Electronic Supplementary Material (ESI) for Soft Matter.

Supporting Information

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

Katarzyna M. Tych[a,b], 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*]

[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.

MATERIALS AND METHODS

Protein Engineering and Expression

The sequences for each polyprotein construct are shown below. While each polyprotein contains distinct CSP domains, other features such as the $(His)_6$ tag (shown in purple in the sequence below), inter-domain linker sequences (shown in black), fingerprint I27 domains (shown in yellow) and cysteine residues at the C-terminus are identical (shown in black). For the $(I27\text{-}CTM)_3$-I27 construct the mutations in the CTM are shown in red.

$(I27\text{-}BsCSP)_3$-I27

MHHHHHHSSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELLSVGATIMLEGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTK
EATVIGLASLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELALSGTIVMLEGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTK
EAVITGSLALIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELSALGIVTMLEGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTK
EAITAGVSLLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELCC*

$(I27\text{-}TmCSP)_3$-I27

MHHHHHHSSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELLSVGATIMRGKVKWFDSKKGYGFITKDEGGDVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVV
ETVIGLASLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAAN
AKSAANLKVKELALSGTIVMRGKVKWFDSKKGYGFITKDEGGDVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
VITGSLALIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAANA
KSAANLKVKELSALGIVTMRGKVKWFDSKKGYGFITKDEGGDVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVEI
TAGVSLLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAANAK
SAANLKVKELCC*

$(I27\text{-}CTM)_3$-I27

MHHHHHHSSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELLSVGATIMLRGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKK
EATVIGLASLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELALSGTIVMLRGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKK
EAVITGSLALIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELSALGIVTMLRGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKK
EAITAGVSLLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGEVSFQAA
NAKSAANLKVKELCC*

RESULTS

Comparing interactions in BsCSP, TmCSP and CTM using molecular dynamics simulations

The results from molecular dynamics (MD) simulations are described below. The CHARMM param19 united-atom force field is used with the FACTS implicit solvation model.[1] The model uses a surface tension-like parameter of 0.015 kcal mol^{-1} Å^{-2}. This value is recommended for the investigation of structured peptides.[1] With such a model the root mean square deviation (RMSD) values for backbone atoms from the PDB structures plateau at ~0.37–0.41 nm.

n=5 simulations	*Bs*CSP	*Tm*CSP	CTM
Beta content (%)	49 ± 4	48 ± 4	50 ± 4
Helical content (%)	6 ± 1	2 ± 3	1 ± 1
Radius of gyration (nm)	1.099 ± 0.006	1.135 ± 0.007	1.106 ± 0.004

Table S1: Secondary structure and radius of gyration of the three cold shock proteins. Data are presented as the mean ± standard deviation from values averaged over five independent 200 ns simulations. β-content includes E and B DSSPcont assignments, helix content includes G, H and I assignments.

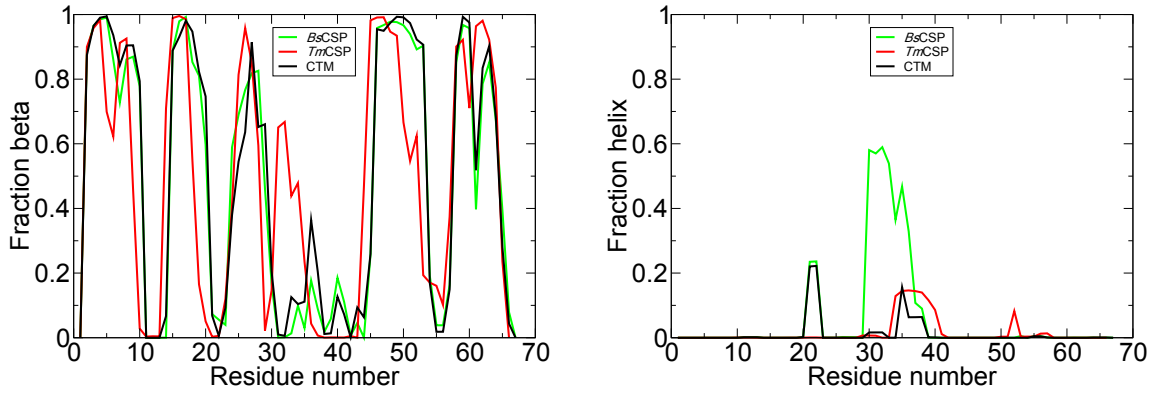

Figure S1: Comparison of the average secondary structure for each of the three cold-shock proteins as a function of position within the sequence. All show similar secondary structures, dominated by 5 β-strands (left). The model shows increased β-character in the long-loop region (#30-35) for *Tm*CSP. Some helical content is also observed in the long loop region for *Bs*CSP (bottom, left). A small helix is observed just after β-strand 3 in the PDB crystal structure of *Bs*CSP and several other cold shock proteins (*e.g.* see PDB 1C9O, 1MJC).

