PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Mohanram H, Kumar A, Verma CS, Pervushin K, Miserez A. 2019 Three-dimensional structure of *Megabalanus rosa* Cement Protein 20 revealed by multi-dimensional NMR and molecular dynamics simulations. *Phil. Trans. R. Soc. B* **374**: 20190198.
http://dx.doi.org/10.1098/rstb.2019.0198

Accepted: 29 April 2019

One contribution of 15 to a theme issue 'Transdisciplinary approaches to the study of adhesion and adhesives in biological systems'.

Subject Areas:
structural biology, biomaterials, computational biology, biophysics, biomechanics

Keywords:
barnacle cement proteins, tertiary structure, MrCP20, NMR, molecular dynamics simulations, disulfide bonds

Author for correspondence:
Ali Miserez
e-mail: ali.miserez@ntu.edu.sg

†Equal author contribution.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4587203.

THE ROYAL SOCIETY PUBLISHING

Three-dimensional structure of *Megabalanus rosa* Cement Protein 20 revealed by multi-dimensional NMR and molecular dynamics simulations

Harini Mohanram[1,†], Akshita Kumar[1,2,†], Chandra S. Verma[2,3,4], Konstantin Pervushin[4] and Ali Miserez[1,4]

[1]School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[2]Bioinformatics Institute, Agency for Science, Technology and Research, Singapore 138671, Singapore
[3]Department of Biological Sciences, National University of Singapore, Singapore 117558, Singapore
[4]School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore

HM, 0000-0002-8179-2665; AM, 0000-0003-0864-8170

Barnacles employ a protein-based cement to firmly attach to immersed substrates. The cement proteins (CPs) have previously been identified and sequenced. However, the molecular mechanisms of adhesion are not well understood, in particular, because the three-dimensional molecular structure of CPs remained unknown to date. Here, we conducted multi-dimensional nuclear magnetic resonance (NMR) studies and molecular dynamics (MD) simulations of recombinant *Megabalanus rosa* Cement Protein 20 (*r*MrCP20). Our NMR results show that *r*MrCP20 contains three main folded domain regions intervened by two dynamic loops, resulting in multiple protein conformations that exist in equilibrium. We found that 12 out of 32 Cys in the sequence engage in disulfide bonds that stabilize the β-sheet domains owing to their placement at the extremities of β-strands. Another feature unveiled by NMR is the location of basic residues in turn regions that are exposed to the solvent, playing an important role for intermolecular contact with negatively charged surfaces. MD simulations highlight a highly stable and conserved β-motif (β7-β8), which may function as nuclei for amyloid-like nanofibrils previously observed in the cured adhesive cement. To the best of our knowledge, this is the first report describing the tertiary structure of an extracellular biological adhesive protein at the molecular level.

This article is part of the theme issue 'Transdisciplinary approaches to the study of adhesion and adhesives in biological systems'.

1. Introduction

Underwater bioadhesion is a survival mechanism evolved by diverse marine invertebrates such as mussels, barnacles, tubeworms, etc. [1]. These macrofoulers secrete proteinaceous adhesive holdfast to successfully achieve adhesion under immersed environments [2,3]. Barnacles are one of the most efficient underwater sessile macro-fouling organisms by producing a multi-protein complex 'cement' to firmly attach to solid substrates [4–6]. Unlike many other bioadhesives, the barnacle cement is not known to contain post-translated amino acids, such as Di-hydroxphenylalanine (Dopa) identified in mussel adhesive proteins [5,7]. In *Megabalanus rosa* (*M. rosa*), an acorn barnacle that has been extensively investigated, the adhesive cement is made of at least five different cement proteins (CPs) (called MrCP100, MrCP68, MrCP52, MrCP20 and MrCP19, where the number indicates the molecular weight (MW) of the CP in kDa), that have been sequenced by Kamino [4]. Among these, MrCP19 and 20 have been suggested to act as interfacial layers between solid substrates and the basal plate of barnacles [4], thus playing a central role in the robust bonding that is characteristic of barnacles.

MrCP20 (figure 1*a*) is characterized by a peculiar amino acid composition consisting of high content of charged residues (30% of Asp, Glu, Lys, Arg) with threefold excess of negatively charged residues over positively charged ones. It also contains a relatively high His content (10%) and is uncommonly very abundant in Cys (17%) [8]. Hydrophobic residues are limited, making up 11% (Ala, Val, Leu and Ile) while the remaining is constituted by Ser/Thr/Tyr (11%) and Pro (6%) [8]. Based on the unusually high Cys content, it has been speculated that disulfide bonds may stabilize the protein to maintain a specific spatial arrangement of side-chains involved in molecular interfacial interactions [9]; however, only limited experimental evidences corroborate this hypothesis. Furthermore, the alignment of Cys residues have revealed the presence of six conserved repeats [8], a feature that has been proposed to be involved in the recognition by MrCP20 of specific calcite crystal faces [10]. Indeed, owing to its high acidic residue content (21% of Asp and Glu) MrCP20 has a calculated pI of 4.7 and has been shown to exhibit strong affinity towards calcite [9], in particularly recognizing specific crystal planes [10].

In addition to barnacles, the amino acid sequences of many protein-based biological adhesives have been elucidated [3,11–13]. Surprisingly, there remains a paucity of structural details of adhesive proteins at the molecular level. Adhesive proteins are generally considered to be intrinsically disordered [14,15], although a few spectroscopic and computational studies have pointed out at least some partial ordering, for example, in the mussel adhesive proteins [16] as well as MrCP20 where a certain level of order was suggested [10]. The implication is that functional properties of adhesive proteins are deemed to be largely governed by their chemical activity [17–19], but molecular level structural features are also likely to play a functional role in mediating strong molecular interactions at the adhesive/substrate interface. However, adhesive proteins have seldom been investigated using structural biology methods such as protein nuclear magnetic resonance (NMR) or crystallography, explaining why there is only limited knowledge on their tertiary molecular architecture. In the case of MrCP20, while self-assembly studies have been undertaken [8,10], knowledge of its architecture at the molecular level including specific folds of the repetitive sequences remains unknown. Initial attempts were made by Suzuki *et al.* [20], but the three-dimensional structure was not elucidated. Important structural insights have been obtained by studying the entire cement complex with a combination of atomic force microscopy observations with circular dichroism (CD) and attenuated total reflection Fourier transform infrared spectroscopy [21,22], which have shown that the cement surface comprises amyloid-like nanofibrils. Interestingly, this structural feature is used by other organisms for their adhesive strategies, most notably in bacterial curli fibres [23] and sub-aerial algae [24]. Because functional amyloid fibrils arise from the self-assembly of proteins into cross-β structures (where β-strands are stacked perpendicular to the fibrous axis with inter-strand hydrogen bonds oriented parallel to the axis [25,26]), in the case of the barnacle cement, these nanofibrils must be formed by CPs. However, it remains unclear which CPs are present in these identified nanofibrils and what their adhesive role is.

