).

[80] D. Jiang, K. L. Dinh, T. C. Ruthenburg, Y. Zhang, L. Su, D. P. Land, and F. Zhou, J. Phys. Chem. B **113**, 3160 (2009).

[81] D. Lopes, A. Meister, A. Gohlke, A. Hauser, A. Blume, and R. Winter, Biophys. J. **93**, 3132 (2007).

[82] C. Wang, N. Shah, G. Thakur, F. Zhou, and R. M. Leblanc, Chem. Comm. **46**, 6702 (2010).

[83] Q. Hua and M. A. Wesss, Biochemistry **30**, 5505 (1991).

[84] V. Babenko and W. Dzwolak, FEBS Letters **587**, 625 (2013).

[85] Y. Levy, J. N. Onuchic, and P. G. Wolynes, J. Am. Chem. Soc. **129**, 738 (2007).

[86] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, and A. L. Fink, Biochemistry **40**, 6036 (2001).