From the β-content plots in Fig. S1 and using sequence alignment between proteins, regions of the sequence were separated into β-strand and loop regions for subsequent hydrogen bonding analysis as follows:

For *Tm*CSP
β-strand 1: 1–10 (red). β-strand 2: 13–20 (blue). β-strand 3: 22–30 (red). Loop: 31–42 (black). β-strand 4: 43–54 (blue). β-strand 5: 55–66 (red).
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE

For *Bs*CSP and the CTM
β-strand 1: 1–11 (red). β-strand 2: 14–21 (blue). β-strand 3: 23–31 (red). Loop: 32–43 (black). β-strand 4: 44–55 (blue). β-strand 5: 56–67 (red).
*Bs*CSP
MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA

CTM

MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

This was in reasonable agreement with the original PDB-based assignments.[2,3]

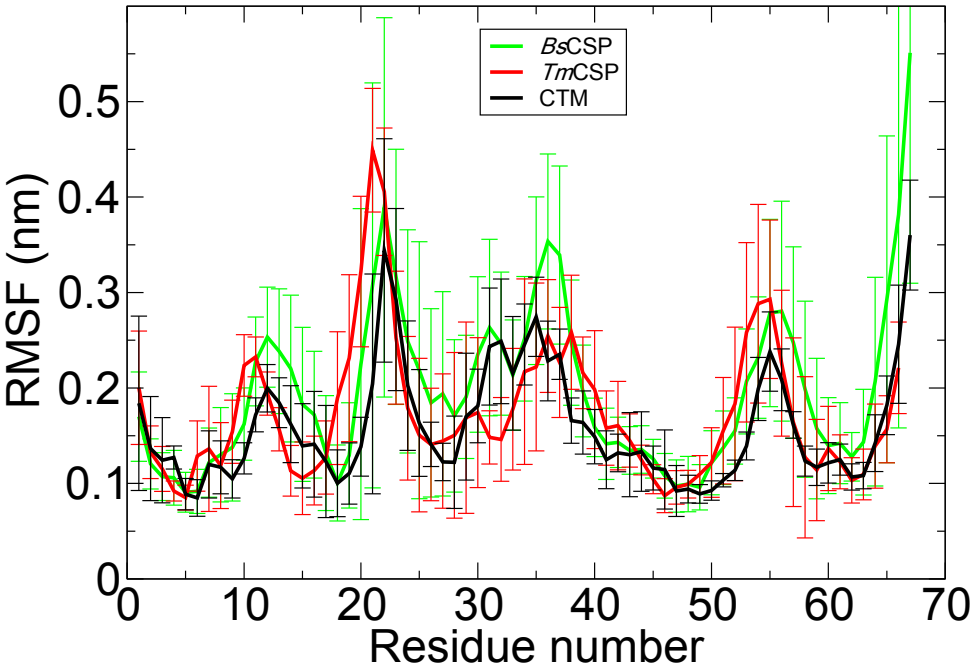

Figure S2: Comparison of the root mean square fluctuations (RMSF) of Cα atoms from each residue in the three cold-shock proteins. Mean values from 5 independent simulations are shown and the error bars represent the standard deviation between runs. All three proteins show similar regions of the sequence with increased flexibility, these relate to the loop regions between β-strands. There is slightly reduced flexibility on average in the long loop region (around residue 36) for *Tm*CSP and the CTM compared to *Bs*CSP. This feature is similar to that observed by Kalimeri *et al.*[4] on comparison of a different pair of mesophilic and thermophilic pair of proteins.

Comparison of hydrophobic core of the proteins using molecular dynamics

We used the program 'NACCESS' to ascertain which residues make up the hydrophobic core of the protein (see http://www.bioinf.manchester.ac.uk/naccess/ Hubbard, S. J. and Thornton, J. M. (1993), 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London, which is based on previous work by Lee *et al.*[5]). NACCESS assesses the solvent accessibility of atoms within each residue. Core residues are here defined as those with values of relative side-chain surface accessibility less than 10%. Cα were included as part of the side chain. Glycine residues were excluded from analysis. Results are shown for core residues: (i) determined using just the PDB structure of the protein, and (ii) through analysing 2000 snapshots from each simulation and averaging. The results are shown as a sequence alignment with residues highlighted in purple that have accessibilities of 0–1% and those residues in pink are those with accessibilities of 1–10%.

