J Comput Aided Mol Des (2015) 29:47–58
DOI 10.1007/s10822-014-9803-6

Computational study of elements of stability of a four-helix bundle protein biosurfactant

Andrea Schaller · Natalie K. Connors · Mirjana Dimitrijev Dwyer · Stefan A. Oelmeier · Jürgen Hubbuch · Anton P. J. Middelberg

Received: 3 June 2014 / Accepted: 10 October 2014 / Published online: 17 October 2014

Abstract Biosurfactants are surface-active molecules produced principally by microorganisms. They are a sustainable alternative to chemically-synthesized surfactants, having the advantages of being non-toxic, highly functional, eco-friendly and biodegradable. However they are currently only used in a few industrial products due to costs associated with production and purification, which exceed those for commodity chemical surfactants. DAMP4, a member of a four-helix bundle biosurfactant protein family, can be produced in soluble form and at high yield in *Escherichia coli*, and can be recovered using a facile thermal phase-separation approach. As such, it encompasses an interesting synergy of biomolecular and chemical engineering with prospects for low-cost production even for industrial sectors. DAMP4 is highly functional, and due to its extraordinary thermal stability it can be purified in a simple two-step process, in which the combination of high temperature and salt leads to denaturation of all contaminants, whereas DAMP4 stays stable in solution and can be recovered by filtration. This study aimed to characterize and understand the fundamental drivers of DAMP4 stability to guide further process and surfactant design studies. The complementary use of experiments and molecular dynamics simulation revealed a broad pH and temperature tolerance for DAMP4, with a melting point of 122.4 °C, suggesting the hydrophobic core as the major contributor to thermal stability. Simulation of systematically created in silico variants of DAMP4 showed an influence of number and location of hydrophilic mutations in the hydrophobic core on stability, demonstrating a tolerance of up to three mutations before a strong loss in stability occurred. The results suggest a consideration of a balance of stability, functionality and kinetics for new designs according to their application, aiming for maximal functionality but at adequate stability to allow for cost-efficient production using thermal phase separation approaches.

Keywords MD simulation · Stability · Biosurfactants · Four-helix bundle

Introduction

Biosurfactants, surfactant molecules produced by a wide range of different microorganisms [1], have advanced as a current trend in industry, as they are a sustainable alternative to petroleum-based chemically synthesized surfactants [2–5]. As surfactants are applied worldwide in innumerable different products and industrial sectors, ranging from food, personal care and hygiene, cosmetics, cleaning and washing, and paints to the pharmaceutical industry, the advantages of biosurfactants in those products of being biodegradable, non-toxic, highly surface active and functional, renewable and ecofriendly are close at hand [3, 4]. Despite these advantages and the current movement

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9803-6) contains supplementary material, which is available to authorized users.

A. Schaller · N. K. Connors (✉) · M. D. Dwyer · A. P. J. Middelberg
Centre for Biomolecular Engineering, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St Lucia, QLD 4072, Australia
e-mail: n.connors@uq.edu.au

S. A. Oelmeier · J. Hubbuch
Section IV: Biomolecular Separation Engineering, Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

toward industrial sustainability, biosurfactants are so far only used in a few commercial products [5–7]. This is due to the cost of production and recovery or purification, which frequently includes the use of expensive chromatographic steps and makes them at present not competitive with cheaper chemical surfactants. Further, difficulties with expression yield, and the design of molecules for a specific function, need to be overcome [6].

To address these problems, Middelberg et al. recently designed a four-helix bundle biosurfactant protein, DAMP4 (MD-[PS-MKQLADS-LHQLARQ-VSRLEHA-D]$_4$), a highly functional surface-active molecule with unique switchable foaming behaviour [8, 9]. Recombinant expression of DAMP4 in the microbial cell factory *Escherichia coli* (*E. coli*) results in a stable product at high yield [8, 10]. DAMP4 is expressed intracellularly but can be purified cheaply in a simple two-step process with thermal treatment followed by solid–liquid separation. Costly chromatography or mechanical cell disruption is not necessary, which allows this process to be used even in low-cost industrial sectors [10]. In detail, following expression, Na_2SO_4 is added to the culture broth to a concentration of 0.5 M and the entire medium heated to 90 °C for 30 min. The kosmotropic salt in combination with the high temperature causes cell lysis and denaturation of host cell proteins, while DAMP4 is salted in and stays stable in solution. Two filtration steps (1.2 and 0.2 μm) remove precipitated impurities and allow the recovery of industrial grade target protein [10]. This process takes advantage of the above-average thermostability of DAMP4 in comparison to the majority of the bacterial protein contaminants. Understanding the structural fundamentals of stability of this four-helix bundle protein can help facilitate the design of new molecules to ensure the applicability of the low-cost purification process while maintaining functionality.

Four-helix bundles have been studied widely, as a naturally-occurring protein fold [11–14] as well as a designed structural motif [8, 15–18]. The stability of four-helix bundles can vary and is influenced by different factors. Several studies demonstrate they can have extremely high thermal stability. Schafmeister et al. [18] reported the design of a 108 amino acid long four-helix bundle (DHP_1) with native-like structure and a melting point similar to that of proteins from thermophilic microorganisms of 122 °C. They suggested the design with a reduced set of alternating hydrophilic and hydrophobic amino acids influenced the properties of the bundle. Wei et al. (2003) showed with their 2nd generation combinatorial library of four-helix bundles, in which they increased the molecule length from 74 to 102 amino acids and followed their binary pattern of hydrophobic and hydrophilic residues, that it is possible to design stable and functional proteins that have a native-like structure without specifically designing the side chains.

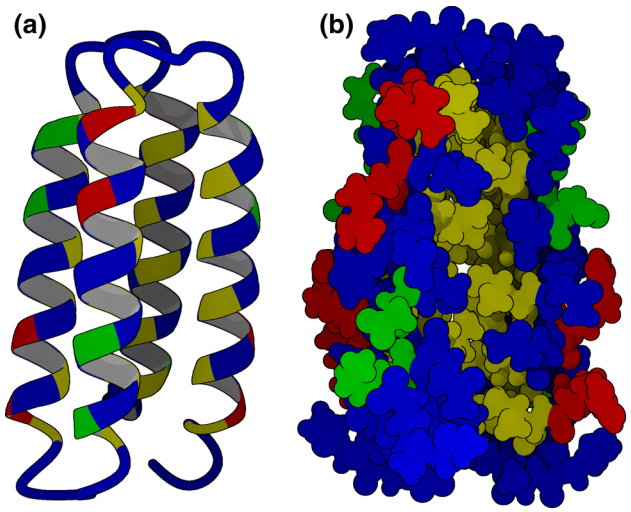

Fig. 1 Fold structure of the biosurfactant protein DAMP4. The hydrophobic core is *highlighted* in *yellow*, residues forming salt bridges in *green* and arginine residues as strong contributors to the hydration shell in *red*. **a** Ribbon, **b** ball representations

Compared to the original library, the proteins revealed considerably more stability and ordered side chains, and they were able to design a second-generation protein with a melting point of ~100 °C as shown by DSC [19]. Munson et al. [20] studied the packing of the hydrophobic core, and reported the melting point of the widely used model four-helix bundle protein Rop wild type to be 64 °C. Among suggested mechanisms influencing the stability of the four-helix bundle fold are the hydrophobic packing [21], loop length [13], electrostatic interactions including hydrogen bonding, ion pair (salt bridges) and dipole interactions [22], sequence length and polarity [21].