A first step to answer this central question is to obtain the 3D conformation (tertiary structure) of putative interfacial CPs (MrCP19 and MrCP20) in solution, which represents the precursor state prior to adsorption and self-assembly into nanofibrils onto solid substrates. In this study, we used multi-dimensional solution NMR combined with molecular dynamics (MD) simulations to obtain the solution structure of MrCP20 as previous studies have suggested that this CP is located at the interface, thus playing a key adhesive role. Structural studies of MrCP19, which is also known to be involved in adhesion, will be presented in a separate publication. Our study unveils for the first time, to our knowledge, the tertiary structure of a barnacle CP and provides key insights into its adhesive properties, including its propensity to form amyloid-like fibrils, its dynamic conformation, and the precise role of disulfide bonds that have so far remained elusive.

2. Material and methods

(a) Recombinant protein purification and biophysical characterization of *r*MrCP20

The recombinant expression and purification of recombinant MrCP20 (*r*MrCP20) is presented in detail in the electronic supplementary material, methods. The MW of the purified protein was verified by Matrix Assisted Laser Deionized Time-of-Flight (MALDI-ToF) mass spectrometry using an AXIMA ToF2 (Shimadzu) equipped with an N_2 laser (337 nm, 4 ns pulse width). The particle size distribution was measured by dynamic light scattering (DLS) using a 90Plus Particle size analyzer (Brookhaven Instruments) equipped with a 658 nm monochromatic laser. The measurements were taken at a scattering angle of 90° and the number-weighted histogram profiles were plotted in Origin Pro [27]. Far-ultraviolet CD spectra of protein at various concentrations were collected using a Chirascan spectropolarimeter (Model 420, AVIV Biomedical Inc.). A quartz supracil cell (0.2 mm path length; Helma Analytics) was used for all measurements. Measurements were conducted at 25°C, down to wavelengths between 190 nm to 260 nm using a 1 nm step size and 1 nm bandwidth, with instrument dynode voltage of less than 600. In each case, the background was corrected against the buffer (20 mM Tris, 150 mM NaCl at pH 8.3) employed to solubilize the recombinant protein. Three scans for each sample were averaged, subtracted from the background and plotted in Origin Pro [27].

(b) Nuclear magnetic resonance experiments

All NMR spectra were acquired using a Bruker DRX 800 MHz spectrometer equipped with a cryo-probe and pulse field gradients at 25°C. Two-dimensional ^{1}H-^{15}N HSQC spectra were initially acquired to check the integrity and folding of the protein. A target acquisition-non-uniform sampling method was employed to acquire various backbone (HNCA, HNCACB, HNCO, HN(CA)CO, HN(CO)CA, CACB(CO)NH) and side-chain (NOESY-HSQC) spectra to facilitate resonance assignment and structure calculations of *r*MrCP20. 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was used as an internal reference to calibrate the proton signals. All spectra were processed using Topspin 3.2 (Bruker). Peak picking and spectra analysis were carried out using NMR-FAM Sparky [28]. Automated peak assignment and structure calculations were carried out using CYANA-FLYA [29]. Initial rounds of CYANA-FLYA assigned around 20% of *r*MrCP20 residues. Further assignments were carried out manually by finding correlations between single peaks in ^{15}N-HSQC with backbone NMR spectra. The 3D NOESY spectra were analysed using automated analysis in CYANA. The dihedral angles phi (Φ) and psi (Ψ) were calculated from TALOS+ [30]. All of

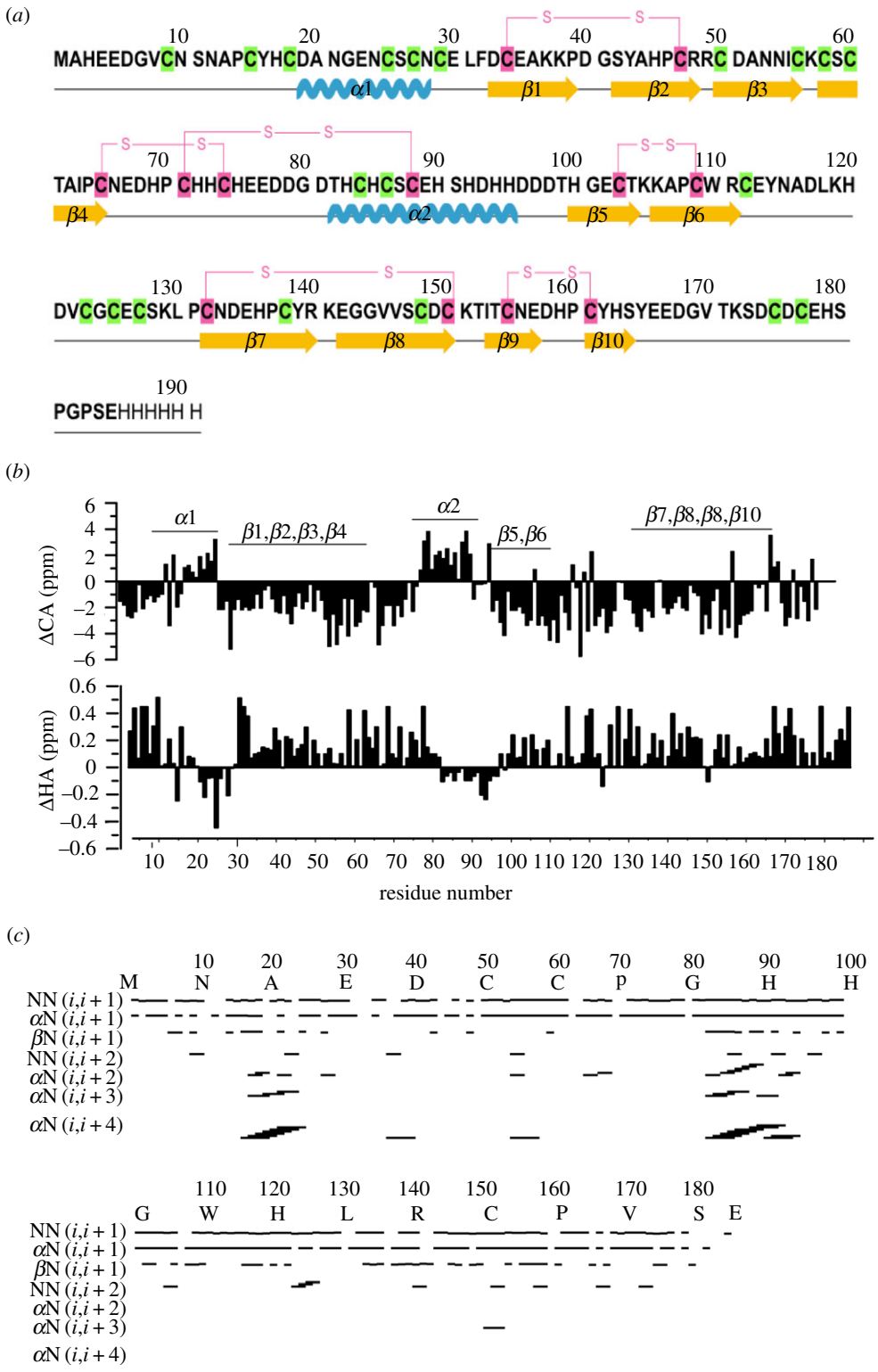

Figure 1. Primary structure and NMR-derived parameters of *r*MrCP20. (*a*) Primary structure of *r*MrCP20. The sulfhydryl groups are highlighted in green and disulfide bonds are highlighted in pink with respective connectivities shown. The secondary structure elements are shown below the respective residues as springs (α-helices), arrows (β-strands) and lines (loops). (*b*) Chemical shift deviations of $^{13}C_\alpha$ (top) and $^{1}H_\alpha$ (bottom) from random coil values. The secondary structure elements of *r*MrCP20 are marked above the bars. (*c*) Bar diagram representation of NOE connectivities used for structure calculation of *r*MrCP20. The helical conformation is supported by strong *i,i* + 4 (αH-NH) connectivities.

