Article:
Tych, KM, Batchelor, M, Hoffmann, T et al. (2016) Tuning protein mechanics through an ionic cluster graft from an extremophilic protein. Soft Matter, 12 (10). pp. 2688-2699. ISSN: 1744-683X

https://doi.org/10.1039/C5SM02938D

Journal Name

ARTICLE

Tuning protein mechanics through an ionic cluster graft from an extremophilic protein

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Katarzyna M. Tych[a,b,c], Matthew Batchelor[a,b], Toni Hoffmann[a,b], Michael C. Wilson[a,b], Emanuele Paci[b], David J. Brockwell[b] and Lorna Dougan[a,b*]

Proteins from extremophilic organisms provide excellent model systems to determine the role of non-covalent interactions in defining protein stability and dynamics as well as being attractive targets for the development of robust biomaterials. Hyperthermophilic proteins have a prevalence of salt bridges, relative to their mesophilic homologues, which are thought to be important for enhanced thermal stability. However, the impact of salt bridges on the mechanical properties of proteins is far from understood. Here, a combination of protein engineering, biophysical characterisation, single molecule force spectroscopy (SMFS) and molecular dynamics (MD) simulations directly investigates the role of salt bridges in the mechanical stability of two cold shock proteins; *Bs*CSP from the mesophilic organism *Bacillus subtilis* and *Tm*CSP from the hyperthermophilic organism *Thermotoga maritima*. Single molecule force spectroscopy shows that at ambient temperatures *Tm*CSP is mechanically stronger yet, counter-intuitively, its native state can withstand greater deformation before unfolding (i.e. it is mechanically soft) compared with *Bs*CSP. MD simulations were used to identify the location and quantify the population of salt bridges, and reveal that *Tm*CSP contains a larger number of highly occupied salt bridges than *Bs*CSP. To test the hypothesis that salt-bridges endow these mechanical properties on the hyperthermophilic CSP, a charged triple mutant (CTM) variant of *Bs*CSP was generated by grafting an ionic cluster from *Tm*CSP into the *Bs*CSP scaffold. As expected CTM is thermodynamically more stable and mechanically softer than *Bs*CSP. We show that a grafted ionic cluster can increase the mechanical softness of a protein and speculate that it could provide a mechanical recovery mechanism and that it may be a design feature applicable to other proteins.

Introduction

Mechanical properties, such as strength, toughness and elasticity, are fundamental considerations for the design of robust biomaterials.[1] In recent years, studies on the mechanical properties of proteins found in nature have provided inspiration for the design of biomimetic biopolymers that have a balance of advanced material properties.[2-4] This includes the remarkable combination of high mechanical strength, fracture toughness and elasticity in the muscle protein titin[5-7] and the intriguing mechanical properties of natural silk fibres.[1, 8, 9] These studies are revealing that the mechanical characteristics of proteins are determined by non-covalent interactions which define their unique molecular structure. Understanding the importance and role of these non-covalent interactions is therefore of critical importance for the future development of protein- and biopolymer-based biomaterials. Further inspiration can be gained from extremophilic organisms, or extremophiles, which have the ability to survive and thrive in environments that are considered to be extreme in terms of temperature, pressure, salinity, pH, radiation, or having low levels of oxygen or nutrients.[10] Organisms that are adapted to high temperatures are known as thermophiles (having an optimal growth temperature, T_{OPT}, between ~ 45 and 80 °C) or hyperthermophiles (T_{OPT} above ~ 80 °C). More complex multi-cellular organisms can adapt to these environmental challenges at the tissue or whole-organism level, but micro-organisms usually adapt at the molecular level.[11] While proteins expressed by thermophilic and hyperthermophilic organisms may in some cases be protected against thermoinactivation by the employment of extrinsic stabilising compounds, [12] many others are of more considerable interest as they are able to retain their fold and function at elevated temperatures.[13] These molecules thus offer a unique opportunity to develop the building blocks for biomaterials with advanced mechanical properties and extreme resilience at high temperatures. In addition to their industrial

[a] School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, United Kingdom. Fax: 0113 343 3900; Tel: 0113 343 8958; E-mail: L.Dougan@leeds.ac.uk
[b] Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom.
c Current address: Physik-Department: Lehrstuhl für Biophysik E22 Technische Universität München, Garching, 85748, Germany

† Electronic Supplementary Information (ESI) available: Figures S1–S11 , Tables S1–S7. See DOI: 10.1039/b000000x/

applications, these proteins, by their comparison to mesophilic homologues, (proteins expressed by mesophiles, which have a T_{OPT} between ~ 20 and 45 °C) offer attractive model systems to understand the origin of protein stability.

While there is no clear consensus on the specific structural adaptations that enable proteins from (hyper)thermophiles to retain their fold and function at elevated temperatures, some general principles have been obtained through structural comparisons between thermophiles, hyperthermophiles and their mesophilic structural homologues.[13-17] Relative to their mesophilic counterparts, (hyper)thermophiles exhibit: an increase in the packing density of the hydrophobic core, a decrease in the length and number of unstructured loop regions and an increased number or optimised distribution of salt bridges (bonds between oppositely charged residues that are sufficiently close to each other to experience electrostatic attraction).[17-19] All of these structural adaptations are thought to result in more closely-packed structures, secured by stronger and longer-ranging interactions. A survey of high resolution structures of eighteen protein family pairs from mesophilic and thermophilic organisms identified increased numbers of salt bridges in the thermophilic protein to be the most common difference.[18] Salt bridges in (hyper)thermophiles are optimised by the presence of specific pairwise interactions, such as positively and negatively charged surface amino acid residue pairs, extended ionic networks or global effects such as the net charge of the protein. Enhanced thermostability of proteins can thus be achieved through the optimisation of the long-range electrostatic interactions in salt bridges.[20, 21]

Proteins are dynamic and most are “mechanically soft” as the forces that maintain their structure are weak. This means that, while the rate-limiting activation barrier to unfolding is relatively high, under moderate mechanical forces these proteins can be extended yet maintain a near-native like state and return to the native state upon removal of the perturbation.[22] A conceptually attractive idea is that the high thermal stability of thermophiles is correlated with high mechanical rigidity (or low mechanical softness) of the protein matrix. However, research has both supported and questioned this view.[18, 23-27] For example, hydrogen/deuterium exchange experiments reported enhanced flexibilities of hyperthermophilic proteins, relative to their mesophilic counterparts[23], neutron scattering experiments measured larger atomistic fluctuations in thermophilic proteins than mesophilic proteins,[26] while simulations found that a hyperthermophilic protein exhibited comparable or even enhanced flexibility with respect to a mesophilic domain under ambient temperature conditions.[27]

Given the observed structural adaptations in proteins from hyperthermophiles and the conflicting views of their mechanical properties, we wished to directly examine the impact of specific non-covalent interactions on protein stability (both thermodynamic and mechanical) and mechanical softness. Single molecule force spectroscopy (SMFS) using the atomic force microscope (AFM) is a powerful technique for examining the mechanical characteristics of bio-molecules at the single molecule level.[28-31] In SMFS, the force required to unfold a protein indicates its mechanical stability (*i.e*. the force at which the unfolding energy barrier is overcome) while the sensitivity of the mechanical unfolding force, F_U, to the extension rate can yield coarse insights into the underlying mechanical energy landscape and, consequently, protein softness. Specifically, from the speed dependence of F_U, parameters can be accessed which describe the underlying unfolding energy landscapes of the proteins, namely ΔG_U^* the height of the activation energy barrier and Δx_U the distance between the folded state and the transition barrier on the mechanical unfolding pathway.[29] The term Δx_U provides a measure of the deformation of the native state before unfolding or the mechanical softness of a protein. A large Δx_U describes a mechanically soft protein while a small Δx_U describes a mechanically stiff protein. In the last decade this technique has provided valuable insight into the impact of specific structural adaptations on protein mechanical stability.[32-43] For example, studies have determined the importance of interactions between residues distal in sequence,[41] packing in the hydrophobic core of a proteins,[40] the role of hydrogen bonds[44], solvent accessibility of hydrogen bonds,[45] non-native interactions as well as the identification of strong and weak sequence motifs in protein families.[38, 42, 45-47] The Ainavarapu group used this technique to show that the protein SUMO I can withstand less deformation before unfolding, exhibiting a reduction in mechanical softness upon the binding of a small polypeptide ligand.[37] A recent survey of the experimental literature on the mechanical unfolding of proteins showed a robust correlation between F_U and Δx_U with mechanically strong proteins (large F_U) having a small Δx_U, and mechanically weak proteins (small F_U) a large Δx_U [48]. Using this SMFS approach to measure the impact of specific non-covalent interactions on the mechanical properties, F_U and Δx_U, has the potential to uncover the ‘design features’ of extreme-adapted proteins.