DAMP4 consists of four identical helices, which are arranged to build an antiparallel four-helix bundle with high helical content (see Fig. 1). DAMP4 is 98 amino acids long, has a molecular weight of 11.1 kDa and a theoretical pI of 6.7. Its helices are derived from the carboxyl-terminal oligomerization domain of the lac repressor protein (Lac21), which self-assembles to form tetramers [23], connected with an Asp-Pro-Ser linker, Asp to promote electrostatic repulsion between adjacent helices, Proline to ensure helix termination due to its low helix propensity, and Ser for flexibility [8]. Due to energetic reasons the four helices will arrange to a bundle to prevent water penetration of the hydrophobic residues based also on known behaviour of heptad peptides containing hydrophobic residues in the a and d positions of the heptad. Considering behaviour, structure and sequence, we suggested three mechanisms, which might be important for DAMP4's stability: (1) the pronounced hydrophobic core; (2) ion pair interactions, and; (3) a tightly-held hydration shell. Each of DAMP4's helices consists of three heptads, in which every

1st and 4th amino acid is hydrophobic. To minimize the solvent exposure in agreement with the hydrophobic effect, these are oriented towards the inside of the bundle and prevent any solvent penetration into the core. Charged residues on the outer surface of the molecule interact electrostatically with each other and thus drive folding and stability. In DAMP4, positively-charged lysine (at positions 6, 30, 54 and 78) is interacting with negatively charged aspartic acid (at positions 10, 34, 58 and 82), giving four salt bridges, one on each helix. Previous results [8] suggested the importance of the water structure around DAMP4 for its functionality and therefore stability, which might be promoted by positively charged arginine on the solvent exposed surface creating a stabilizing hydrogen-bonded water network [24, 25]. Arginine is with five possible hydrogen-bond donor sites unique [26, 27] and strongly interacts with water molecules, whose mobility is lowered around the guanidinium group [24].

This study aims to understand the stability and contribution of each of these mechanisms by a combination of experiment and molecular dynamics (MD) simulation. The experiments show a high thermal stability at all pHs with a melting point of 122 °C, but average stability against chemical denaturation. In agreement, MD simulations reveal the hydrophobic core as the most important structural feature for DAMP4 stability, and further show the influence of number and location of hydrophilic mutations in the core on the overall stability of the molecule.

Materials and methods

DAMP4 was expressed [8] and purified [10] in *E. coli* as described before.

Differential scanning calorimetry (DSC)

DSC measurements were used to assess the thermal stability of DAMP4. Measurements were performed with a MicroCal VP-DSC Microcalorimeter (GE Healthcare Life Sciences, Fairfield, CT, USA) with the VPViewer 2000 DSC software. To determine the melting point (T_m) the sample (1 mg ml^{-1} DAMP4 in Milli-Q water) was treated with a ThermoVac degassing station at 25 °C for 5 min to remove air bubbles. In the DSC measurement a scan rate of 90 °C h^{-1} was applied for a temperature ramp from 30 to 130 °C, and samples analysed with the MicroCal-enabled Origin data analysis package.

Circular dichroism (CD)

CD spectra were determined with a J-810 Spectropolarimeter (Jasco, Easton, MD, USA) to analyse secondary structure of DAMP4 (0.25 mg ml^{-1} in 1 mM HEPES buffer, 200 μM EDTA) at 20 and 90 °C at a pH range from 3.2 to 10.3. To determine stability against chemical denaturants, CD signals at a wavelength of 222 nm as a measure for helix content were recorded. Samples with 1 mg ml^{-1} DAMP4 were prepared by ratio-based mixing of Milli-Q water and 8 M GdmCl-solution to reach final concentrations of 0, 0.5, 1, 2, 3, 3.2, 3.5, 3.8, 4, and 5 M GdmCl. The signal at 222 nm was measured for 300 s at 20 °C and the average for this time interval calculated for each of the GdmCl-concentrations.

Homology modelling (HM)

All homology modelling experiments and simulations were conducted in the MD software package YASARA structure version 12.6.28 (YASARA Biosciences GmbH, Vienna, Austria) [28–30]. As no homologous sequences using PSI-BLAST [31] for DAMP4 were found in the protein data bank (PDB) a single helix based on the aligned sequence of template 3EDC [32] (sequence alignment in figure S1 in supplementary material) was modelled first. Three copies were created and the four helices were connected manually. Therefore peptide bonds were added manually to the arranged helices, energy minimized and the side chain rotamers were refined in a following 500 ps MD simulation with the YASARA2 force field [33]. The model quality was validated with YASARA's internal structure validation tool [34] and the external ProSA (protein structure analysis) program [35, 36]. YASARA's structure validation tool is based on a library of gold standard reference proteins normalized in knowledge-based potentials. It compares the model to the average of these references and calculates the deviation from that standard expressed in a Z-score. This Z-score is 2.732, therefore the DAMP4 model is considered optimal. The ProSA overall model quality Z-score was −6.69, which is in the range typical for native molecules of the same size. The knowledge-based energies are below zero showing no problematic parts. A Ramachandran plot to assess the geometrical validation around the C_α-atoms can be found in the supplementary material [37]. *In silico* variants were created with homology modelling with standard parameters by providing the DAMP4 structure as a template. This was followed by a refinement simulation to find the structure with the lowest energy.

Molecular dynamics (MD) simulation

Simulations were conducted in a cubic simulation box with an extension of 10 and 20 Å around all atoms for the starting structure generation and production simulation, respectively. Periodic boundary conditions, Particle mesh Ewald for long-range electrostatics [38], the Amber03 force field [39] and a cutoff for van der Waals interactions of 7.86 Å were applied.

To increase computational speed the equations of motion were integrated with a multiple timestep algorithm (impulse method [40] or VERLET-I [41]), where intramolecular forces were calculated every 1.25 fs, whereas non-bonded Van der Waals and electrostatic interactions were updated only every 2nd step and then added with a scaling factor of two. A comparison to other multiple timestep algorithms [40] revealed the most stable trajectories for this approach under the premise that non-bonded interactions are calculated frequently (e.g. every 2nd step as in this case). Validation in YASARA was performed on the dihydrofolate reductase molecule by calculating the resulting energy shift. This energy shift was reported 0.02 per degree of freedom (K_BT/ns) at 298 K, which is comparable to drifts reported for other MD programs [42]. Further validation was conducted by simulating 25 protein crystal structures (like those described by Krieger et al. [28]) and no influence on the root mean square deviation (rmsd) was observed compared to the simulations without multiple timesteps. The temperature was kept constant by rescaling atom velocities using a Berendsen thermostat and the pressure was set to 1 bar and controlled by the time averaged solvent density and adjusted by resizing the cell isotropically to the target solvent density set according to the experimental value [43]. The simulations were run in triplicates. To generate the starting conformations, the homology models were simulated at 298 K and pH 7.4 in a simulation box filled with TIP3P water molecules in a neutralization experiment [44]. Trajectories as starting structures were saved at 3, 4 and 5 ns. For the production simulation these were energy minimized, and simulated at 423 K for 100 ns, to allow for a faster unfolding at lower computational expense [45]. The first 100 ps of the simulation were considered as equilibration and discarded from data analysis and trajectories saved every 5 ps for study purposes.