the relaxation experiments were carried out in uniformly labelled ^{15}N *r*MrCP20. For T1 (longitudinal spin-lattice relaxation) measurements, the spectra with relaxation delays of 0.05, 0.2, 0.3, 0.5, 0.7, 0.9, 1.2, 1.5, 1.8, 2.0, 2.2 s were recorded. For T2 (spin-spin relaxation) measurements delays of 17, 34, 51, 68, 85, 102, 119, 136, 153 and 170 ms were recorded. The ^{15}N heteronuclear Overhauser effects (NOEs) cross-peaks were obtained with a relaxation delay of 3 s and the ratio of peak intensities with and without saturation of amide protons were plotted against residue number. Urea and 1,4-dithiothreitol (DTT) unfolding studies of *r*MrCP20 were carried out by adding 8 M urea and/or 2 mM DTT solution to ^{15}N-labelled *r*MrCP20 solution. A 2D ^{1}H-^{15}N HSQC spectrum was recorded after each addition.

Hydrogen/deuterium exchange experiments were carried out by adding the appropriate volume of D_2O to freeze-dried ^{15}N-labelled *r*MrCP20 and a series of ^{1}H-^{15}N HSQC spectra

were acquired every 30 min. The extrinsic exchange rates were obtained by fitting the peak intensity versus time to a single-exponential decay equation. The protection factor (PF) was calculated as the ratio of intrinsic exchange rates (calculated from SPHERE [31,32]) to the extrinsic exchange rates.

(c) Molecular dynamics simulation studies

MD simulations were carried out to refine and analyse the NMR structure of *r*MrCP20 using AMBER14 [33] (protocol discussed in detail in the electronic supplementary material, methods). The root mean squared deviation (RMSD) and room mean squared fluctuation (RMSF) calculations were done using *cpptraj* module in AMBER14. Each structure sampled during the MD simulations was superimposed onto the α-carbons of the reference structure (minimized starting structure) and a single PDB file containing all the structures was created. Principal component analysis (PCA) was carried out on the Cα atoms with the Bio3D [34] package using R [35] statistical programming language. The dictionary of secondary structure of proteins (DSSP) algorithm was applied to assign secondary structures using VMD [36]. Structural visualizations were constructed using PyMOL [37] and VMD. RMSD and RMSF plots were constructed using XMGRACE.

3. Results and discussion

(a) Expression, purification and biophysical characterization of *r*MrCP20

*r*MrCP20 was initially purified from proteins expressed in *E. coli* using nickel affinity chromatography. Subsequently, the monomeric His-tagged protein was separated from the oligomeric fractions by size exclusion chromatography. Thus, using a two-step purification protocol, the protein was purified to homogeneity. The purity of the protein was checked with SDS-PAGE gel and the MW was confirmed by MALDI-ToF (electronic supplementary material, figure S1*a*).

Preliminary biophysical characterizations were performed on the purified protein in order to investigate the acquired structural conformation in the aqueous environment. The particle size distribution measured using DLS indicated that *r*MrCP20 existed in a monomeric state, with a mean hydrodynamic diameter D_H of about 2.4 nm (electronic supplementary material, figure S1*b*) and polydispersity of 0.3 (±0.01) (values above 0.7 indicating aggregation). CD spectra of *r*MrCP20 at 4–12 mg ml^{-1} at pH 8.3 (electronic supplementary material, figure S1*c*) exhibited minima at 208 and 222 nm, (characteristic of α-helices [38]) as well as a broad minimum around 218 nm and a maximum at 195 nm (characteristic of β-sheets), the latter being more prominent as the protein concentration increased. These features qualitatively indicated a structural conformation of *r*MrCP20 containing both β-sheets and α-helices and gave us the confidence to pursue a comprehensive analysis of *r*MrCP20 by solution NMR spectroscopy.

(b) Resonance assignment and conformational characteristics of *r*MrCP20

The ^{1}H-^{15}N HSQC spectrum of *r*MrCP20 displayed sharp, intense peaks that spanned through a broad range of chemical shifts from 10.5 to 6.0 ppm, suggesting a preferentially structured conformation of the protein (electronic supplementary material, figure S2). Around 130 well-resolved peaks were identified from 173 residues (excluding 11 Pro). Ninety six per cent of the main chain carbon atoms were assigned by combined analyses of the 3D backbone NMR (HNCA, HNCACB, HNCO, HN(CA)CO, HN(CO)CA, CACB(CO)NH) spectra. The chemical shifts of Pro residues were identified from the CACB(CO)NH spectrum.

An initial secondary structure assignment of *r*MrCP20 was deduced by plotting H$^{\alpha}$ and C$^{\alpha}$ chemical shift deviations obtained from backbone resonance assignments (figure 1*b*). Helical conformations display negative H$^{\alpha}$ and positive C$^{\alpha}$ deviations, whereas β-sheet conformations display positive H$^{\alpha}$ and negative C$^{\alpha}$ deviations of at least five consecutive residues [39]. Based on these rules, residues D^{19}-C^{27} and T^{82}-H^{95} displayed helical conformations whereas the remaining residues were confined to β-sheet or random coil categories (figure 1*b*). The helical conformations were further confirmed with the presence of medium range H$^{\alpha}$ *(i,i + 2; i,i + 3; and i,i + 4)* and NH–NH *(i,i + 1 and i,i + 2)* NOEs (figure 1*c*). The β-sheet conformations were supported by strong NH–NH *(i,i + 1)* resonances and long-range cross-strand NOEs (figure 1*c*).