Here, using a combination of SMFS, fluorescence spectroscopy and molecular dynamics (MD) simulations we examine and compare the cold shock protein (*Bs*CSP) from the mesophilic organism *Bacillus subtilis* with the benchmarked cold shock protein B (*Tm*CSP) from the hyperthermophilic organism *Thermotoga maritima.*[36, 43] We obtain information about their thermodynamic and mechanical stability, including information on F_U and Δx_U, and quantify the importance of salt bridges. Informed by these studies, we graft an ionic cluster from *Tm*CSP into *Bs*CSP to make a charged triple mutant (CTM) protein and measure its thermodynamic and mechanical stabilities. Our experiments illustrate the remarkable ability of SMFS to capture previously inaccessible details and demonstrate the potential in using an ionic cluster graft to tune biomaterial properties in hyperthermophilic and mesophilic cold shock proteins.

Methods

Protein engineering and expression

Three different pairs of protein monomer and polyprotein constructs were made, purified and characterised. The polyproteins were constructed using a method we have recently developed which makes use of Gibson Assembly cloning.[49] The polyproteins were purified using a method described previously.[36] An additional anion-exchange stage was included to remove any bound DNA or RNA from the CSP. Accordingly, the protein was bound to a 6 mL Resource Q column (GE Healthcare) in 20 mM phosphate 50 mM NaCl buffer and the protein and DNA eluted separately by a NaCl gradient from 50 mM to 1 M NaCl. For this study, three $(His)_6$-tagged chimeric polyprotein constructs, each containing three domains of a cold shock protein interdigitated with four domains of I27 (the 27th immunoglobulin-like domain of titin) were produced: (i) containing the cold shock protein from the mesophilic organism *Bacillus subtilis* (*Bs*CSP), $(I27\text{-}BsCSP)_3$-I27 (ii) containing the cold shock protein from the hyperthermophilic organism *Thermotoga maritima* (*Tm*CSP), $(I27\text{-}TmCSP)_3$-I27 and (iii) containing the mutated protein which we refer to as a charge triple mutant (CTM) cold shock protein, $(I27\text{-}CTM)_3$-I27. While each polyprotein contains distinct CSP domains, other features such as the $(His)_6$ tag, inter-domain linker sequences, fingerprint I27 domains and two cysteine residues at the C-terminus are identical (see Supporting Information). All three proteins were also expressed as a N-terminal $(His)_6$-tagged monomer and purified as described above.

Protein thermodynamic stability

Chemically induced unfolding transitions of the cold shock proteins were followed using a PTI fluorimeter (Photon Technology International, UK) with a Peltier temperature control and LPS-100 lamp. Protein samples (0.1 mg ml^{-1} in 63 mM sodium phosphate buffer pH 7.4 containing different concentrations of GdnHCl as denaturant) were equilibrated at 23 °C overnight before measurements were recorded. Fluorescence spectra were measured in a 1 cm pathlength quartz cuvette using an excitation wavelength of 280 nm and emission range of 320–380 nm with a 1 nm step size. Unfolding transitions were followed by a change in the barycentric median (BCM). The BCM 'centre of mass' of each spectrum between 320 nm and 380 nm was calculated by equation (1) where $I(\lambda)$ is the fluorescence value at a given wavelength. Each intensity value is multiplied by the respective wavelength and the sum of these values divided by the sum of total intensities.

$$\lambda_{BCM} = \frac{\sum \lambda . I(\lambda)}{\sum I(\lambda)} \quad (1)$$

The BCM values for each spectrum were plotted against denaturant concentration ([D]) and the unfolding transition followed by an increase in BCM due to a shift to a higher wavelength of the unfolded peak. Differences between the relative sum of fluorescence intensities of the folded ($\sum I(\lambda)_F$) and the unfolded ($\sum I(\lambda)_U$) states mean that the BCM signal does not vary linearly with the fraction of folded protein. This is corrected using the quantum yield (Q) as detailed in ref. [50] and using equation 2.

$$Q = \frac{\sum I(\lambda)_F}{\sum I(\lambda)_U} \quad (2)$$

Chemical equilibrium curves were fitted to a two-state unfolding model in Igor Pro (Version 6.34, Wavemetrics, Lake Oswego, OR) using equation 3 where a_F and a_U represent the signal at the start and end of the run and b_F and b_U represent the rate of change of signal with guanidine concentration in the pre-transitional and post-transitional baselines. Q is again the quantum yield, R is the ideal gas constant and m (the 'm-value') is a measure of the (linear) dependence of ΔG_U on denaturant concentration [D]. [51]

$$f([\mathrm{D}]) = \frac{(a_\mathrm{F} + b_\mathrm{F}[\mathrm{D}])(\frac{1}{Q})e^{(\frac{-\Delta G + m[\mathrm{D}]}{RT})} + (a_\mathrm{U} + b_\mathrm{U}[\mathrm{D}])}{1 + (\frac{1}{Q})e^{(\frac{-\Delta G + m[\mathrm{D}]}{RT})}} \quad (3)$$

Force spectroscopy

SMFS was performed using an Asylum MFP-3D AFM (Asylum Research, Santa Barbara, CA, USA). Silicon nitride cantilevers (MLCT) were obtained from Bruker (Billerica, MA, USA). The spring constant of each cantilever was calibrated in buffer, using the equipartition theorem[52] and was found to be within the range of 38 (± 3) pN nm^{-1}. Lyophilized protein (0.1 mg) was reconstituted to a concentration of 0.2 mg ml^{-1} in sterile sodium phosphate buffer (63 mM, pH 7.4) and incubated on a gold substrate for 10 minutes. Mechanical unfolding experiments were performed at pulling velocities of 100, 200, 600 and 2000 nm s^{-1} at room temperature (23 °C) over a distance of 400 nm. Three datasets, each containing at least 34 total unfolding events were accumulated at each pulling velocity using a new sample and cantilever for every dataset. Single molecule force-extension data were filtered to only include traces where one polyprotein chain was seen to unfold, characterised by there being seven or fewer unfolding events. Only traces with a minimum of two I27 unfolding events, without non-specific unbinding events at high force, or other sources of noise were used. Given the interdigitated nature of the polyprotein, the presence of two I27 unfolding events meant that force would be applied to at least one CSP domain.[36, 43] The force-extension data were subsequently analysed using in-house software written for Igor Pro.

Analysis of force-extension data

Mechanical unfolding was analysed using a simple two-state energy landscape which describes the process of converting from the folded state to the unfolded state of the protein over a single energy barrier (the transition state (TS)). The probability of unfolding is governed by the rate of unfolding k_U and the distance from the folded state to the TS along the measured reaction co-ordinate, Δx_U, following the Zhurkov-Bell

model.[53] Monte Carlo simulations were used to generate histograms of unfolding forces for protein domains at a particular pulling velocity and the median values and widths of these distributions were compared to those generated experimentally. The pair of k_U and Δx_U values that provided the best global fit to the experimental data over all pulling velocities was obtained. The uncertainty in the experimental data, defined as the standard deviation in the straight line fit to the dependence of F_U on the pulling speed, was used to quantify the uncertainty in k_U and Δx_U. The range of k_U and Δx_U values that provided a fit to the data within the experimental uncertainty gave the value of the uncertainty for each parameter.

MD simulations

The behaviour of *Bs*CSP, *Tm*CSP and CTM protein was simulated using a united-atom force field (CHARMM19) and implicit solvent model (FACTS). An implicit solvent approach was used both for computational efficiency and to ensure that the non-equilibrium steered extension of the protein is not faster then the relaxation of the solvent around the proteins; also, many simulations could be performed and a statistically representative unfolding mechanism determined. All simulations were performed at 300 K, with Langevin dynamics using a timestep of 2 fs and a friction coefficient of 3 ps^{-1}, and run using CHARMM.[54] A surface tension-like parameter of 0.015 kcal mol^{-1} $Å^{-2}$ was used. Trajectory frames were recorded every 500 steps. Starting structures for simulations were prepared by performing a steepest descent minimisation (1000 steps) from the PDB structure of 1CSP for *Bs*CSP or 1G6P (model #1) for *Tm*CSP, followed by a short (20 ps) dynamics run. The charged triple mutant protein was created by manual conversion of the appropriate residues (E3R, S48E, T64K) from the *Bs*CSP structure, incorporating missing atoms using CHARMM, followed by a steepest descent minimisation/short dynamics run as before. MD simulations were used to model the dynamics of the protein in the native state; five independent 200 ns simulations were performed. To verify the robustness of these simulations, additional simulations were run using (i) the same model but modified implicit solvent parameters, and (ii) an all-atom model with explicit solvent (see Supporting Information for details and results). Simulations to mimic constant velocity force spectroscopy experiments were achieved by applying an external force between the N-terminal N atom and C-terminal C atom; an attached 'cantilever' being moved away at constant velocity (v = 10^8 nm s^{-1}).[55] The spring constant of the cantilever roughly matched those used in AFM experiments (k_c = 30 pN nm^{-1}). AFM-MD simulations were initiated from each of 20 different starting structures extracted from the equilibrium simulations. Wordom (version 21) was used to analyse the trajectories.[56] Secondary structure content values were calculated using continuous secondary structure assignments 'DSSPcont' averaged for each residue and each trajectory frame.[57] Salt bridges were quantified using VMD,[58] and defined using a simple yes/no criterion based on whether the distance between the centre-of-mass of side chain O or N atoms was <0.7 nm apart, allowing for salt bridges to be separated by a gap the size of one water molecule.[59] VMD was also used to locate and quantify hydrogen bonds between strands based on the criterion of an O–N distance <4 Å and an O–H–N angle >150°. Snap-shot images were rendered using VMD.