Simulation data was analysed using Matlab Version R2012a (Mathworks, Natick, MA, USA). To analyse and reduce the data set a smoothing moving average function was applied. For DAMP4 and each variant all data points of all three simulations were averaged in 50 ps simulation time intervals, in which two snapshots overlapped with the previous interval, respectively.

Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

Results and discussion

Experimental characterization of DAMP4 stability

Thermal stability of DAMP4

Thermal stability was characterized using differential scanning calorimetry (DSC) to obtain the melting point (T_m) and to investigate tertiary structure changes with increasing temperature. The melting temperature was determined in Milli-Q water, eliminating the influence of salt ions on stability. The results are depicted in Fig. 2. The endothermic peak in heat capacity occurs at approximately 122.4 °C and corresponds to the protein's transition point.

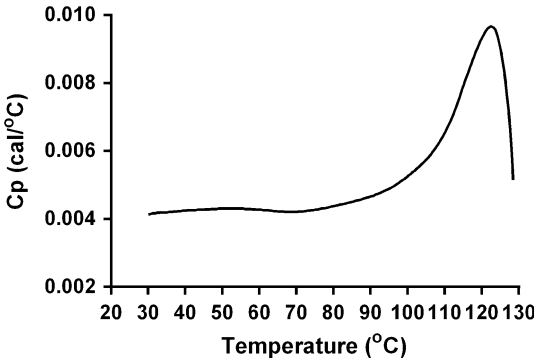

Fig. 2 Differential scanning calorimetry showing a melting point above 120 °C

The melting point lies above the average of proteins and is in the region of proteins in thermophilic organisms [46]. It is a single sharp peak indicating a cooperative process, as unfolding and denaturation occur simultaneously. A second scan of the same sample showed no distinct peak similar to the first one in the course of temperature progression, which indicates that thermal unfolding is irreversible likely due to the formation of insoluble aggregates after denaturation.

DAMP4 has a CD spectrum characteristic for helical proteins, in agreement with its designed four-helix bundle structure in bulk (Fig. 3) [47]. This helicity is preserved over the entire pH range at 20 °C, and even at a high temperature of 90 °C, with exception of pH 3.2, where some loss of helical content occurs. It has been previously reported a low pH causes a lowering of the thermal denaturation temperature [48] and partial unfolding occurs likely due to a combination of acid denaturation due to the high net positive charge of the molecule and the high temperature. This finding of weaker helicity at lower pH may indicate a significance of charges for the thermal stability of DAMP4.

Stability against chemical denaturants

To test DAMP4's stability against chemical denaturation a number of CD spectra with varying guanidinium chloride (GdmCl)-concentration were measured.

The sigmoidal transition curve shows unfolding of DAMP4 in a simple two-state folded–unfolded process (Fig. 4). Transition of DAMP4 occurs between 3 and 4 M GdmCl, whereby 50 % of the native structure is lost at

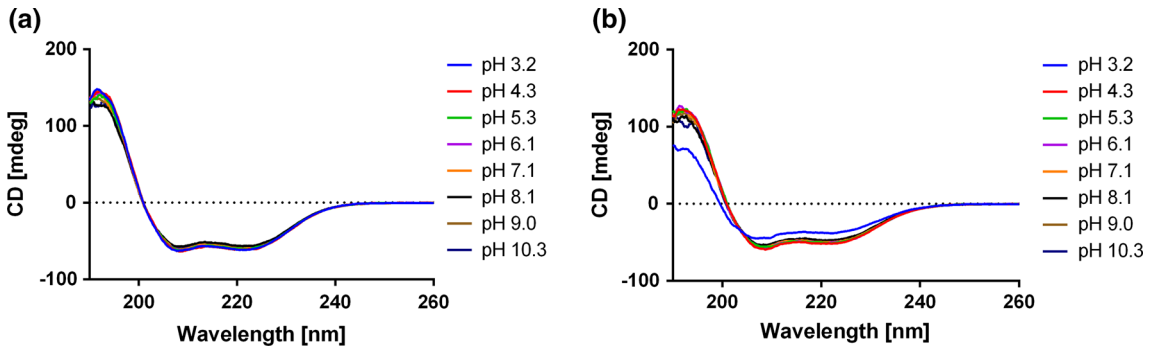

Fig. 3 Circular dichroism spectra of DAMP4 at **a** 20 °C and **b** 90 °C. The spectra show DAMP4 is stable over the entire pH range even at elevated temperatures

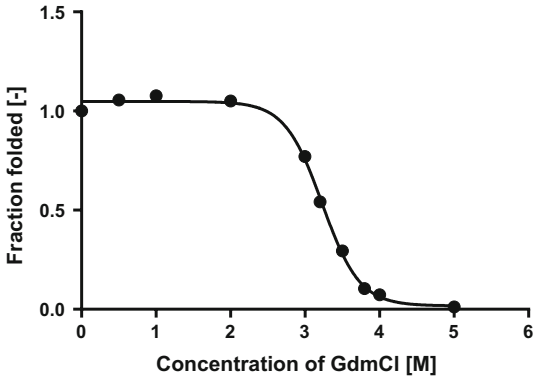

Fig. 4 Stability of DAMP4 at different concentrations of GdmCl at 20 °C. Stability against chemical denaturants is typical of that for proteins

3.23 M elucidating a moderate stability against chemical denaturants.

Investigation of DAMP4 stability with molecular dynamics simulation

Sequence design of stability variants for MD simulation

To test which of the previously-suggested stabilization mechanisms applies for the high thermal stability observed experimentally, a number of different stability variants of DAMP4 were created. To account for the influences of salt bridges, the hydrophobic core and water structure, variants in which each of these mechanisms was eliminated completely while maintaining the other effects to the greatest possible extent were created in silico. The variants are all labelled with mechanism [hydrophobic core (hc), salt bridge (sb), water shell (ws)], number of mutations in comparison to DAMP4, and the number of the variant with a capital letter when several variants with the same number of mutations exist. The sequences of the variants can be found in Table 1.

To disrupt the hydrophobic core, two hydrophobic leucines (GRAVY index [49] 3.8) on each helix were replaced with hydrophilic asparagine (GRAVY index -3.5), which has a similar side chain geometry and reduces the risk for destabilization due to steric or packing reasons (variant hc8). To break salt bridges three different variants were designed: two, where the replacement with the same charge, lysine with aspartic acid (variant sb4A) in the first heptad of each helix and vice versa (variant sb4B) causes a destabilization due to charge repulsion, and one with the effect neutralized by insertion of uncharged glutamine for aspartic acid (variant sb4C). Glutamine was selected as the side chains of glutamine and aspartic acid have the same capability to build H-bonds to water and the amino acids have the same GRAVY indices. To cause a distortion in the hydration network around the molecule, two arginine residues on each helix, which have five hydrogen-bond donor groups, were replaced with lysines, which maintains the charge structure, but the interactions with water molecules are less (variant ws8A). Biswas et al. [25] described that Lys → Arg substitutions can lead to increased thermostability, likely due to an improvement in electrostatic interactions and hydrogen-bonding network. A second water shell variant contains alanine residues in the same location, as alanine is known to have a high helix propensity [50] but does not interact with water (variant ws8B). To maintain the net charge this variant was simulated at pH 5.