(c) *r*MrCP20 adopts structural propensities interrupted by dynamic regions

The tertiary structure of *r*MrCP20 was calculated using 1226 NOE constraints, along with 54 hydrogen bond and 18 disulfide bond restraints (table 1). The monomeric structure exhibits a multi-domain conformation composed of 12 structural motifs: 2 α-helices (α1: D^{19}-C^{27}; α2: T^{82}-H^{95}) and 10 β -strands (β1: D^{33}-P^{39}; β2: S^{42}-C^{47}; β3: C^{50}-C^{56}; β4: C^{58}-C^{65}; β5: H^{100}-T^{104}; β6: A^{107}-C^{112}, β7: C^{132}-R^{140}; β8: E^{142}-C^{150}; β9: I^{153}-E^{157} and β10: H^{159}-H^{163}) as shown in figure 2*a*. About 18 residues in the N- and C-termini were found to adopt loop conformations (figure 2*a*). At the molecular level, these structural regions can be categorized as 'domain 1' (α1 with β1–4), 'domain 2' (α2 with β5–6) and 'domain 3' (β7–10) regions (figure 2*b*). The residues between the aforementioned domains exhibited sequential NOEs disrupting the structural integrity. Hence, they are categorized as dynamic loop 'DL1' rich in charged amino acids (N^{66}-D^{81}) and dynamic loop 'DL2' (E^{113}-P^{131}) with a random distribution of hydrophobic and charged amino acids.

Interestingly, four di-peptide basic residues (K^{37}-K^{38}; R^{48}-R^{49}; K^{105}-K^{106}; R^{140}-K^{141}) occupied the turn regions of β-sheets in domains 1, 2 and 3, with their respective side chains exposed to the solvent (figure 2*a*). This configuration may impart electrostatic attraction for negatively charged mineral oxides. Additionally, domain 1 folds into an amphipathic conformation with both hydrophobic and hydrophilic residues distributed on either side of the structure (electronic supplementary material, figure S3*a*,*b*). An 180° rotation of the surface potential map along the *y*-axis revealed a dense negatively charged surface in the centre, surrounded by sparsely distributed basic residues (electronic supplementary material, figure S3*b*). This region is composed of β-sheets in domains 2 and 3. Mapping of charged residues revealed that there may be ionic interactions between residues in DL1 and domain 3 regions, which could be critical for inter-domain packing (electronic supplementary material, figure S3*c*).

Table 1. Structural statistics of 10 lower energy structures of *r*MrCP20.

distance restraints	
intraresidue ($\|i-j\|=0$)	316
sequential ($\|i-j\|=1$)	505
medium range ($2 \leq \|i-j\| \leq 4$)	205
long range ($\|i-j\| \leq 5$)	200
total NOE constraints	1226
hydrogen bond restraints	54
angular restraints	
Φ	145
Ψ	145
deviation from mean structure	
backbone atoms (Å)	6.65
heavy atoms (Å)	7.33
backbone RMSD of structural elements[a]	
helix 1	0.206
helix 2	0.392
β1–2	1.218
β3–4	1.704
β5–6	1.203
β7–8	2.222
β9–10	0.920
Ramachandran plot for the mean structure[b]	
% residues in the most favourable and additionally allowed regions	92.6
% residues in the generously allowed region	7.4
% residues in the disallowed region	0

[a]RMSD values were calculated using the MOLMOL [40] program.
[b]PROCHECK-NMR [41] was used to calculate the Ramachandran plot.

(d) Tertiary structure of *r*MrCP20 is stabilized by disulfide bridges

One of the unique features of *r*MrCP20 is the high Cys content (32 Cys residues corresponding to 17% of the total amino acid content) whose role has so far remained elusive. It has previously been suggested that all Cys residues were oxidized into disulfide bonds, which may help stabilizing the topology of charged residues on the surface of MrCP20 [8]. Our NMR structure shows that only 12 out of 32 Cys residues are oxidized into disulfide bonds (figure 2*a*) while 20 Cys residues exist as reduced sulfhydryl groups (electronic supplementary material, figure S4*a*), identified from their Cβ chemical shifts. The reduced Cβ exhibits an up-field chemical shift of 30 ppm while the oxidized Cβ resonates at a downfield chemical shift of 40–45 ppm [42]. The chemical shifts of the disulfide bonded Cys are reported in the electronic supplementary material, table S1. Four of these bonds (C^{33}-C^{47}, C^{103}-C^{109}, C^{132}-C^{150} and C^{155}-C^{161}) appear to stabilize the β-strands of domains 1, 2 and 3 while the remaining two (between C^{65}-C^{74} and C^{71}-C^{88}) pack domain 1 with DL1 and DL1 with domain 2, respectively. The spatial location of free thiols in *r*MrCP20 tertiary structure (electronic supplementary material, figure S4a) also indicates that it is unlikely to accommodate an additional 10 disulfide bonds because it would introduce steric violations. Reduced thiol groups have also been demonstrated to strongly interact with the calcite mineral phase [43]. Collectively, our data suggest that only a fraction of Cys needs to be oxidized to stabilize *r*MrCP20, with the remaining free thiols available for other types of inter-molecular interactions, and possibly to participate in barnacle shell biomineralization. The distribution of Pro residues was also mapped because oxidized Cys residues are preceded by Pro residues in the *r*MrCP20 sequence (electronic supplementary material, figure S4*a*). As Pro does not participate in hydrogen bonding, it often disrupts β-sheets resulting in the destabilization of tertiary structures of globular proteins [44]. Critically, among the 11 Pro residues of *r*MrCP20, eight are directly preceded by Cys residues, suggesting that the destabilization by Pro is mitigated by adjacent disulfide bonds.

(e) Slow exchanging NHs and low internal motion govern the conformational landscape of *r*MrCP20

The backbone dynamics of *r*MrCP20 in solution was analysed using T1, T2 relaxation and ^{15}N hetero-nuclear NOE experiments of a uniformly ^{15}N labelled sample. Longitudinal spin-lattice (T1) relaxation is related to the intensity decay of magnetic spin (^{15}N) while returning to equilibrium parallel to the magnetic field, whereas transverse spin-spin (T2) relaxation occurs from the intensity decay perpendicular to the magnetic field [45]. Hence, the residues that are actively engaged in the conformational landscape exhibit higher T1 and lower T2 values, explaining the restricted internal motion [46]. The T1 and T2 values of *r*MrCP20 integrated in the 10.0–6.0 ppm range against increasing values of delay times were 923.4 ms and 80.0 ms, respectively (electronic supplementary material, figure S4*b*,*c*). We then fitted the exponential decay of individual amides over the range of delay times. The residues in domains 1, 2 and 3 experienced higher T1 and lower T2 values owing to their restricted mobility (figure 3*a*,*b*), while the first and last 20 residues in the N- and C-terminal regions and DL1–2 regions experienced lower T1 and higher T2 values, denoting a higher degree of flexibility (figure 3*a*,*b*). The mobility of the residues was further validated with the steady-state experimental ^{15}N hetero-nuclear NOE values. The ratio of peak intensities with and without proton saturation can be directly correlated with the amides exhibiting the NOE. The dynamic loop regions exhibited very low or nearly zero experimental NOEs depicting faster motion of the residues in contrast to domains 1–3, which exhibited NOE values as high as 0.9 (figure 3*c*), indicative of restricted mobility of residues.