Results

Structural and thermodynamic comparison of *Bs*CSP and *Tm*CSP. The family of cold shock proteins (CSPs) belong to a sub-set of the OB (oligonucleotide / oligosaccharide-binding) class of folds, a protein fold that is found in all kingdoms of life.[60, 61] The proteins form a β-barrel structure comprising five anti-parallel β-strands arranged in two β-sheets separated by a loop region. Figure 1(a) shows the three-dimensional structure of *Bs*CSP and *Tm*CSP, illustrating the arrangement of the five antiparallel β-strands. To form the β-barrel, the β-sheets twist and coil to form a closed structure in which the first strand is hydrogen-bonded to the fourth strand. The two proteins have high sequence identity (Fig. 1(b)) with 65% residues conserved, 7% added or removed and 28% changed. *Bs*CSP has 19 charged residues while *TmCSP* has 24. To assess their relative thermodynamic stabilities, His-tagged variants of each protein were over-expressed in *E. coli*, purified as described above and subjected to chemical denaturation at equilibrium at room temperature (Methods). As expected and in agreement with previous studies [62] the hyperthermophilic *Tm*CSP displayed greater stability relative to *Bs*CSP (ΔG_U = 25.86 ± 0.63 kJ mol^{-1} and 11.27 ± 0.65 kJ mol^{-1}, respectively (Fig. 2a and 2b, Table S4)).

Mechanical unfolding of *Bs*CSP and *Tm*CSP. The mechanical strength of these homologous proteins was then compared as the relationship between thermodynamic stability and mechanical properties of a protein remains unclear.[40, 48] SMFS experiments were completed using the chimeric polyprotein (I27-*Bs*CSP)$_3$-I27 (Fig. 2(c)) and compared to previously obtained mechanical unfolding data for *Tm*CSP in an analogous scaffold ((I27-*Tm*CSP)$_3$-I27).

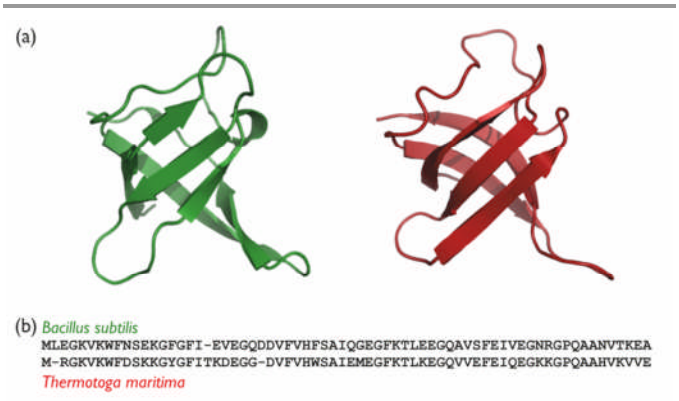

Figure 1. (a) The structure of the cold shock protein domain from the mesophilic organism *Bacillus subtilis* (*Bs*CSP, PDB code 1CSP) (left) and the structure of the cold

shock protein B from the hyperthermophilic bacteria *Thermotoga maritima* (*Tm*CSP, PDB code 1G6P) (right). *Bacillus subtilis* has a T_{OPT} of 37 °C and *Thermotoga maritima* has a T_{OPT} of 80 °C. (b) Sequence alignment of *Bs*CSP and *Tm*CSP.

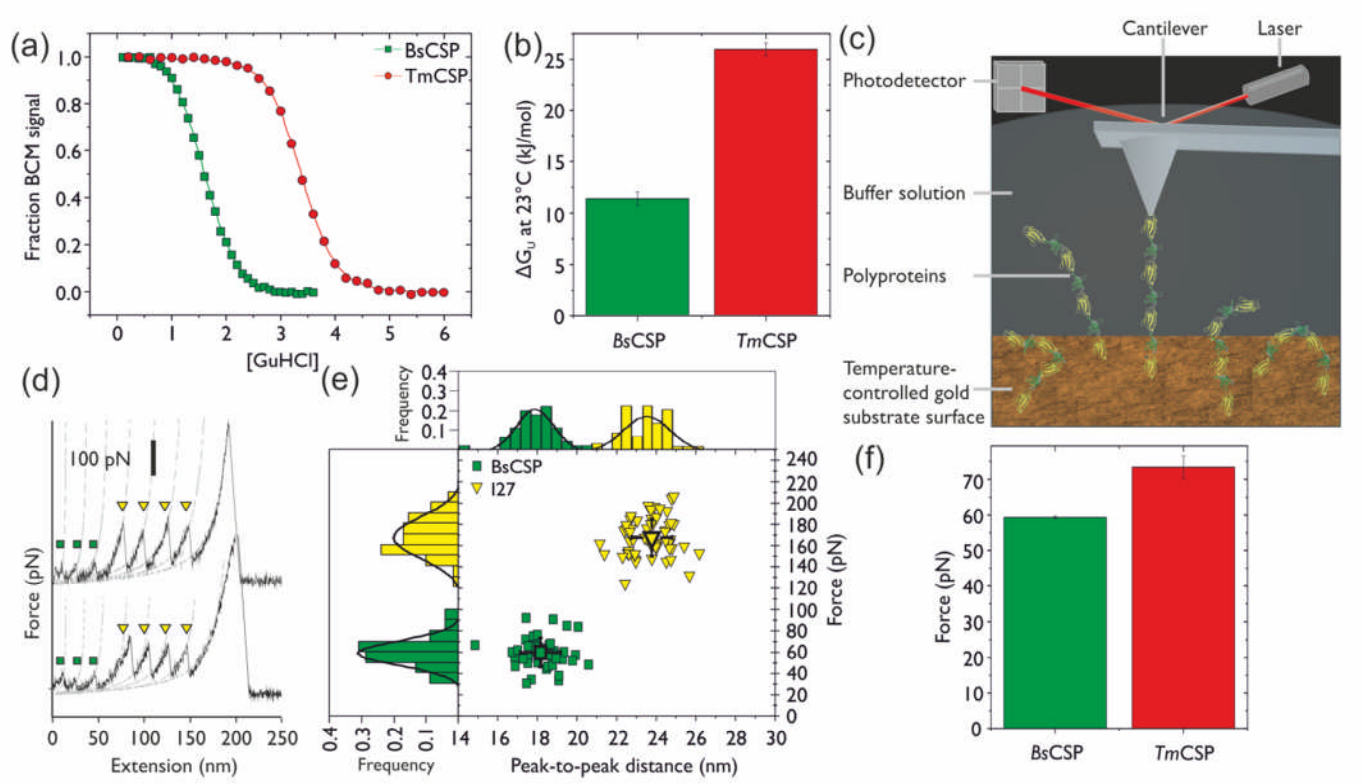

Figure 2 (a) Chemical denaturation experiments monitored by fluorescence emission spectroscopy measure the thermodynamic stability of *Bs*CSP (green) and *Tm*CSP (red) at 23 °C. (b) The mesophilic protein *Bs*CSP (green) is thermodynamically less stable than the hyperthermophilic protein *Tm*CSP (red) at 23 °C. (c) Mechanically unfolding a cold shock protein from a mesophilic organism immobilised on a gold substrate in solution using SMFS. Schematic shows an AFM cantilever picking up a polyprotein ((I27-*Bs*CSP)$_3$-I27) that contains four I27 domains (yellow) and three *Bs*CSP domains (green). (d) A force-extension profile resulting from the mechanical unfolding of the polyprotein at a constant velocity of 600 nm s^{-1}. In the examples, three CSP proteins unfold (green squares) followed by four I27 proteins (yellow triangles). (e) The scatter plots of CSP- and I27-specific unfolding forces and inter-peak distances for 45 CSP unfolding events (green squares) and 58 I27 unfolding events (yellow triangles), are shown together with their respective frequency histograms. Gaussian fits to histograms for each data set are used to obtain a measure of the mode force and inter-peak distance. (f) At a pulling speed of 600 nm s^{-1}, the mesophilic protein *Bs*CSP (green) is mechanically weaker than the hyperthermophilic protein *Tm*CSP [36, 43] (red).