Simulation results for stability variants

A simulation at a high temperature, well above the experimentally measured melting point of DAMP4, was run to investigate the differences in stability of the variants compared to DAMP4. To elucidate the progress of destabilization and unfolding, the content of α-helices and the rmsd of the C_α-atoms were analysed (Fig. 5). At the beginning of the simulation the α-helical content is around 68 % for DAMP4 and its variants, with exception of sb4B,

Table 1 Sequence design for DAMP4 in silico stability variants for MD simulation

Protein	Sequence	Sequence features used
DAMP4	MD-[PS-MKQLADS-LHQLARQ-VSRLEHA-D]$_4$	-
Hydrophobic variant		
hc8	MD-[PS-MKQNADS-LHQNARQ-VSRLEHA-D]$_4$	Hydrophobicity
Salt bridge variants		
sb4A	MD-[PS-MKQLAKS-LHQLARQ-VSRLEHA-D]$_4$	Repulsion between positive charges
sb4B	MD-[PS-MDQLADS-LHQLARQ-VSRLEHA-D]$_4$	Repulsion between negative charges
sb4C	MD-[PS-MKQLAQS-LHQLARQ-VSRLEHA-D]$_4$	Neutral amino acid inserted
Water shell variants		
ws8A	MD-[PS-MKQLADS-LHQLAKQ-VSKLEHA-D]$_4$	Arg has five proton donor groups, Lys only three
ws8B	MD-[PS-MKQLADS-LHQLAAQ-VSALEHA-D]$_4$	Ala has a high helix propensity but does not interact with water, simulated at pH 5 to maintain net charge

In these variants one of the suggested stability mechanisms (hydrophobic core, salt bridges, water shell) is eliminated or weakened. Red marks site of mutations in comparison to DAMP4

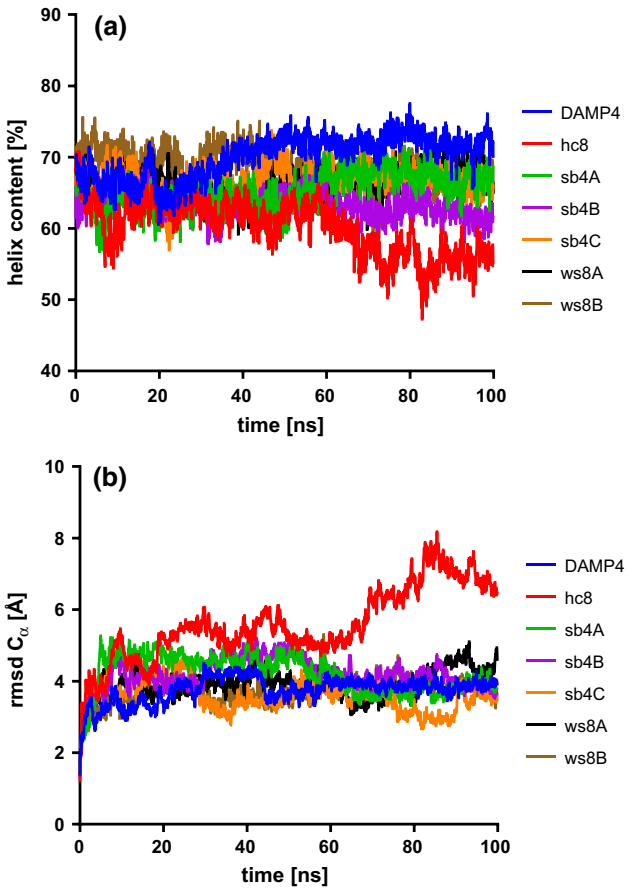

Fig. 5 Simulation results of DAMP4 stability variants. **a** Helical content and **b** rmsd C_α. These results are the moving average values in 50 ps intervals of all three starting structures

with 65 % lower compared to the others likely due to some loss of secondary structure by repulsion of similar charged residues, and ws8B, with 71 % higher consistent with the high helix propensity of alanine. After a similar course of α-helical content for all structures up to around 50 ns, the differences emerge towards the end of the simulation. DAMP4 has the highest helical content, around 72 %. Destabilization occurs during the 100 ns simulation time mainly around the turn regions, but the structure stays mainly intact with minor loss of tertiary structure. The same effects can be seen for the salt bridge and water shell variants, for which only minor loss in helical content occurs. Thereby variant sb4B has the second lowest helical content, with average 5 % difference to the other two salt bridge variants. This observation can be explained by the position of the mutation at the beginning of each helix and the lower helix propensity of aspartic acid (0.69) compared to lysine (0.26) [50]. As seen from the simulation conformations, loss of secondary structure starts from the turn regions and open beginning or end of the strands, which makes this variant particularly susceptible for lower helical content.

However, hc8 clearly reveals the greatest influence of the temperature on the structure, with a drop in secondary structure starting after approximately 60 ns to 50–55 %. The same trends can be seen considering the rmsd C_α. The salt bridge and water shell variants behave within the general fluctuations very similar to DAMP4, especially from 60 to 100 ns, whereas hc8 noticeably sets off with rmsd values intermittent double that of DAMP4, in

Table 2 Sequence design for hydrophobic variants with 1–8 Leu to hydrophilic Asp mutations in the hydrophobic core

Protein	Sequence	Mutation
DAMP4	MD- [PS-MKQLADS-LHQLARQ-VSRLEHA-D]$_4$	-
Hydrophobic variants		
hc1	MD- (PS-MKQNADS-LHQLARQ-VSRLEHA-D [PS-MKQLADS-LHQLARQ-VSRLEHA-D]$_3$)	Single mutation on the first helix
hc2A	MD- (PS-MKQNADS-LHQNARQ-VSRLEHA-D [PS-MKQLADS-LHQLARQ-VSRLEHA-D]$_3$)	Two mutations on the first helix
hc2B	MD- (PS-MKQNADS-LHQLARQ-VSRLEHA-D PS-MKQLADS-LHQNARQ-VSRLEHA-D [PS-MKQLADS-LHQLARQ-VSRLEHA-D]$_2$)	Two mutations, one on helix 1, one on helix 2
hc3	MD-([PS-MKQNADS-LHQLARQ-VSRLEHA-D]$_3$ PS-MKQLADS-LHQLARQ-VSRLEHA-D)	Three mutations, one on helices 1 to 3
hc4A	MD- (PS-MKQNADS-LHQLARQ-VSRLEHA-D]$_4$	Four mutations, one on each helix, segregated
hc4B	MD- (PS-MKQLADS-LHQNARQ-VSRLEHA-D PS-MKQNADS-LHQLARQ-VSRLEHA-D PS-MKQLADS-LHQNARQ-VSRLEHA-D PS-MKQNADS-LHQLARQ-VSRLEHA-D)	Four mutations, one on each helix, clustered
hc8	MD- [PS-MKQNADS-LHQNARQ-VSRLEHA-D]$_4$	Eight mutations, two on each helix

Red marks site of mutations in comparison to DAMP4

particular in this time frame. DAMP4 has a rather stable rmsd, whereas the other variants are more subjected to structural fluctuations, indicating the lower stability of the variants compared to DAMP4. These results are also reflected in the variance, relative variance and mean rmsd values which can be found in Table 3, where variant hc8 has distinctively higher values. Conclusively, these results clearly show the hydrophobic core as most important factor contributing to DAMP4's stability, whereas salt bridges and the hydration shell contribute in only a minor way to the overall stability.

Sequence design of hydrophobic stability variants for MD simulation

To investigate the previous findings further and to study the influence of number and position of hydrophilic mutations in the core, another set of DAMP4 variants was designed. 1–8 hydrophobic leucine residues were replaced systematically with hydrophilic asparagine residues (Table 2).