Slow exchanging NH (amide protons) are reliable indicators of hydrogen bonds. These protons are protected from the solvent [32] and are represented as PFs. The exchange rates of the amide protons were monitored over a range of time after re-suspending the protein in D_2O. The fast exchanging protons on the terminal loops disappeared from the spectrum after 30 min. The amide protons that were involved in hydrogen bonding and subsequent conformational formation were preserved even after 12 h of exchange (figure 3*d*). Interestingly, some amide protons from the dynamic DL1 and DL2 regions were categorized into an intermediate exchange regime owing to their packing with the domains (figure 3*d*).

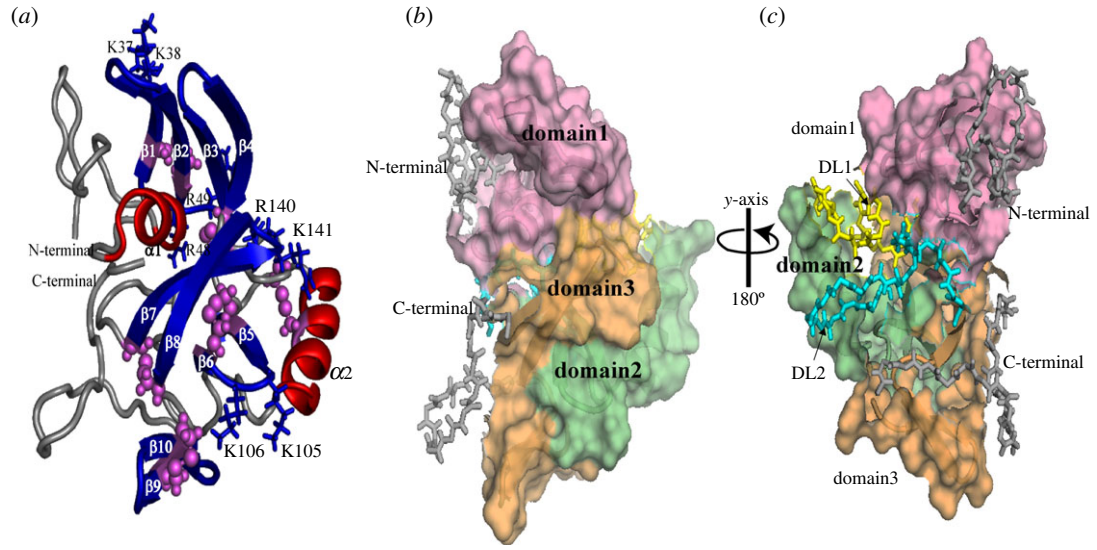

Figure 2. Representative solution structure and surface representation of structural domains and dynamic loop regions of *r*MrCP20. (*a*) Cartoon representation of the tertiary structure of *r*MrCP20 in solution with disulfide linkages represented as pink spheres, α-helices in red and β-sheets in blue. The di-peptide basic residues that occupy the turns of the β-sheets and exposed to the solvent are represented as blue sticks. (*b*) Surface representation of *r*MrCP20 in the same orientation as in (*a*) highlighting the structural core regions (domain 1, domain 2 and domain 3). (*c*) 180° rotation around *y*-axis of model (*b*) displaying the packing of structural domains with dynamic loop regions DL1 and DL2, highlighted in yellow and cyan coloured sticks, respectively.

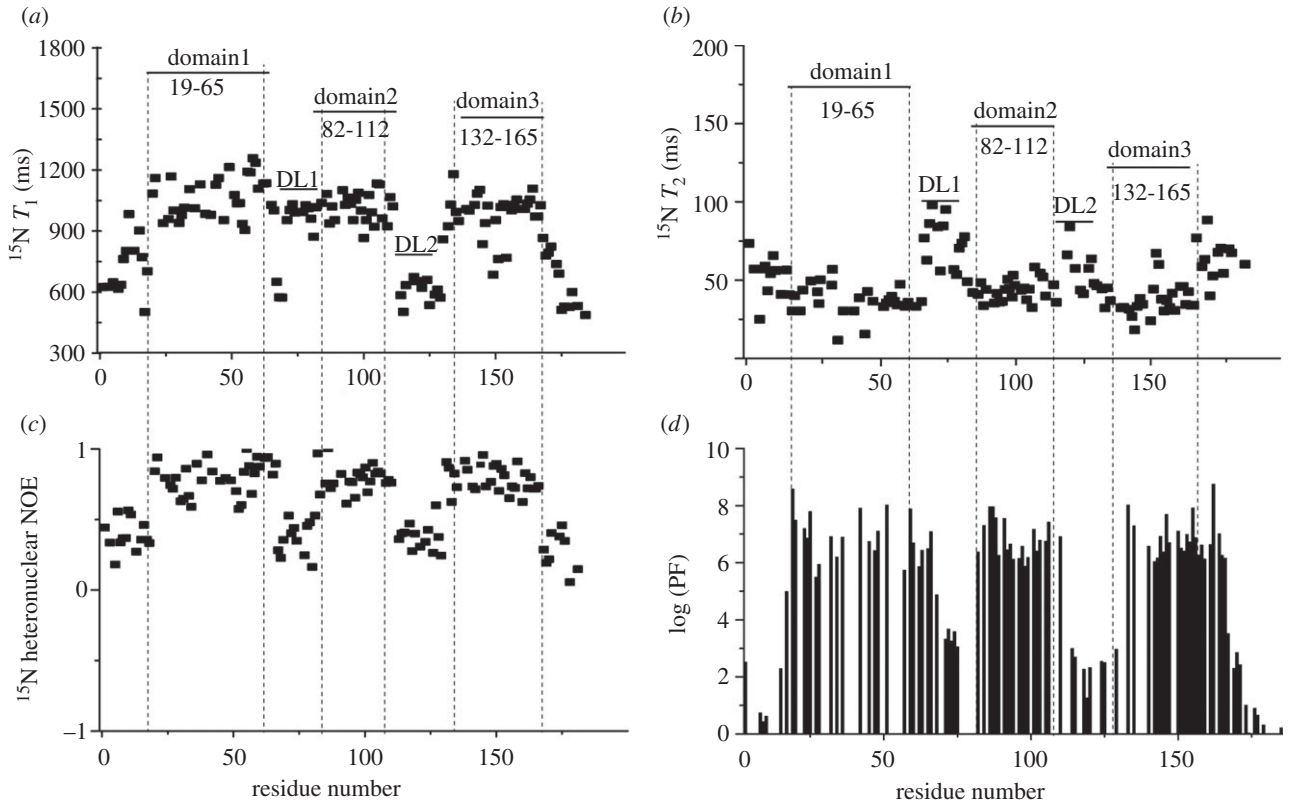

Figure 3. Backbone dynamics of *r*MrCP20. Plots of (*a*) longitudinal spin-lattice T1 relaxation; (*b*) transverse spin-spin T2 relaxation; and (*c*) ^{1}H-^{15}N hetero-nuclear NOEs, as a function of residue number. (*d*) Amide hydrogen-deuterium exchange protection factors shown on a logarithmic plot versus residue number. The positions of the domains and dynamic loop regions of *r*MrCP20 are indicated in *(a)* and *(b)* and separated using vertical dotted lines.