I27 is included in these constructs to improve expression and purification of concatenated CSPs and, in addition, this protein has been extensively investigated using SMFS and consequently has a known mechanical signature. The force-extension (FX) experiments were completed at a pulling speed of 600 nm s^{-1} and a number of FX traces were recorded. Example FX traces are shown in Figure 2(d). The saw-tooth patterns contain two sets of peaks, which differ in both their height (*i.e*. unfolding force, *F*u) and the distance between each unfolding peak (x_{p2p}) defined as the distance between consecutive unfolding peaks at the same force. The x_{p2p} histogram (Fig. 2(e)) shows two distributions centred around 18.2 nm and 23.8 nm and the force-frequency histogram shows two distributions around 59 pN and 166 pN. The unfolding forces and x_{p2p} values are correlated so that two clusters are visible in force-distance scattergrams (Fig. 2e). An unfolding peak with a x_{p2p} ~ 24.0 nm and a F_U of ~ 166 pN (±4 pN, standard deviation between the median values of the triplicate datasets) serves as the mechanical fingerprint for the I27 variant (C47S, C63S, I27)[41] used in this chimeric polyprotein. [36, 41, 43, 49, 55]. Prior to the I27 unfolding events, unfolding peaks are observed with x_{p2p} and F_U values of ~18.2 nm and 59 pN (±1 pN), respectively. This cluster corresponds to the unfolding of the smaller (and weaker) *Bs*CSP. While the x_{p2p} is the same as that measured for *Tm*CSP, the F_U is 24% lower ($F_{U,TmCSP}$ ~ 73 pN).[36, 43] This reduction in the average unfolding force (Fig. 2(f)), despite a similar x_{p2p} suggests that additional interactions are contributing towards an enhanced mechanical stability in *Tm*CSP. Full datasets including all histograms and statistics can be found in the Supporting Information.

Investigating the mechanisms of thermodynamic and mechanical stabilisation of *Tm*CSP. The data obtained above shows that the hyperthermophilic cold shock protein *Tm*CSP is thermodynamically more stable than the mesophilic cold shock protein *Bs*CSP, and also mechanically more stable at a pulling speed of 600 nm s^{-1}. To identify which feature(s) endow *Tm*CSP with these properties we used MD simulations to compare the structural and dynamic properties of *Bs*CSP

and *Tm*CSP in the absence of force (to investigate mechanisms of thermodynamic stabilisation) and under extension to investigate mechanical stabilisation (Fig. 3(a) and (b)).

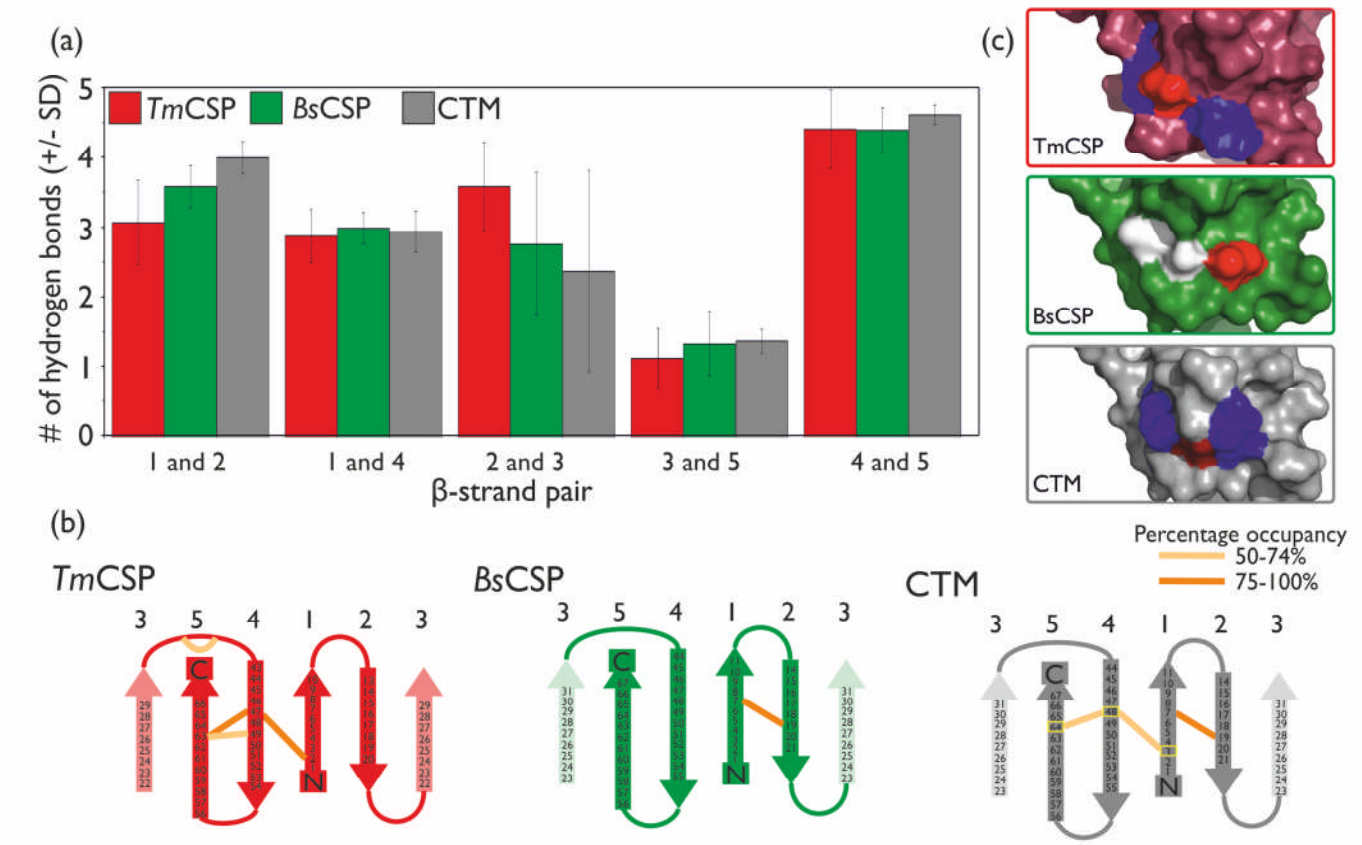

Figure 3 (a) MD simulations of each protein reveal the mean number of hydrogen bonds ± the standard deviation (SD, shown as error bars) over the course of five 200 ns simulations between each pair of β-strands in *Tm*CSP (red bars), *Bs*CSP (green bars) and CTM (grey bars) (b) Topology diagram of the cold shock protein domain highlighting ionic interactions with a mean percentage occupancy > 50 % (orange lines) between β-strands (labelled 1-5) for *Tm*CSP (red), *Bs*CSP (green) and CTM (grey) over five 200 ns simulations. The mutated residues are highlighted by yellow boxes in the CTM protein in positions 3, 48 and 64. On average, the total numbers of salt bridges were: *Tm*CSP 6.3 ± 0.3, *Bs*CSP 3.3 ± 0.4, and CTM 6.3 ± 0.2. (c) An ionic cluster from the hyperthermophilic protein *Tm*CSP (top) is grafted into the mesophile protein *Bs*CSP (middle). The three mutations in CTM (bottom) are designed to form an ionic network spanning the N-terminal β strand 1, β strand 4, and the C-terminal β strand 5. Only residues involved in the ionic cluster mutation are highlighted, where red depicts a negatively charged residue, blue a positively charged residue and white a residue which is neutral.

In the absence of force, the overall secondary structure of the two proteins was very similar, with the same β-sheet content and similar radii of gyration (Table S1 and Fig. S1). We also determined that the hydrophobic cores of the proteins are virtually identical, as discussed in the Supporting Information. Patterns of flexibility in the proteins were highly correlated, as measured by the root mean square fluctuations (RMSF) of Cα atoms, albeit *Tm*CSP showed slightly reduced fluctuations in the loop region in comparison to *Bs*CSP (Fig. S2). We quantified the extent of both hydrogen bond and salt bridge interactions in the two proteins and found that while the mean number of hydrogen bonds between pairs of β-strands (Fig. 3(a)) is largely the same within error, the number of salt bridge interactions and their locations differ significantly (Fig. 3(b)). Plotting the salt bridges which have > 50% probability of occupancy, reveals that there are three such inter-strand salt bridge interactions in *Tm*CSP compared with just one in *Bs*CSP. Importantly, the salt bridges provide an additional link between the terminal strands and strand 4. An increased number of salt bridges has been observed in other studies of thermophilic and mesophilic protein homologues.[13, 63, 64] Additional MD simulations using alternative models (including an all-atom explicitly solvated model) gave very similar results: the secondary structure, solvent-protected core residues, and inter-strand hydrogen bonding are very similar, whilst *Bs*CSP lacks salt bridge interactions between residues on strand 1-4 and 4-5. Details and results of these additional simulations are given in the Supporting Information (Table S2 and S3, Fig. S3-S8).