Hc1 contains a single hydrophilic point mutation on the first helix. Hc2A has two mutations on helix 1, one in the first heptad and one in the second. To see the difference of mutations on different helices, hc2B was designed with one mutation on helix 1 and one mutation on helix 2. Hc3 has three mutations, one on each of the first three helices. Hc4A and hc4B contain four mutations each; in the first case these mutations are segregated within the core, and in the latter they are clustered. Figure 7 shows a schematic illustration of the allocation of the mutations.

Table 3 Relative variance, variance and mean rmsd C_α as numerical descriptors for rmsd C_α plots depicted in Figs. 5 and 6

	Relative variance rmsd C_α $\frac{STD}{Mean}$ [%]	Variance rmsd C_α STD^2	Mean rmsd C_α
DAMP4	10.0858	0.1439	3.7547
hc1	17.0660	0.4933	4.1155
hc2A	10.0039	0.1688	4.1067
hc2B	8.4170	0.0883	3.5311
hc3	14.5264	0.4130	4.4239
hc4A	16.3894	0.6279	4.8350
hc4B	21.3206	1.1751	5.0844
hc8	18.3966	1.0552	5.5839
sb4A	13.1367	0.3057	4.2088
sb4B	10.7227	0.2012	4.1837
sb4C	12.2776	0.1928	3.5766
ws8A	10.4293	0.1674	3.9227
ws8B	9.9407	0.1357	3.7060

Simulation results for hydrophobic variants

The hydrophobic variants were simulated under the same conditions as the previous stability variants, and the results for α-helix content and rmsd C_α are displayed in Fig. 6. The simulations reveal that variants hc1, hc2A and hc2B lose only minor helical content, and their rmsd C_α values are as well in the same region as DAMP4.

This shows, their structure experienced some distortions and movement of the helices and is marginally

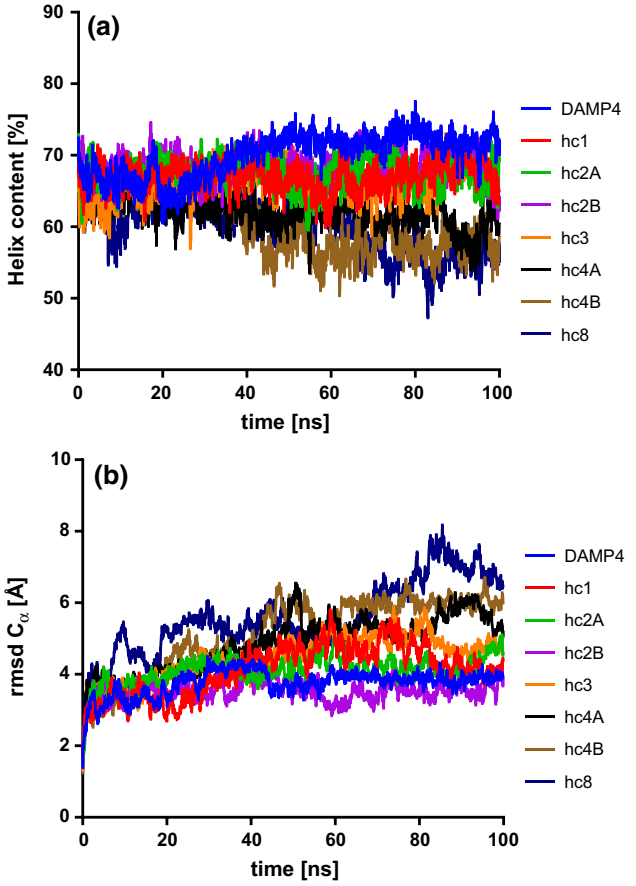

Fig. 6 Simulation results of DAMP4 and hydrophobic core variants. **a** Helical content, and **b** rmsd C_α. These results are the moving average values in 50 ps intervals of all three starting structures. They show strong influence on stability when three or more hydrophobic residues are mutated

affected in their tertiary structure, but the secondary structure remains mainly unaffected in the evaluated time span. It has to be noted, in comparison to salt bridge and water shell variants, rmsd values are comparable or even higher for the hydrophobic variants with 1–3 hydrophilic replacements, despite having a lower number of mutations (see mean rmsd values in Table 3). This shows that the effect of mutations in the hydrophobic core is stronger than mutations disrupting salt bridges or altering the water shell. Variants hc4A and hc4B are more affected by the high temperature and have distinctly lower helical content than DAMP4, whereby hc4B reaches the same extent as hc8, whose core is completely disrupted. Rmsd C_α values are approximately 5 and 6 Å after 100 ns simulation for hc4A and hc4B, respectively, lower than for hc8 with 7 Å. This deviation shows that the effects for hc8 are locally (secondary structure) and globally (tertiary structure) whereas for hc4A and hc4B the local effects are stronger.

To further study the influence of the mutations on the destabilization pathway and the structure, the final simulation conformations were analysed (Fig. 7). To better characterize these changes, Fig. 8 additionally contains the top views, distances between the four helices at the bottom (C_α 5, 46, 53, 94) and the top (C_α 22, 29, 70, 77) as a numerical way to capture tertiary movement in the course of the simulation, and the change in solvent accessible surface area (SASA) compared to DAMP4 for selected variants after 100 ns.

DAMP4 loses some secondary structure in the proceeding simulation, and the structure becomes disordered mainly in the turn regions and at the beginning and end of the strands. The four-helix bundle fold is not affected and only little distortion occurs. In the final structure of hc1 helix 1 clearly stands out by a shift versus the remaining helices. Helix 1 contains the single hydrophilic mutation, which is likely the trigger for this movement and detachment. The extent of this effect is also shown in the top view of hc1 (Fig. 8) and the top distances (Fig. 8c) for helix 1–helix 3, and helix 1–helix 4, increase. Defined secondary structure is lost around the turn regions, but to a low extent. Exposure of residues which are still buried in DAMP4 occurs on helix 4 at the bottom part due to the movement of helix 1, but the effect is minor. Hc2A has, compared to hc1, an additional mutation on the second heptad of helix 1. Considering the three final conformations of this variant, this clearly enhances the destabilization and the influence on the tertiary four-helix bundle fold structure is stronger. Helix 1 shows the same shift as observed for hc1 but due to the stronger nature of this shift it influences helix 2, which gets displaced. The top view of hc2A (Fig. 8a) shows the displacement and influence on the four helices and the loss of their ideally parallel arrangement. The lower red coloured residue on helix 1 corresponds to the mutated Asn 8 and the middle red coloured residue on the same helix to Asn 15 (Fig. 8d), showing a twist of helix 1 exposing these residues to the surrounding water which is not observed for DAMP4. The same shift and extent of distortion is seen in hc2B, with the movement of helix 1 influencing the other three helices, whereas the loss of defined secondary structure is low. With four mutations, one on each helix segregated (hc4A), the effect and the structure become more spherical. The destabilization is stronger on the less protected 'outer' helices 1 and 4, and more loss of helical content occurs. The distance plots show how the structure becomes more flexible after around 50 ns. SASA data shows a twist in helix 1 whereby normally buried hydrophobic residues become exposed, and in turn the relative hydrophilic residues are then buried on the opposite site of the helix. Due to the complete loss of helical structure on the lower part of helix 4, this part becomes completely exposed to the surrounding water. In variant hc4B four