In order to confirm the contribution of hydrogen and disulfide bonds on the structural integrity, *r*MrCP20 was treated with 8 M urea and/or DTT. In the presence of 8 M urea, the ^{1}H-^{15}N HSQC spectra revealed that the protein remained partially folded (electronic supplementary material, figure S5*b*). When the protein was treated with 5 mM DTT (without urea), minimal shifts in the spectrum were observed (electronic supplementary material, figure S5*c*). Complete unfolding of *r*MrCP20 was only observed when both urea and DTT were added (electronic supplementary material, figure S5*d*), which further confirms that the structural integrity of *r*MrCP20 is controlled by

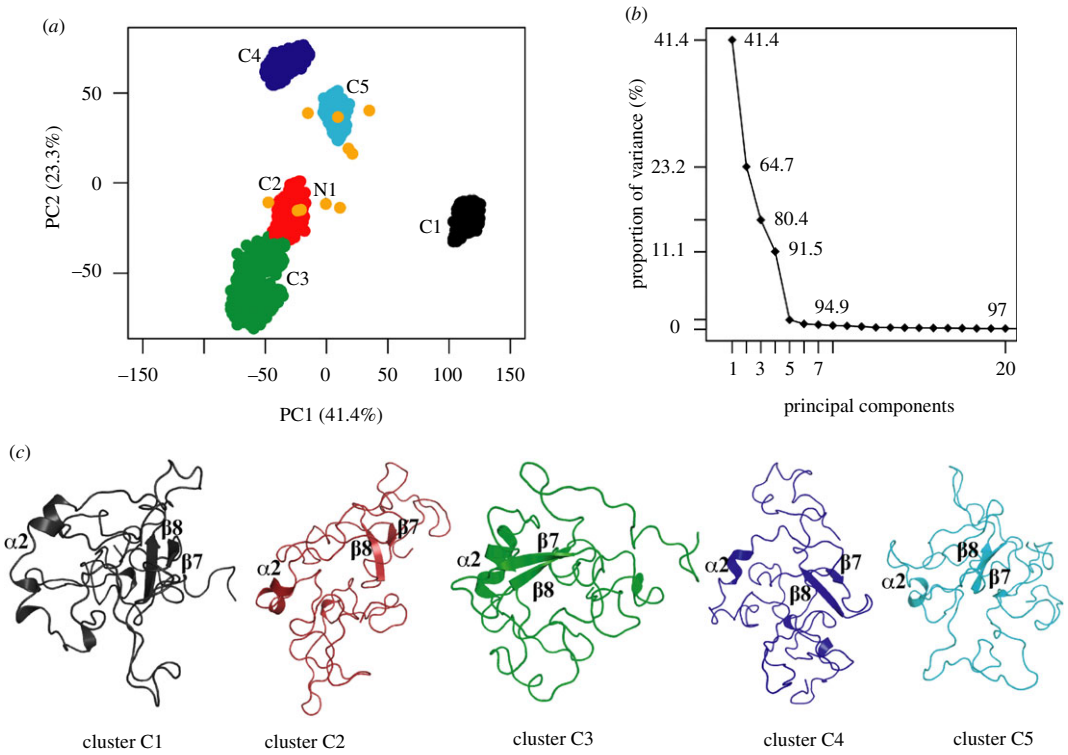

Figure 4. Principal component analysis (PCA) of MD trajectory of *r*MrCP20 and scatter plot of five clusters. (*a*) Distribution of individual conformations along the first two principal component (PC) directions. The MD dataset shows five distinct clusters (C1: black; C2: red; C3: green; C4: blue; C5: cyan). The NMR dataset has 10 structures (represented as orange circles), which are projected onto the MD scatter plot. The starting NMR structure (model 1) for MD simulations is indicated. (*b*) Proportion of variance (in %) captured by each PC. (*c*) Extracted medoid structures from each cluster shown in cartoon representation. The conserved motifs (α2 and β7–8) are marked in each representative structure.

both hydrogen bonds (disrupted by urea) and disulfide bonds (disrupted by DTT).

(f) Structural dynamics of *r*MrCP20 using molecular dynamics simulations

The atomic structure of *r*MrCP20 obtained by NMR (the lowest energy structure) was subjected to energy minimizations using the Amber force field. The minimized structure was subsequently subjected to MD simulations using the AMBER ff14SB force field. The RMSD of the C^{α} atoms relative to the starting structure (electronic supplementary material, figure S6*c*) suggested that the simulations had equilibrated at 300–500 ns region. The spread of RMSD values and the general pattern of RMSF during the MD simulations (electronic supplementary material, figure S6) mirror the structural heterogeneity of the NMR derived ensemble. The average C^{α}-RMS fluctuations of the NMR ensemble showed motions of amplitude greater than 3 Å for the β-motifs and higher mobility in the connecting and flanking loops (electronic supplementary material, figure S6*b*). This pattern shows that *r*MrCP20 is characterized by a dynamic ensemble of tertiary conformations. The coordinates from the 300–500 ns region were subsequently combined into one trajectory. We excluded parts of the N-terminus (residues 1–13) and C-terminus (residues 165–185) from our analysis as they exhibited very high fluctuations (electronic supplementary material, figure S6*d*).

The combined trajectory was subjected to PCA to identify the conformational landscape characterizing the dynamics of the protein in relation to the NMR ensemble (figure 4). The first three principal components (PCs) of the MD simulations accounted for 80% of the motions (figure 4*b*), suggesting that most of the internal motions of the protein can be captured by only a few principal motions [47]. Projections of the sampled conformations along PC1 and PC2 showed five distinct clusters (figure 4*a*) in conformational space and it is interesting to note that the clusters are evenly populated (20% in each cluster).

Regions of the PC1–PC2 plots such as C1 or C3–4 that are sampled during MD in figure 4*a* characterize additional metastable states of this dynamic protein, which may be associated with specific functions tied to varied conditions. Interestingly, of the five simulations carried out, individual PCA and comparison with the landscape of the NMR ensemble showed that at least three simulations mostly covered the landscape of the NMR ensemble (electronic supplementary material, figure S7). Representative conformations (medoids) from each cluster of the combined trajectory were extracted and their structural features analysed (figure 4*c*). It is apparent that the clusters from the simulations had a higher population of disordered structures, with certain secondary structure elements conserved. The partial loss of helical content in the α2 region could be attributed to the high content (greater than 50%) of hydrophilic residues and to the presence of a disulfide bridge connecting the centre of the helix (C^{88}) to the flexible loop DL1 (C^{71}). For the β-sheet motifs identified, the representative conformers exhibited structural destabilization of β1–6 and β9–10 (figure 4*c*). We observed rearrangements of β-sheets resulting in β–bridges stabilized by single hydrogen bonds and disulfide bridges in the vicinity of the respective regions. Even though we observed variable conformational stabilities within the aforementioned motifs, the structural integrity of the β7–β8 motifs were conserved in all conformers. A relatively higher number of NOE signals were observed between the β7–β8 motifs (in comparison to other β motifs) and were used as restraints for NMR structure

calculations. These restraints were not used explicitly for the MD simulations; however, the associated interactions between the *β*7–*β*8 motifs were conserved during the MD simulations (with inter-atomic distances in the range of 2.9–3.5 Å). Thus, it can be speculated that the *β*7–8 motifs are probably the folding nuclei for *r*MrCP20. This is substantiated by observations made by Nakano *et al.* [48] where a peptide from *r*MrCP20 containing the *β*7–8 motifs self-assembled into a mesh-like mesoscopic structure with the help of an intra-molecular disulfide bond at alkaline pH.