It is well known that the thermodynamic stability of a protein does not correlate with protein mechanical strength.[40] This is because the former is a global measure of the difference in ground state stability at equilibrium while the latter reports on the kinetic stability of the localised regions of the protein that resist unfolding (the mechanical clamp).[48] The presence of additional salt bridges would thus be expected to only increase the mechanical strength of *Tm*CSP relative to *Bs*CSP if these stabilised the mechanical clamp of this protein. To assess this possibility we completed MD simulations of forced mechanical

unfolding of *Bs*CSP and *Tm*CSP (see Supporting Information for details). The unfolding pathways for *Bs*- and *Tm*CSP were found to be similar and are summarised in Figure 4. For both *Bs*CSP and *Tm*CSP, rupture of β-strands 1-4 or β-strands 4-5 or near simultaneous rupture of both underlie the initial peak in force after which the protein lengthens. Interestingly, ionic interactions are observed in the equilibrium simulations described above for *Tm*CSP but not *Bs*CSP across both of these interfaces (whose relative flux is broadly the same between the homologues). This suggests that ionic interactions may increase the mechanical strength of CSPs. We next sought to test this hypothesis by attempting to enhance the mechanical stability of *Bs*CSP.

Design of the CTM protein. Based on the data presented above, we designed a new protein based on the mesophilic protein *Bs*CSP in which additional features were introduced mimicking that of the hyperthermophilic protein *Tm*CSP. Based on the theoretical predictions of Motono *et al*.,[62] and previous studies where the thermodynamic stability of the mesophilic cold shock protein was tuned by making a range of amino acid substitutions[65-67], we elected to modify the ionic network of the mesophilic protein *Bs*CSP to more closely match that of the hyperthermophilic protein *Tm*CSP. In order to achieve this, we mutated a negatively charged glutamic acid (E) in position 3 to a positively charged arginine (R), a neutral serine (S) at position 48 to a negatively charged glutamic acid (E), and a neutral threonine (T) at position 64 to a positively charged lysine (K). The three substitutions were designed to form an ionic network spanning the terminal β-strands (1 and 5) via their mutual interaction with β-strand 4 (Figs. 3(b) and 3(c)). The equivalent ionic network in the *Tm*CSP structure is part of an ionic cluster, thought to be key in stabilising *Tm*CSP against high temperature denaturation.[62] These mutations were selected on the basis of three considerations: (i) that the stabilisation of the N- and C-terminal β-strands is critical for providing enhanced mechanical stability when a force is applied from the N- and C-termini,[68] and (ii) that ionic interactions between charged amino acid side-chains act over longer distances than hydrogen bonds.[69] Thus, by introducing ionic interactions between strands 1 (the N-terminal strand), 4 and 5 (the C-terminal strand), which have been identified as being important in the MD simulations (Fig. 4), we can predict that we will be forming a structure that is thermodynamically more stable, and can be deformed more before unfolding under an applied force.[36, 70] MD simulations were performed for CTM in analogous fashion to those for *Tm*CSP and *Bs*CSP. The secondary structure, hydrophobic core, RMSF and patterns of inter-strand hydrogen bonding in CTM were equivalent to those found in the other CSPs (see Fig. 3(a) and the Supplementary Information). The only area of real difference was confirmation of the addition of R3-E48 and E48-K64 salt bridges to the *Bs*CSP pattern as per the design (Fig. 3(b)). Similar unfolding pathways were also observed for CTM as for *Bs*CSP/*Tm*CSP in simulations performed under the application of force (Fig. 4).

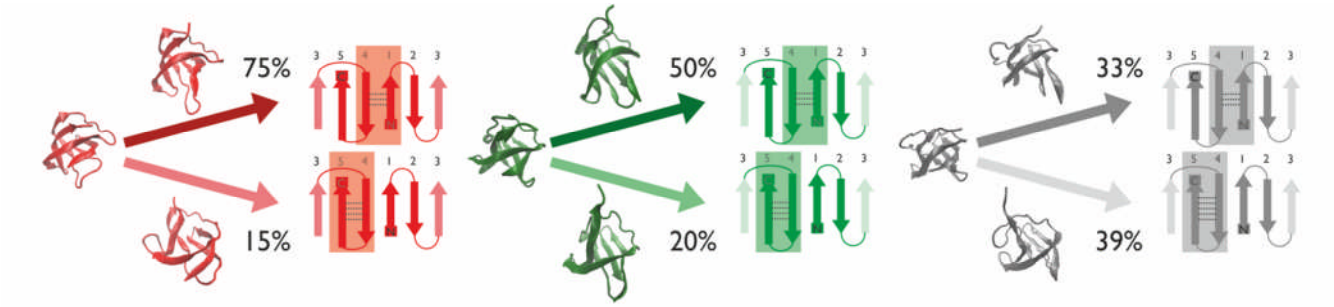

Figure 4 MD simulations of the mechanical unfolding pathway of the cold shock proteins. Topology diagrams of the cold shock protein domain highlighting β-strands 1-5 for *Tm*CSP (red, left), *Bs*CSP (green, middle) and CTM (grey, right). The initial unfolding event occurs as rupture of hydrogen bonds (dashed lines) in β-strands 1-4 or β-strands 4-5 (shaded regions) or near simultaneous rupture of both, identified as an initial peak in force and subsequent lengthening of the protein in the simulations. The flux through each pathway is shown calculated as a percentage.

Thermodynamic and mechanical stability of CTM. CTM was generated by encoding the triple amino-acid substitution described above into the gene for *Bs*CSP. CTM was over-expressed and purified as described for *Bs*- and *Tm*CSP. CD and fluorescence spectroscopies showed that CTM was folded and subsequent equilibrium denaturation experiments using fluorescence revealed that CTM displayed intermediate thermodynamic stability (ΔG_U = 12.34 ± 78 kJ mol^{-1} at 23 °C, see Fig. S9 and Table S4) as expected. MD simulations showed *Bs*CSP and CTM have similar RMSD values for atoms in the backbone regions relevant to the mechanical clamp of the proteins (see Supporting Information). Having characterised the thermodynamic stabilities of all three proteins and having confirmed that the CTM is stable and folded at room temperature, we completed SMFS experiments to obtain information about its mechanical properties. Unfolding the polyprotein construct containing the CTM ((I27-CTM)$_3$-I27) at a constant velocity of 600 nm s^{-1} revealed the familiar saw-tooth pattern, where each tooth corresponds to a single protein domain unfolding (Fig. 5(a)). By examination of the force-extension curves, we identified a clear mechanical fingerprint for the unfolding of CTM, again using I27 as an internal reference in the experiments. Two clusters of correlated F_U–x_{p2p} values were evident once more: events with F_U ~ 176 pN and x_{p2p} ~23.5 nm and events with F_U ~ 51 pN and x_{p2p} ~ 18.6 nm (Fig. 5(b)). The first cluster gives the expected force and

distance values for the reference protein I27.[36, 41, 43, 49, 55] The second cluster yields a distance value matching that found for *Bs*CSP and *Tm*CSP (18.2 and 19.0 nm respectively), suggesting that the mechanical unfolding pathway and the mechanical clamp region for CTM is the same as that of *Bs*CSP and *Tm*CSP with an unfolding force of ~51 pN. This force is significantly lower than that of the hyperthermophilic *Tm*CSP [36, 43] (73 pN) and, more surprisingly, also lower than that of the mesophilic *Bs*CSP at this pulling speed (59 pN).

Pulling speed dependence of unfolding force for *Bs*CSP, *Tm*CSP and CTM. To understand why the CTM variant appeared to be mechanically destabilised despite the presence of the ionic cluster, we performed measurements in triplicate at pulling speeds of 100, 200 and 2000 nm s^{-1} to obtain the pulling speed dependence of F_U from which information about the underlying energy landscape could be deduced (Methods).

These data were compared with those obtained previously for *Tm*CSP under the same experimental conditions.[36] We recorded F_U for each unfolding peak, and constructed separate histograms for every experiment. The unfolding force distributions are a consequence of the thermally assisted stochastic nature of mechanical unfolding and their widths are a conflation of the underlying energy landscape and experimental error.[71] To obviate these problems we adopted the robust approach of completing each experiment in triplicate.[41, 43, 48, 49] The histograms were fitted with Gaussian distributions and the median values of F_U for I27 and both *Bs*CSP and CTM from each of the three replicate experiments were found (Fig. 5(c), S10, and Tables S5 and S6).