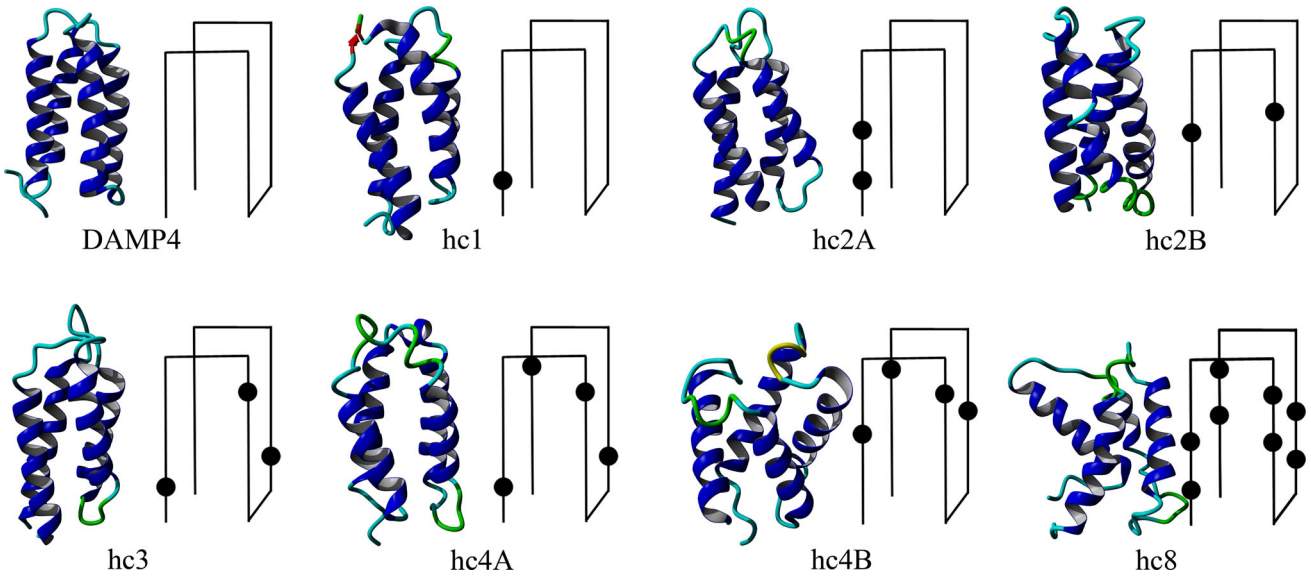

Fig. 7 Selected structures of hydrophobic variants after 100 ns simulation

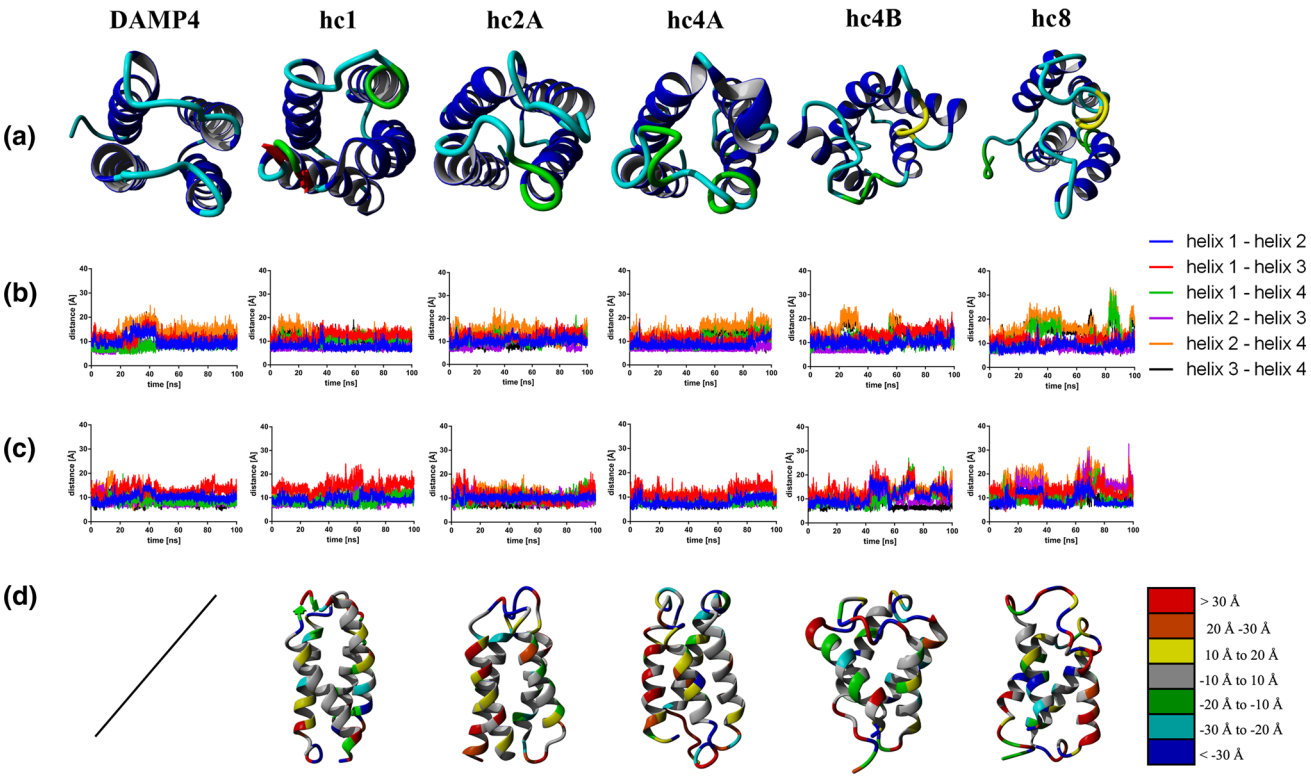

Fig. 8 **a** Top view, designated C_α distances between the four helices at the bottom **b** and top **c** of four-helix bundles and **d** difference in solvent accessible surface area compared to DAMP4 for selected hydrophobic variants

mutations, one on each helix, are clustered at the upper part of the four-helix bundle (see Fig. 7). These mutations have a strong global effect and the four-helix bundle fold is mostly lost. The structure reveals a bursting from the top, in which region the helical structure is lost, showing the local and global effects. The distance plot of the top region reflects this trend, with strong fluctuations in the distances, showing the flexibility of the structure. In hc8, the variant with eight mutations, two on each helix, the effect on the tertiary and secondary structure is obvious. The four

helices completely lose their parallelism, as seen in the top view in Fig. 8a, and defined fold. Starting from the turn regions, major parts of helical content are lost. Strong fluctuations in the distance plots in upper and lower part of the bundle (Fig. 8b, c) suggest the high flexibility and instability of this variant.

These results demonstrated the strong impact on stability of DAMP4 when three or more hydrophobic core residues are mutated. Investigation of the position of the mutations revealed an influence on the unfolding pathway, and that the effect is stronger when the hydrophilic residues are clustered compared to segregation within the core.

Further discussion

In this study we presented a four-helix bundle biosurfactant protein, DAMP4, with extraordinary stability within a broad pH range and even at temperatures above 90 °C.