Numerous salt bridge interactions were observed between residues of dynamic loops with those of domain regions (electronic supplementary material, table S2) facilitating structural disorder-order transition. It is conceivable that during cement formation, the *β*7–8 motifs cooperatively facilitate intra (or inter) -structural interactions with other small *β*-motifs (*β*1–6, *β*9–10) providing a certain inherent order. In addition, the formation of transient salt-bridges provides dynamic flexibility concomitant with conformational switching in this protein. It is compelling to speculate that these dynamic structural features facilitate the protein assembly into structural scaffolds that in turn act as nuclei to enable the build-up of larger protein structures/assemblies.

4. Conclusion

The solution structure of *r*MrCP20, one of the key components of the barnacle adhesive cement complex, was determined by solution NMR and subsequently analysed by MD simulations. The monomeric structure of *r*MrCP20 is organized into three domains (domains 1, 2 and 3), interrupted by two flexible dynamic (DL1 and DL2) loops. The residues in the structured domains were observed to interact closely with those in the dynamic regions, imparting structural integrity. The partially folded structure of *r*MrCP20 was further evidenced by measuring a D_H value of 2.4 nm, which corresponds to the expected range for 20 kDa globular proteins in the monomeric form [49]. A prominent feature of the tertiary structure of *r*MrCP20 is the packing interactions exerted by the disulfide bonds. Notably, disulfide bonds are often located directly adjacent to Pro residues, which may help mitigate the destabilization effect of Pro on the tertiary structure.

Although *r*MrCP20 is negatively charged at neutral pH, the electrostatic surface potential map reveals a central acidic core surrounded by small clusters of positively charged residues. In particular, dipeptide basic residues within the *β*-sheets in domains 1, 2 and 3 appear to be solvent exposed, which may engage in electrostatic interactions with negatively charged solid surfaces to promote adhesion. In addition, the presence of salt bridges as revealed by both NMR and MD data may contribute to stabilizing interactions between the domains and loops of *r*MrCP20.

PCA of MD simulation results indicate that *r*MrCP20 exists as a dynamic equilibrium of multiple conformations, which could help barnacles adapt to a wide range of substrates. Within these conformations, the motifs *β*7–8 appear to be the most stable of all *β*-strands. Based on the established presence of amyloid-like nanofibrils on the surface of the cured cement, it is tempting to suggest that these stable *β*-sheet motifs could act as a seed for fibrillization of CPs into nanofibrils, but this remains to be validated experimentally.

This study represents a significant step towards understanding the mechanisms of underwater barnacle adhesion at the molecular level, notably by providing valuable insights into the sequence/structure/function relationships of barnacle CP MrCP20 that was previously proposed to play a critical adhesive role.

Data accessibility. Additional data are provided in the electronic supplementary material. Data to support the manuscript can also be obtained from the corresponding author upon reasonable request.

Authors' contributions. A.M., C.S.V. and K.P. designed the study. H.M. and A.K. performed the experiments. H.M., K.P. and A.M. analysed NMR results. A.K. and C.S.V. analysed MD results. H.M., A.K. and A.M. wrote the manuscript.

Competing interests. The authors declare no competing interests.

Funding. This study was funded by the US Office of Naval Research – Global (ONR-G), grant no. N62909-17-1-2045.

Acknowledgement. We thank A*STAR for computational support.

References

1. Flammang P, Santos R. 2015 Biological adhesives: from biology to biomimetics. *Interface Focus* **5**, 20140086. (doi:10.1098/rsfs.2014.0086)
2. Stewart RJ, Ransom TC, Hlady V. 2011 Natural underwater adhesives. *J. Polym. Sci., Part B: Polym. Phys.* **49**, 757–771. (doi:10.1002/polb.22256)
3. Lee BP, Messersmith PB, Israelachvili JN, Waite JH. 2011 Mussel-inspired adhesives and coatings. *Annu. Rev. Mater. Res.* **41**, 99–132. (doi:10.1146/annurev-matsci-062910-100429)
4. Kamino K. 2013 Mini review: barnacle adhesives and adhesion. *Biofouling* **29**, 735–749. (doi:10.1080/08927014.2013.800863)
5. Lee H, Scherer NF, Messersmith PB. 2006 Single-molecule mechanics of mussel adhesion. *Proc. Natl Acad. Sci. USA* **103**, 12 999–13 003. (doi:10.1073/pnas.0605552103)
6. Kamino K, Odo S, Maruyama T. 1996 Cement proteins of the acorn barnacle, *Megabalanus rosa*. *Biol. Bull.* **190**, 403–409. (doi:10.2307/1543033)
7. Naldrett MJ, Kaplan DL. 1997 Characterisation of barnacle (*Balanus eburneus* and *B. cenatus*) adhesive proteins. *Mar. Biol.* **127**, 629–635. (doi:10.1007/s002270050053)
8. Kamino K. 2001 Novel barnacle underwater adhesive protein is a charged aminoacid rich protein constituted by a cys-rich repetitive sequence. *Biochem. J.* **356**, 503–507. (doi:10.1042/bj3560503)
9. Mori Y, Urushida Y, Nakano M, Uchiyama S, Kamino K. 2007 Calcite-specific coupling protein in barnacle underwater cement. *FEBS J.* **274**, 6436–6644. (doi:10.1111/j.1742-4658.2007.06161.x)
10. So CR, Liu J, Fears KP, Leary DH, Golden JP, Wahl KJ. 2015 Self-assembly of protein nanofibrils orchestrates calcite step movement through selective nonchiral interactions. *ACS Nano* **9**, 5782–5791. (doi:10.1021/acsnano.5b01870)
11. Waite JH, Holten-Andersen N, Jewhurst SA, Sun C. 2005 Mussel adhesion: finding the tricks worth mimicking. *J. Adhes.* **81**, 297–317. (doi:10.1080/00218460590944602)
12. Guerette PA *et al.* 2013 Accelerating the design of biomimetic materials by integrating RNA-Seq with proteomics and materials science. *Nat. Biotechnol.* **31**, 908–915. (doi:10.1038/nbt.2671)
13. Hennebert E, Wattiez R, Demeuldre M, Ladurner P, Hwang DS, Waite JH, Flammang P. 2014 Sea star tenacity mediated by a protein that fragments, then aggregates. *Proc. Natl Acad. Sci. USA* **111**, 6317–6322. (doi:10.1073/pnas.1400089111)
14. Hwang DS, Yoo HY, Jun JH, Moon WK, Chan HJ. 2004 Expression of functional recombinant mussel adhesive protein Mgfp-5 in *Escherichia coli*. *Appl.*