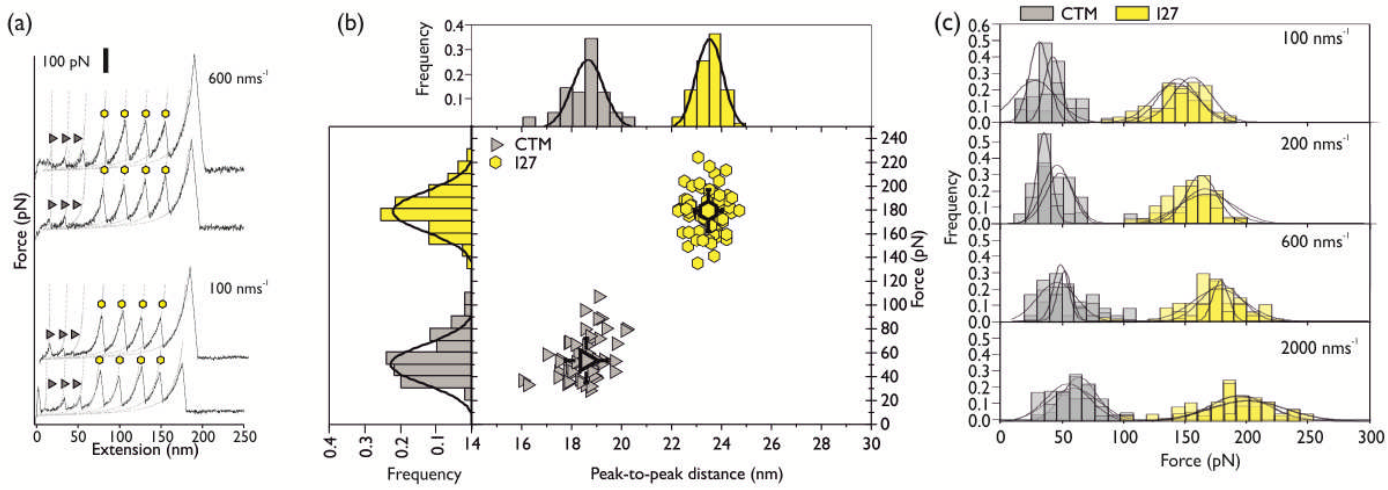

Figure 5 (a) Mechanically unfolding a designed charge triple mutant (CTM) cold shock protein using SMFS. Force-extension profiles resulting from the mechanical unfolding of $(I27\text{-}CTM)_3$-I27 at a constant velocity of 600 nm s^{-1} and 100 nm s^{-1}. In the example traces, three CTM proteins unfold (grey triangles) followed by four I27 proteins (yellow hexagons). (b) The scatter plots of CTM- and I27-specific unfolding forces and inter-peak distances for 50 CTM unfolding events (grey triangles) and 74 I27 unfolding events (yellow hexagons) at 600 nm s^{-1}, are shown in combination with their respective distribution histograms. Gaussian fits to histograms for each data set are used to obtain the unfolding force and peak-to-peak distance. (c) Unfolding force histograms for experiments conducted in triplicate at pulling speeds of 100, 200, 600 and 2000 nm s^{-1} for the $(I27\text{-}CTM)_3$-I27 polyprotein. The histograms show a clear separation in the distributions of the forces resulting from the mechanical unfolding of I27 and the CTM. Gaussian fits to histograms for each data set are used to obtain a measure of the unfolding forces.

To examine the pulling speed dependence of F_U for each protein, we plotted the natural logarithm of pulling speed against the mean F_U (calculated from the medians of the triplicate experiments) for the two populations observed in the F_U histograms. Figure 6(a) shows this plot for the *Bs*CSP, *Tm*CSP and the CTM proteins, and Figure S11 also includes this plot for I27 from data on unfolding $(I27\text{-}TmCSP)_3$-I27, $(I27\text{-}BsCSP)_3$-I27 and $(I27\text{-}CTM)_3$-I27. Inspection of the pulling speed dependence of the unfolding force for all three polyprotein constructs (Fig. 6(a)) shows that the mechanical hierarchy for the cold shock proteins from strongest to weakest is *Tm*CSP > *Bs*CSP > CTM at all pulling speeds. Conversely, the values of F_U for I27 overlay well, which provides an internal control on comparative force measurements from multiple experiments. Detailed inspection of F_U at each pulling speed (Tables S5 and S6) does not identify any mechanical hierarchy for the I27 in the three polyprotein constructs. Therefore, the variation in forces measured for the cold shock proteins is real. The slope of the pulling speed dependence (Fig. 6(a)) is steeper for *Bs*CSP than for *Tm*CSP, meaning that towards lower pulling speeds the difference in the unfolding force tends to increase. Interestingly, although the magnitude of the unfolding force for CTM is less than that of *Bs*CSP at 600 nm s^{-1}, the pulling speed dependency for CTM more closely resembles that of *Tm*CSP. Thus, if the trend observed in Figure 6(a) continues, at lower pulling speeds the unfolding force of CTM will match and then become larger than that of *Bs*CSP.

Unfolding energy landscape of *Bs*CSP, *Tm*CSP and CTM. From the speed dependence of the unfolding force we can access parameters which describe the underlying unfolding energy landscapes of the proteins, namely ΔG_U^* (the height of the

activation energy barrier) and Δx_U (the distance between the folded state and the TS barrier on the mechanical unfolding pathways, see Fig. 6(b).[29] The term Δx_U provides a measure of the deformation of the native state before unfolding or the mechanical softness of a protein. The values of Δx_U, k_U and ΔG_U^* (assuming a value of the pre-factor[72] $A = 10^6$ s^{-1}) are shown in Figure 6(b) and Table S7. Note that the value for *Tm*CSP is different to the previously published value because of an improved Monte Carlo fit. These results show that the hyperthermophilic protein *Tm*CSP has a larger activation energy barrier to unfolding than that of the mesophilic protein *Bs*CSP. Further, the increased value of Δx_U for *Tm*CSP suggests that this protein is mechanically softer than *Bs*CSP.[70] Values for Δx_U, k_U, ΔG_U^* were also obtained for CTM (grey, Fig. 6(b)). We find that the mechanical stability of CTM is relatively unchanged from that of *Bs*CSP, as demonstrated by the small change in the activation energy barrier to unfolding ΔG_U^*.

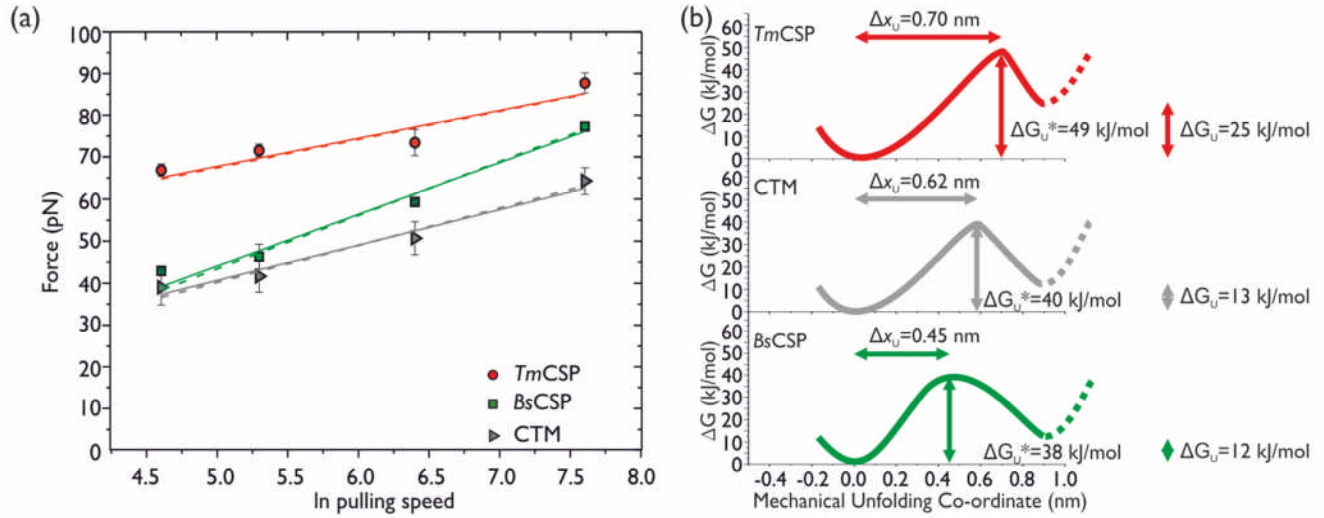

Figure 6 The pulling speed dependence of the three different cold shock proteins reveals details of their underlying unfolding energy landscape. (a) The unfolding force as a function of the logarithm of the pulling speed is shown for the cold shock proteins for each of the three polyprotein constructs studied. Each set of data points at a given pulling speed show the median value of the unfolding force for the cold shock proteins from three experiments conducted under the same conditions. The error bars indicate the standard deviation between the three experiments. Solid lines show the line of best fit to the data. Dashed lines show the Monte Carlo fits to the experimental data. (b) Schematics of the energy landscape with parameters obtained using the experimental results. These include the activation energy barrier height (ΔG^*) and the distance to the unfolding transition state (Δx_U).

However, we have made significant changes to Δx_U where we measure an increase from Δx_U = 0.45 to 0.62 nm (~38%). A larger Δx_U can be interpreted as the protein having a mechanically softer structure, and therefore being able to sustain a greater deformation before unfolding.[70] Note, the changes observed in Δx_U and k_U for the CSP variants contrast with the similar values calculated for the I27 domains present in each scaffold (Table S7).