A wide pH and temperature tolerance is important for industrial processes and applications, and stability in harsh process conditions may be required particularly in downstream processing [51]. Among advantages associated with elevated process temperatures are higher solubility, lower viscosity, higher diffusion rates and a speed increase of kinetic reactions [51]. Further, a highly stable protein is less susceptible to proteolytic degradation, and thus a higher expression yield in a heterologous expression host system is to be expected [15]. In the case of DAMP4, recombinant expression in *E. coli* results in high yield and good solubility, and DAMP4's stability against high temperatures can be used for a simple and cheap purification strategy [10]. These characteristics are of particular interest considering the current demand for low-cost products which are industrially sustainable, eco-friendly, and renewable. Effectiveness at wide pH and temperature range is an advantage compared to chemical surfactants [5].

In a combination of experiments and MD simulation, this study revealed the hydrophobic core as the most important structural feature for the high thermal stability of the four-helix bundle biosurfactant protein. The strong hydrophobic interactions in the core explain why DAMP4 has above-average thermostability, but is only moderately stable against chemical denaturation with GdmCl. Several studies suggest a weakening of hydrophobic interactions by GdmCl [52, 53] and an interaction of the guanidinium molecules with hydrophobic residues in the protein core and thus a shift in equilibrium towards the unfolded state [53–55]. In contrast, kosmotropic salts, for example Na_2SO_4 as used in the thermal process, are known to strengthen hydrophobic interactions, as the energy to solvate hydrophobic residues in the salt solution is even more unfavourable [56]. This results in a decrease in solubility [56], as reinforced hydrophobic interactions on protein surfaces lead to aggregation, as might be seen for bacterial contaminant proteins in the DAMP4 purification process. In contrast, DAMP4 only contains hydrophobic residues in the core, thus Na_2SO_4 stabilizes DAMP4, but leads to a faster denaturation of the bacterial host cell proteins at the elevated temperature. The thermal process is carried out at neutral pH where DAMP4 charges are neutralized and equally distributed in the bundle. However, CD results showed that DAMP4 is stable over the entire pH range at 20 °C, even at extremely low pH where strong electrostatic interactions occur and the molecule is positively charged. It is known that repulsive charges strongly oppose folding [56], confirming the importance of hydrophobic over electrostatic interactions for DAMP4 stability.

From previous studies it is known that DAMP4 folds into a four-helix bundle structure in bulk, whereas DAMP4's monomer DAMP1 (PS-MKQLADS-LHQLARQ-VSRLEHA-D) does not adopt a defined secondary fold and remains unstructured [57]. Whereas electrostatic attraction makes only a minor contribution to the folding into three-dimensional structures of proteins, the main driving forces are polar, van der Waals and hydrophobic interactions [56]. In DAMP4 the hydrophobic interactions are the major contributors to the formation of secondary and tertiary structure through the gain in free energy upon formation of α-helices with a hydrophobic and a hydrophilic plane, and the burial of the water-averse residues in the protein core by folding of the four helices into a bundle. The hydrophobic and hydrophilic planes, built by α-helix formation, provide DAMP4 its surface active characteristics. DAMP4 adsorbs to hydrophobic-hydrophilic interfaces, for example air–water or oil–water, as a monolayer with the thickness of a single helix, indicating an unfolding of the four-helix bundle into four single helices [57]. However, compared to DAMP1, the adsorption kinetic of DAMP4 to the interface is slow, slower than expected for a diffusion controlled process, indicating a large energy barrier for DAMP4 that does not exist for DAMP1 [57]. This barrier is very likely associated with the free energy of unfolding which has to be overcome for adsorption for DAMP4 and is influenced by the strong hydrophobic interactions in the core. This balance of different interactions and forces has to be considered according to the application of new designs which are stable enough for high expression yield and purification, but highly functional with fast adsorption kinetics.

Once adsorbed to the interface, DAMP4 is able to build a stable foam in the presence of kosmotropic salt (Na_2SO_4), but not in the presence of chaotropic salt (NaSCN) [8]. As suggested previously, kosmotropic Na_2SO_4 may share DAMP4's hydration layer and electrostatically stabilize the thin film, and at the same time increase hydrophobic interactions, so the barrier of desorption is higher and the

foam more stable. Compared to that, chaotropic NaSCN may weaken DAMP4's hydration layer and hydrophobic interactions, which leads to a decreased barrier for desorption from the interface and almost no foaming is observed.

As can be seen from DSC and MD, DAMP4 unfolding is a cooperative process, in which secondary and tertiary structure are lost simultaneously. Once its structure is denatured, unfolding in bulk is not reversible, as upon unfolding, exposed hydrophobic residues of different molecules interact strongly with each other, which ultimately leads to the formation of strong aggregates. It is not clear if this process operates only in bulk, and whether interfacial desorption might be associated with a refolding process.

The MD results suggest, the mutation of more than three hydrophobic core residues leads to a rapid loss of stability. However, up to three hydrophilic residues in the core can be tolerated in terms of thermostability. This can be particularly useful in new designs, when higher solubility or faster adsorption kinetics are required.

Conclusion

This research uniquely linked sequence design to stability, process behaviour, and functionality and aimed to broaden the application of sustainable biosurfactants in more industrial products including low-cost industrial sectors. We investigated the stability of a unique biosurfactant four-helix bundle protein, DAMP4, using experimental techniques and MD simulation of the protein and different stability variants. The results suggested that the hydrophobic core is the most important feature for DAMP4's stability and leads to a melting point of 122.4 °C. This fact is exploited in a simple "bake-to-break and precipitate" (BPP) process which allows quick and economical purification to industrial grade DAMP4. Other results, namely the stability against GdmCl, adsorption behaviour, and pH tolerance, support the importance of the hydrophobic core for DAMP4's characteristics. The broad pH and temperature tolerance demonstrated is of high significance for industrial processes and applications and helps to target current problems of biosurfactants. The MD simulations showed an influence of number and location of mutations in the hydrophobic core on stability and unfolding behaviour which may be considered in future sequence designs.

Acknowledgments The authors would like to gratefully acknowledge funding by an Australian Research Council Discovery Grant (DP120103683). A.S. was financially supported by an UQI Tuition Fee Scholarship and an AIBN RHD Living Scholarship. A.P.J.M. thanks the Queensland Government award of the 2010 Smart State Premier's Fellowship. We acknowledge Lei Yu for executing DAMP4 expression and purification. Further, we would like to thank the HPC support from the University of Queensland Research Computing Centre (RCC), the Queensland Cyber Infrastructure Foundation (QCIF) and NCI National Facility for their support and the supercomputer resource allocation.