Environ. Microbiol. **70**, 3352–3359. (doi:10.1128/AEM.70.6.3352-3359.2004)

15. Waite JH. 2017 Mussel adhesion – essential footwork. *J. Exp. Biol.* **220**, 517–530. (doi:10.1242/jeb.134056)
16. Petrone L *et al.* 2015 Mussel adhesion is dictated by time-regulated secretion and molecular conformation of mussel adhesive proteins. *Nat. Commun.* **6**, 8737. (doi:10.1038/ncomms9737)
17. Maier GP, Rapp MV, Waite JH, Israelachvili JN, Butler A. 2015 Adaptive synergy between catechol and lysine promotes wet adhesion by surface salt displacement. *Science* **349**, 628–631. (doi:10.1126/science.aab0556)
18. Yu J, Wei W, Menyo MS, Masic A, Waite JH, Israelachvili JN. 2013 Adhesion of mussel foot protein-3 to TiO_2 surfaces: the effect of pH. *Biomacromolecules* **14**, 1072–1077. (doi:10.1021/bm301908y)
19. Yu J, Wei W, Danner EW, Israelachvili JN, Waite JH. 2011 Effects of interfacial redox in mussel adhesive protein films on mica. *Adv. Mater.* **23**, 2362–2366. (doi:10.1002/adma.201003580)
20. Suzuki R, Mori Y, Kamino K, Yamazaki T. 2005 Letter to the editor: assignment of 1H, 13C and 15N resonances of barnacle cement protein Mrcp-20 k. *J. Biomol. NMR* **32**, 257. (doi:10.1007/s10858-005-7029-6)
21. Barlow DE, Dickinson GH, Orihuela B, Rittschof D, Wahl KJ. 2009 *In situ* ATR–FTIR characterization of primary cement interfaces of the barnacle *Balanus amphitrite*. *Biofouling* **25**, 359–366. (doi:10.1080/08927010902812009)
22. Barlow DE, Dickinson GH, Orihuela B, Kulp III JL, Rittschof D, Wahl KJ. 2010 Characterization of the adhesive plaque of the barnacle *Balanus amphitrite*: amyloid-like nanofibrils are a major component. *Langmuir* **26**, 6549–6556. (doi:10.1021/la9041309)
23. Barnhart MM, Chapman MR. 2006 Curli biogenesis and function. *Annu. Rev. Microbiol.* **60**, 131–147. (doi:10.1146/annurev.micro.60.080805.142106)
24. Mostaert AS, Giordani C, Crockett R, Karsten U, Schumann R, Jarvis SP. 2009 Characterisation of amyloid nanostructures in the natural adhesive of unicellular subaerial algae. *J. Adhes.* **85**, 465–448. (doi:10.1080/00218460902996366)
25. Kajava AV, Squire JM, Parry DAD. 2006 β-Structures in fibrous proteins. *Adv. Protein Chem.* **73**, 1–15. (doi:10.1016/S0065-3233(06)73001-7)
26. Nelson R, Eisenberg D. 2006 Structural models of amyloid-like fibrils. *Adv. Protein Chem.* **73**, 235–282. (doi:10.1016/S0065-3233(06)73008-X)
27. Origin(Pro) 9.0. Northampton, MA, USA, Origin Lab Corporation.
28. Lee W, Tonelli M, Markley JL. 2015 NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327. (doi:10.1093/bioinformatics/btu830)
29. Schmidt E, Güntert P. 2012 A new algorithm for reliable and general NMR resonance assignment. *J. Am. Chem. Soc.* **134**, 12 817–12 829. (doi:10.1021/ja305091n)
30. Shen Y, Delaglio F, Cornilescu G, Bax A. 2009 TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223. (doi:10.1007/s10858-009-9333-z)
31. Zhang Y-Z. 1995 Protein and peptide structure and interactions studied by hydrogen exchange and NMR. PhD thesis, ProQuest Dissertations Publishing, University of Pennsylvania, Philadelphia, PA, USA.
32. Bai Y, Milne JS, Mayne L, Englander SW. 1993 Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86. (doi:10.1002/prot.340170110)
33. Case DA *et al.* 2014 *AMBER14*. San Francisco, CA: University of California.
34. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006 Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696. (doi:10.1093/bioinformatics/btl461)
35. R Development Core Team. 2011 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
36. Humphrey W, Dalke A, Schulten K. 1996 VMD: visual molecular dynamics. *J. Mol. Graph* **14**, 27–38. (doi:10.1016/0263-7855(96)00018-5)
37. DeLano WL. 2002 The PyMOL Molecular Graphics System. See https://pymol.org/2/.
38. Greenfield NJ. 2006 Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890. (doi:10.1038/nprot.2006.202)
39. Wishart DS, Sykes BD, Richards FM. 1992 The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647–1651. (doi:10.1021/bi00121a010)
40. Koradi R, Billeter M, Wüthrich K. 1996 MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55. (doi:10.1016/0263-7855(96)00009-4)
41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993 PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291. (doi:10.1107/S0021889892009944)
42. Wang Y *et al.* 2018 Solution structure of extracellular loop of human beta 4 subunit of BK channel and its biological implication on ChTX sensitivity. *Sci. Rep.* **8**, 4571. (doi:10.1038/s41598-018-23016-y)
43. Borukhin S, Bloch L, Radlauer T, Hill AH, Fitch AN, Pokroy B. 2012 Screening the incorporation of amino acids into an inorganic crystalline host: the case of calcite. *Adv. Funct. Mater.* **22**, 4216–4224. (doi:10.1002/adfm.201201079)
44. Li SC, Goto NK, Williams KA, Deber CM. 1996 Alpha helical but not beta sheet, propensity of proline is determined by peptide environment. *Proc. Natl Acad. Sci. USA* **93**, 6676–6681. (doi:10.1073/pnas.93.13.6676)
45. Charlier C, Cousin SF, Ferrage F. 2016 Protein dynamics from nuclear magnetic relaxation. *Chem. Soc. Rev.* **45**, 2410–2422. (doi:10.1039/C5CS00832H)
46. Caballero-Manrique E, Bray JK, Deutschman WA, Dahlquist FW, Guenza MG. 2007 A theory of protein dynamics to predict NMR relaxation. *Biophys. J.* **93**, 4128–4140. (doi:10.1529/biophysj.107.111849)
47. Lukman S, Lane DP, Verma CS. 2013 Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. *PLoS ONE* **8**, e80221. (doi:10.1371/journal.pone.0080221)
48. Nakano M, Shen JR, Kamino K. 2007 Self assembling peptide inspired by a barnacle underwater adhesive protein. *Biomacromolecules* **8**, 1830–1835. (doi:10.1021/bm0612236)
49. Nobbmann U. 2016 Zetasizer sensitivity-for protein DLS. See http://www.materials-talks.com/blog/2016/01/21/zetasizer-sensitivity-for-protein-dls/.