Discussion

The geometry of the SMFS protein unfolding experiments allows us to define and determine an equivalent "spring constant" of the protein when slightly deformed from its native state under the effect of a force applied to its termini.[37, 43, 70] From the free energy plots in Figure 6(b) it is possible to estimate such a spring constant of the protein along the direction of pulling, using a method described previously.[70] Assuming that the free energy is quadratic in the extension between the two ends:

$$\Delta G^* = 0.5 D \Delta x_U^{\,2} \qquad (6)$$

where D is the spring constant of the protein, we obtain D = 0.32 N m^{-1} for *Tm*CSP and D = 0.60 N m^{-1} for *Bs*CSP. Therefore, the hyperthermophilic protein *Tm*CSP is softer at room temperature than the mesophilic protein *Bs*CSP. Interestingly, the spring constant of the CTM protein reduces by 45% relative to *Bs*CSP to a value of D = 0.33 N m^{-1}. How do the spring constants obtained using SMFS experiments compare with those from other techniques? Neutron scattering experiments measure thermal fluctuations of a protein as a function of temperature and provide access to an effective mean force constant $\langle k \rangle$, reflecting the softness of the protein.[22] Using this approach, values of $\langle k \rangle$ of the order of ~0.1–1.0 N m^{-1} have been measured for a number of proteins.[73] In a study on the protein malate dehydrogenase, the measured value of $\langle k \rangle$ was an order of magnitude higher for the hyperthermophilic protein (1.5 N m^{-1}) than for the mesophilic protein (0.15 N m^{-1}),[26] *i.e.* the mesophilic protein was softer than the hyperthermophilic protein. So while the values of $\langle k \rangle$ are in the same range as our measure of D, the observed trend in increased softness in the hyperthermophilic protein is not the same. However, care must be taken in

comparing *D* and <*k*>. While <*k*> is a measure of the averaged fluctuations around the ground state of a protein, *D* is obtained from SMFS experiments in which the mechanical unfolding pathway is defined by the direction of the applied extension. Previous experiments have shown that the unfolding pathway depends on the pulling geometry, [68] and *D* may thus be highly dependent on the pulling direction. [74] Another significant difference is that <*k*> is obtained by fitting a slope to the measured mean squared displacement as a function of temperature, above the dynamical transition, [22] while in the SMFS experiments, *D* is obtained at a specific temperature, in this case at room temperature. Given the dependence of non-covalent interactions on temperature, the forces in the protein, and the resulting *D* will be temperature dependent. Indeed we recently examined the temperature dependence of the mechanical properties of *Tm*CSP and found that *D* decreased by 55% as the temperature increased from 5 °C to 40 °C, [43] *i.e*. the hyperthermophilic protein became even softer with increasing temperature.

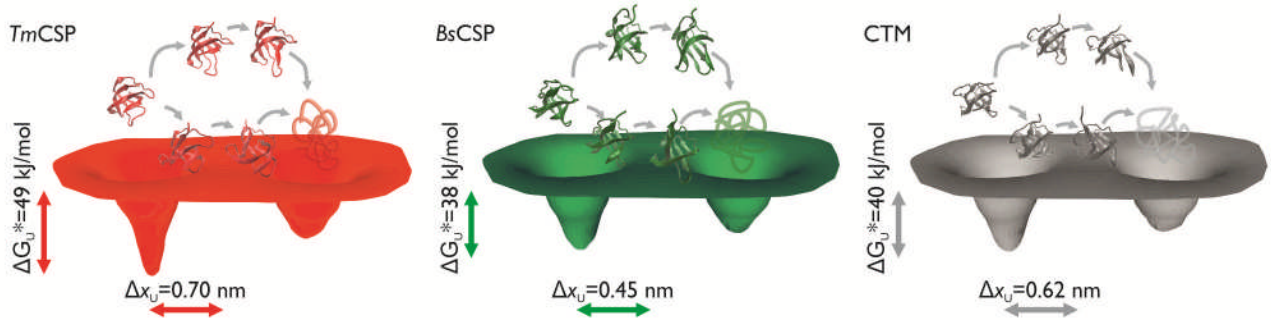

Figure 7 Schematic summary of the results. For each of the three protein domains studied, we have obtained measures of: the free energy difference between the native state and the transition state of the protein along the mechanical unfolding co-ordinate (ΔG_U^*), the free energy difference between the native state and the unfolded state (ΔG_U), and the distance between the native state and the transition state along the mechanical unfolding co-ordinate (Δx_U). Finally, we have identified two key unfolding pathways for each of the protein domains, and used this information alongside that obtained from previous studies to design modifications to a mesophilic protein structure that provide it with enhanced thermodynamic stability and mechanically softness under the application of force.

We hypothesize that enhanced mechanical softness may be an important design feature of some hyperthermophilic proteins, providing them with a mechanical recovery mechanism. This might be particularly important under extreme high-temperature conditions. We propose that the measured increase in the mechanical softness for the CTM protein may reflect the longer-range interactions which are possible due to the grafted ionic cluster. Future experiments are planned to explore other non-covalent interactions to determine their impact on mechanical softness and to understand the range of these interactions in the context of protein mechanical unfolding.

Softness is an essential property of biological scaffolds. [75] Cell behaviour can be controlled by designing material scaffolds that incorporate specific structural and mechanical cues. For example, the mechanics of the extracellular matrix (ECM) has been shown to regulate both short and longer term cell function such as cell motility [76] and recent work has highlighted the importance of the ECM as the main regulator of stem cell differentiation. [77] Multimodular scaffold proteins can act as dynamic switches, for example by assembling proteins into supramolecular complexes, [78] while β-sheet-rich silk proteins possess high toughness for biomaterial applications, such as bone repair. [79] Previous studies have shown that protein softness can be increased by increasing the temperature, [43, 70] or reduced through protein-ligand binding. [37] Here, we demonstrate that protein softness can be tuned through the rational inclusion of salt bridges in the protein. Our experiments illustrate the ability of SMFS to capture important mechanical properties of proteins and quantitatively determine the role of specific non-covalent interactions. Further examination and optimisation of non-covalent interactions in CSPs [62, 65-67] using SMFS could provide a platform of proteins with specific mechanical stability and softness. The ability to tune protein mechanical properties could provide new opportunities to create bespoke scaffolds for biomaterial applications.

Conclusions

We have used a combination of experimental and computational techniques to probe the non-covalent interactions, mechanical and thermodynamic stability of three proteins: *Bs*CSP, *Tm*CSP and a mutated protein, CTM (Fig. 7). In all three proteins we find that the hydrogen bonding between β-strands is similar, and that the mechanical unfolding pathway proceeds through the same mechanism. The hyperthermophilic *Tm*CSP has a larger number of highly occupied salt bridges than *Bs*CSP and is both thermodynamically and mechanically more stable. We observe that the *Tm*CSP is softer than *Bs*CSP. By adding salt bridge interactions to the *Bs*CSP we create the CTM protein and demonstrate that it is thermodynamically stabilised and has increased mechanical softness (Fig. 7). Therefore, we have been successful in re-engineering a protein from a mesophilic organism to exhibit properties more closely mimicking those of a protein from a hyperthermophilic organism. We have also demonstrated that longer-range interactions, such as those

involved in salt bridges, can have a clear effect on the mechanical softness of a protein structure under an applied force.

Acknowledgements

Dr Lorna Dougan is supported by a grant from the European Research Council (258259-EXTREME BIOPHYSICS).