References

1. Sekhon KK, Khanna S, Cameotra SS (2012) Biosurfactant production and potential correlation with esterase activity. J Pet Environ Biotechnol 3:1–10
2. PRNewswire (2012) Biosurfactants industry is expected to reach usd 2,210.5 million globally in 2018: Transparency market research. http://www.prnewswire.com/news-releases/biosurfactants-industry-is-expected-to-reach-usd-22105-million-globally-in-2018-transparency-market-research-163530426.html. Accessed 23/04/2014
3. Mulligan CN, Sharma SK, Mudhoo A (2014) Biosurfactants: research trends and applications. CRC Press, Boca Raton
4. Marchant R, Banat IM (2012) Biosurfactants: A sustainable replacement for chemical surfactants? Biotechnol Lett 34:1597–1605
5. Makkar RS, Cameotra SS, Banat IM (2011) Advances in utilization of renewable substrates for biosurfactant production. AMB Express 1:1–19
6. Marchant R, Banat IM (2012) Microbial biosurfactants: challenges and opportunities for future exploitation. Trends Biotechnol 30:558–565
7. Syldatk C, Hausmann R (2010) Microbial biosurfactants. Eur J Lipid Sci Technol 112:615–616
8. Middelberg APJ, Dimitrijev Dwyer M (2011) A designed biosurfactant protein for switchable foam control. ChemPhysChem 12:1426–1429
9. Dimitrijev Dwyer M, He L, James M, Nelson A, Wang L, Middelberg APJ (2012) The effects of acid hydrolysis on protein biosurfactant molecular, interfacial, and foam properties: Ph responsive protein hydrolysates. Soft Matter 8:5131–5139
10. Dimitrijev Dwyer M, Brech M, Yu L, Middelberg APJ (2014) Intensified expression and purification of a recombinant biosurfactant protein. Chem Eng Sci 105:12–21
11. Ford NB, Shin DW, Gray HB, Winkler JR (2013) Conformational dynamics of a fast folding cytochrome captured by electron transfer, time-resolved fluorescence energy transfer, and microfluidic mixing. Paper presented at the 245th ACS National Meeting and Exposition, New Orleans, April 7–11, 2013
12. Faraone-Mennella J, Gray HB, Winkler JR (2005) Early events in the folding of four-helix-bundle heme proteins. Proc Natl Acad Sci USA 102:6315–6319
13. Nagi AD, Regan L (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. Fold Des 2:67–75
14. De Vos AM, Ultsch M, Kossiakoff AA (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science 255:306–312
15. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262:1680–1685
16. Choma CT, Lear JD, Nelson MJ, Dutton PL, Robertson DE, DeGrado WF (1994) Design of a heme-binding four-helix bundle. J Am Chem Soc 116:856–865
17. Fukuda M, Komatsu Y, Yamada H, Morikawa R, Miyakawa T, Takasu M, Akanuma S, Yamagishi A (2013) Evaluation of the protein interfaces that form an intermolecular four-helix bundle as studied by computer simulation. Mol Simul 40:498–503

18. Schafmeister CE, LaPorte SL, Miercke LJ, Stroud RM (1997) A designed four helix bundle protein with native-like structure. Nat Struct Mol Biol 4:1039–1046
19. Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH (2003) Stably folded de novo proteins from a designed combinatorial library. Protein Sci 12:92–102
20. Munson M, Regan L, O'Brien R, Sturtevant JM (1994) Redesigning the hydrophobic core of a four-helix-bundle protein. Protein Sci 3:2015–2022
21. Bellesia G, Jewett AI, Shea JE (2011) Relative stability of de novo four-helix bundle proteins: insights from coarse grained molecular simulations. Protein Sci 20:818–826
22. Robinson CR, Sligar SG (1993) Electrostatic stabilization in four-helix bundle proteins. Protein Sci 2:826–837
23. Middelberg AP, Radke CJ, Blanch HW (2000) Peptide interfacial adsorption is kinetically limited by the thermodynamic stability of self association. Proc Natl Acad Sci 97:5054–5059
24. Freites JA, Tobias DJ, von Heijne G, White SH (2005) Interface connections of a transmembrane voltage sensor. Proc Natl Acad Sci USA 102:15059–15064
25. Biswas S, Chakrabarti C, Kundu S, Jagannadham MV, Dattagupta JK (2003) Proposed amino acid sequence and the 1.63 å x-ray crystal structure of a plant cysteine protease, ervatamin b: some insights into the structural basis of its stability and substrate specificity. Proteins Struct Funct Bioinf 51:489–497
26. Shimoni L, Glusker JP (1995) Hydrogen bonding motifs of protein side chains: descriptions of binding of arginine and amide groups. Protein Sci 4:65–74
27. Bedford MT, Clarke SG (2009) Protein arginine methylation in mammals: who, what, and why. Mol Cell 33:1–13
28. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins Struct Funct Bioinf 57:678–683
29. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292:195–202
30. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with yasara nova—a self-parameterizing force field. Proteins Struct Funct Bioinf 47:393–402
31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped blast and psi-blast: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
32. Stenberg KA, Vihinen M (2009) Crystal structure of a 1.6-hexanediol bound tetrameric form of *Escherichia coli* lac-repressor refined to 2.1 å resolution. Proteins Struct Funct Bioinf 75:748–759
33. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in casp8. Proteins Struct Funct Bioinf 77:114–122
34. Hooft R, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272
35. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins Struct Funct Bioinf 17:355–362
36. Wiederstein M, Sippl MJ (2007) Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410
37. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by cα geometry: ϕ, ψ and cβ deviation. Proteins Struct Funct Bioinf 50:437–450
38. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh ewald method. J Chem Phys 103:8577–8593
39. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012
40. Grubmüller H, Tavan P (1998) Multiple time step algorithms for molecular dynamics simulations of proteins: How good are they? J Comput Chem 19:1534–1552
41. Grubmüller H, Heller H, Windemuth A, Schulten K (1991) Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Simul 6:121–142
42. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447
43. Kell G (1967) Precise representation of volume properties of water at one atmosphere. J Chem Eng Data 12:66–69
44. Krieger E, Nielsen JE, Spronk CA, Vriend G (2006) Fast empirical pka prediction by Ewald summation. J Mol Graph Model 25:481–486
45. Day R, Bennion BJ, Ham S, Daggett V (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. J Mol Biol 322:189–203
46. Kumar S, Nussinov R (2001) How do thermophilic proteins deal with heat? Cell Mol Life Sci 58:1216–1233
47. Chen Y-H, Yang JT, Chau KH (1974) Determination of the helix and β form of proteins in aqueous solution by circular dichroism. Biochemistry 13:3350–3359
48. Renkema J, Lakemond CM, de Jongh HH, Gruppen H, van Vliet T (2000) The effect of ph on heat denaturation and gel forming properties of soy proteins. J Biotechnol 79:223–230
49. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132
50. Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 75:422–427
51. Toplak A, Wu B, Fusetti F, Quaedflieg PJ, Janssen DB (2013) Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium coprothermobacter proteolyticus. Appl Environ Microbiol 79:5625–5632
52. Gunari N, Balazs AC, Walker GC (2007) Force-induced globule-coil transition in single polystyrene chains in water. J Am Chem Soc 129:10046–10047
53. Godawat R, Jamadagni SN, Garde S (2010) Unfolding of hydrophobic polymers in guanidinium chloride solutions. J Phys Chem B 114:2246–2254
54. Mason PE, Neilson GW, Enderby JE, Saboungi M-L, Dempsey CE, MacKerell AD, Brady JW (2004) The structure of aqueous guanidinium chloride solutions. J Am Chem Soc 126: 11462–11470
55. Mehrnejad F, Khadem-Maaref M, Ghahremanpour MM, Doustdar F (2010) Mechanisms of amphipathic helical peptide denaturation by guanidinium chloride and urea: A molecular dynamics simulation study. J Comput Aided Mol Des 24: 829–841
56. Jelesarov I, Dürr E, Thomas RM, Bosshard HR (1998) Salt effects on hydrophobic interaction and charge screening in the folding of a negatively charged peptide to a coiled coil (leucine zipper). Biochemistry 37:7539–7550
57. Dimitrijev Dwyer M, He L, James M, Nelson A, Middelberg AP (2013) Insights into the role of protein molecule size and structure on interfacial properties using designed sequences. J R Soc Interface. doi:10.1098/rsif.2012.0987
58. Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R (2007) Clustal w and clustal x version 2.0. Bioinformatics 23:2947–2948