Notes and references

1. J. C. Johnson and L. T. J. Korley, *Soft Matter*, 2012, **8**, 11431-11442.
2. J. Fang and H. B. Li, *Langmuir*, 2012, **28**, 8260-8265.
3. A. M. Kushner, J. D. Vossler, G. A. Williams and Z. B. Guan, *J Am Chem Soc*, 2009, **131**, 8766-8768.
4. S. Lv, D. M. Dudek, Y. Cao, M. M. Balamurali, J. Gosline and H. B. Li, *Nature*, 2010, **465**, 69-73.
5. M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier and C. Bustamante, *Science*, 1997, **276**, 1112-1116.
6. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109-1112.
7. L. Tskhovrebova, J. Trinick, J. A. Sleep and R. M. Simmons, *Nature*, 1997, **387**, 308-312.
8. M. Cetinkaya, S. B. Xiao and F. Grater, *Soft Matter*, 2011, **7**, 8142-8148.
9. F. Vollrath, D. Porter and C. Holland, *Soft Matter*, 2011, **7**, 9595-9600.
10. L. Rothschild, *Nature*, 2002, **417**, 593-593.
11. K. Horikoshi and W. D. Grant, *Extremophiles:Microbial life in extreme environments*, Wiley-Liss, 1998.
12. R. Sterner and W. Liebl, *Crit Rev Biochem Mol*, 2001, **36**, 39-106.
13. A. Szilagyi and P. Zavodszky, *Structure*, 2000, **8**, 493-504.
14. S. A. Wells, S. J. Crennell and M. J. Danson, *Proteins-Structure Function and Bioinformatics*, 2014, **82**, 2657-2670.
15. J. Hollien and S. Marqusee, *Biochemistry-Us*, 1999, **38**, 3831-3836.
16. S. Kumar, C. J. Tsai and R. Nussinov, *Biochemistry-Us*, 2001, **40**, 14152-14165.
17. M. Robinson-Rechavi, A. Alibés and A. Godzik, *Journal of Molecular Biology*, 2006, **356**, 547-557.
18. S. Kumar, C. J. Tsai, B. Y. Ma and R. Nussinov, *J Biomol Struct Dyn*, 2000, **1**, 79-85.
19. P. Zavodszky, J. Kardos, A. Svingor and G. A. Petsko, *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**, 7406-7411.
20. D. Perl and F. X. Schmid, *Journal of Molecular Biology*, 2002, **316**, 213-213.
21. D. Perl, U. Mueller, U. Heinemann and F. X. Schmid, *Nat Struct Biol*, 2000, **7**, 380-383.
22. G. Zaccai, *Science*, 2000, **288**, 1604-1607.
23. G. Hernandez, F. E. Jenney, M. W. W. Adams and D. M. LeMaster, *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 3166-3170.
24. R. Jaenicke, *Proceedings of the National Academy of Sciences*, 2000, **97**, 2962-2964.
25. L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney and J. C. Smith, *Biophysical Journal*, 2008, **94**, 4812-4818.
26. M. Tehei, D. Madern, B. Franzetti and G. Zaccai, *Journal of Biological Chemistry*, 2005, **280**, 40974-40979.
27. M. Kalimeri, O. Rahaman, S. Melchionna and F. Sterpone, *J Phys Chem B*, 2013, **117**, 13775-13785.
28. T. Bornschlogl and M. Rief, *Methods Mol Biol*, 2011, **783**, 233-250.
29. T. Hoffmann and L. Dougan, *Chem Soc Rev*, 2012, **41**, 4781-4796.
30. X. T. Hu and H. B. Li, *Febs Lett*, 2014, **588**, 3613-3620.
31. P. E. Marszalek and Y. F. Dufrene, *Chemical Society Reviews*, 2012, **41**, 3523-3534.
32. F. Berkemeier, M. Bertz, S. B. Xiao, N. Pinotsis, M. Wilmanns, F. Grater and M. Rief, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 14139-14144.
33. M. Bertz, J. Chen, M. J. Feige, T. M. Franzmann, J. Buchner and M. Rief, *Journal of Molecular Biology*, 2010, **400**, 1046-1056.
34. X. Gao, M. Qin, P. G. Yin, J. Y. Liang, J. Wang, Y. Cao and W. Wang, *Biophysical Journal*, 2012, **102**, 2149-2157.
35. C. Z. He, G. Lamour, A. Xiao, J. Gsponer and H. B. Li, *J Am Chem Soc*, 2014, **136**, 11946-11955.
36. T. Hoffmann, K. M. Tych, D. J. Brockwell and L. Dougan, *J Phys Chem B*, 2013, **117**, 1819-1826.
37. H. C. Kotamarthi, A. Yadav and R. K. Ainavarapu, *Biophysical Journal*, 2015, **108**, 360-367.
38. W. Lu, S. S. Negi, A. F. Oberhauser and W. Braun, *Proteins*, 2012, **80**, 1308-1315.
39. M. Mickler, R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai and M. Rief, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 20268-20273.
40. S. P. Ng, K. S. Billings, T. Ohashi, M. D. Allen, R. B. Best, L. G. Randles, H. P. Erickson and J. Clarke, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 9633-9637.
41. D. P. Sadler, E. Petrik, Y. Taniguchi, J. R. Pullen, M. Kawakami, S. E. Radford and D. J. Brockwell, *J Mol Biol*, 2009, **393**, 237-248.
42. D. Sharma, O. Perisic, Q. Peng, Y. Cao, C. Lam, H. Lu and H. B. Li, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 9278-9283.
43. K. M. Tych, T. Hoffmann, D. J. Brockwell and L. Dougan, *Soft Matter*, 2013, **9**, 9016-9025.
44. L. Dougan, K. R. Ainavarapu, G. Genchev, H. Lu and J. M. Fernandez, *Chemphyschem*, 2008, **9**, 2836-2847.
45. D. L. Guzmán, A. Randall, P. Baldi and Z. Guan, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 1989-1994.
46. H. Li and J. M. Fernandez, *Journal of Molecular Biology*, 2003, **334**, 75-86.
47. M. M. Balamurali, D. Sharma, A. Chang, D. Khor, R. Chu and H. Li, *Protein science : a publication of the Protein Society*, 2008, **17**, 1815-1826.
48. T. Hoffmann, K. M. Tych, M. Hughes, D. J. Brockwell and L. Dougan, *Physcial Chemistry Chemical Physics*, 2013, **15**, 15767-15780.

49. T. Hoffmann, K. M. Tych, T. Crosskey, R. Schiffrin, D. J. Brockwell and L. Dougan, *Acs Nano*, 2015.
50. C. P. Moon and K. G. Fleming, *Method Enzymol*, 2011, **492**, 189-211.
51. J. M. Scholtz, G. R. Grimsley and C. N. Pace, *Methods in Enzymology, Vol 466: Biothermodynamics, Pt B*, 2009, **466**, 549-565.
52. E. L. Florin, M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, V. T. Moy and H. E. Gaub, *Biosens Bioelectron*, 1995, **10**, 895-901.
53. G. I. Bell, *Science*, 1978, **200**, 618-627.
54. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545-1614.
55. M. Wolny, M. Batchelor, P. J. Knight, E. Paci, L. Dougan and M. Peckham, *Journal of Biological Chemistry*, 2014, **289**, 27825-27835.
56. M. Seeber, M. Cecchini, F. Rao, G. Settanni and A. Caflisch, *Bioinformatics*, 2007, **23**, 2625-2627.
57. P. Carter, C. A. F. Andersen and B. Rost, *Nucleic Acids Research*, 2003, **31**, 3293-3295.
58. W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph Model*, 1996, **14**, 33-38.
59. J. Dzubiella, *J Am Chem Soc*, 2008, **130**, 14000-14007.
60. P. L. Graumann and M. A. Marahiel, *Trends Biochem Sci*, 1998, **23**, 286-290.
61. G. Horn, R. Hofweber, W. Kremer and H. Kalbitzer, *Cell Mol Life Sci*, 2007, **64**, 1457-1470.
62. C. Motono, M. M. Gromiha and S. Kumar, *Proteins*, 2008, **71**, 655-669.
63. C.-F. Lee, M. D. Allen, M. Bycroft and K.-B. Wong, *Journal of Molecular Biology*, 2005, **348**, 419-431.
64. J. M. Sanchez-Ruiz and G. I. Makhatadze, *Trends in Biotechnology*, 2001, **19**, 132-135.
65. B. N. Dominy, D. Perl, F. X. Schmid and C. L. Brooks, *Journal of Molecular Biology*, 2002, **319**, 541-554.
66. D. Perl, C. Welker, T. Schindler, K. Schroder, M. A. Marahiel, R. Jaenicke and F. X. Schmid, *Nat Struct Mol Biol*, 1998, **5**, 229-235.
67. M. Wunderlich, A. Martin and F. X. Schmid, *Journal of Molecular Biology*, 2005, **347**, 1063-1076.
68. D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham and S. E. Radford, *Nat Struct Biol*, 2003, **10**, 731-737.
69. M. F. Thorpe, M. Lei, A. J. Rader, D. J. Jacobs and L. A. Kuhn, *Journal of Molecular Graphics and Modelling*, 2001, **19**, 60-69.
70. M. Schlierf and M. Rief, *Journal of Molecular Biology*, 2005, **354**, 497-503.
71. R. B. Best, S. B. Fowler, J. L. Toca-Herrera and J. Clarke, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 12143-12148.
72. M. S. Li, D. K. Klimov and D. Thirumalai, *Polymer*, 2004, **45**, 573-579.
73. M. Tehei, B. Franzetti, D. Madern, M. Ginzburg, B. Z. Ginzburg, M. T. Giudici-Orticoni, M. Bruschi and G. Zaccai, *Embo Rep*, 2004, **5**, 66-70.
74. D. K. West, P. D. Olmsted and E. Paci, *J Chem Phys*, 2006, **124**.
75. S. Kumar, *Nat Mater*, 2014, **13**, 918-920.
76. O. Chaudhuri, S. T. Koshy, C. B. da Cunha, J. W. Shin, C. S. Verbeke, K. H. Allison and D. J. Mooney, *Nat Mater*, 2014, **13**, 970-978.
77. J. H. Wen, L. G. Vincent, A. Fuhrmann, Y. S. Choi, K. C. Hribar, H. Taylor-Weiner, S. C. Chen and A. J. Engler, *Nat Mater*, 2014, **13**, 979-987.
78. M. Zhang, *Nat Chem Biol*, 2007, **3**, 757-758.
79. B. Mandal, A. Grinberg, E. S. Gil, B. Panilaitis and D. Kaplan, *J Tissue Eng Regen M*, 2012, **6**, 181-181.