The hydrophobic core of the cold shock proteins determined using just the PDB structure of the protein:

*Bs*CSP

MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA

*Tm*CSP
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
CTM
MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

The hydrophobic core of the cold shock proteins determined through analysing 2000 snapshots from each simulation and averaging:

*Bs*CSP
MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA
*Tm*CSP
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
CTM
MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

The hydrophobic cores of the three proteins are closely related (virtually identical for *Bs*CSP and the CTM). Many of the core residues indicated in the PDB are loosened over the course of the simulation. An interesting outlier from the analysis is the tryptophan residue (W29), which is exposed in the PDB of *Tm*CSP but to some degree buried during the simulations. Interestingly this W residue is only present in *Tm*CSP (typically F in most other cold shock proteins).

Comparison of cold shock protein behaviour using a more conservative implicit solvent MD model

Altering model parameters allows us to gauge the robustness of the overall conclusions from simulation. Using the same CHARMM param19 united-atom force field with the FACTS implicit solvation model,[1] we performed a second set of simulations using a higher surface tension-like parameter of 0.025 kcal mol^{-1} $Å^{-2}$ (all other model parameters were maintained). This value is recommended for modelling larger, globular systems[1] and along with 0.015 kcal mol^{-1} $Å^{-2}$ was the value originally used in accessing the FACTS solvation model. This model gave backbone atom RMSD values that plateaued at 0.27–0.32 nm and is therefore referred to as the conservative model. This following describes the results from analyses equivalent to those performed for the original model.

n=5 simulations	*Bs*CSP	*Tm*CSP	CTM
Beta content (%)	52 ± 2	49 ± 2	49 ± 3
Helical content (%)	1 ± 0	1 ± 1	0 ± 0
Radius of gyration (nm)	1.068 ± 0.005	1.085 ± 0.007	1.068 ± 0.004

Table S2: Secondary structure and radius of gyration comparisons for the three cold shock proteins using the conservative model. Data are presented as the mean ± standard deviation from values averaged over 5 independent 200 ns simulations.

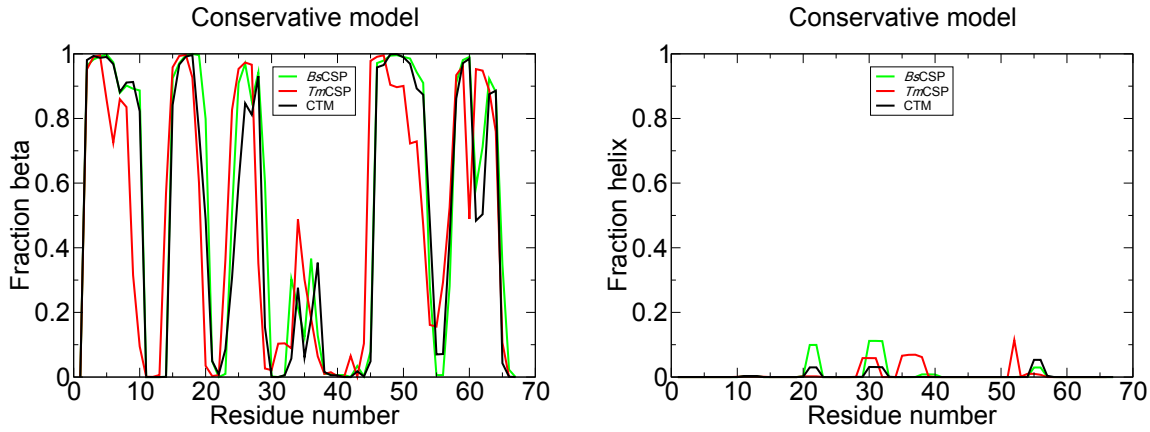

Figure S3: Comparison of the average secondary structure for each of the three cold-shock proteins as a function of position within the sequence, using the conservative model. All show similar secondary structures, dominated by 5 β strands (left). The conservative model shows no significant helical content (right) and greater similarity between proteins.

As a reminder, the hydrophobic core of the cold shock proteins determined using just the PDB structure is:

*Bs*CSP
MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA
*Tm*CSP
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
CTM
MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

The hydrophobic core of the cold shock proteins determined through analysing 2000 snapshots from each conservative model simulation and averaging (as above):

*Bs*CSP
MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA

*Tm*CSP
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
CTM
MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

The hydrophobic cores of all three proteins are again closely related (virtually identical for *Bs*CSP and the CTM). Many of the core residues indicated in the PDB are loosened over the course of the simulation. An interesting outlier from the analysis is the glutamine residue (Q59) in *Bs*CSP and the CTM, which is exposed in the PDB structures but is to some degree buried during the simulations.

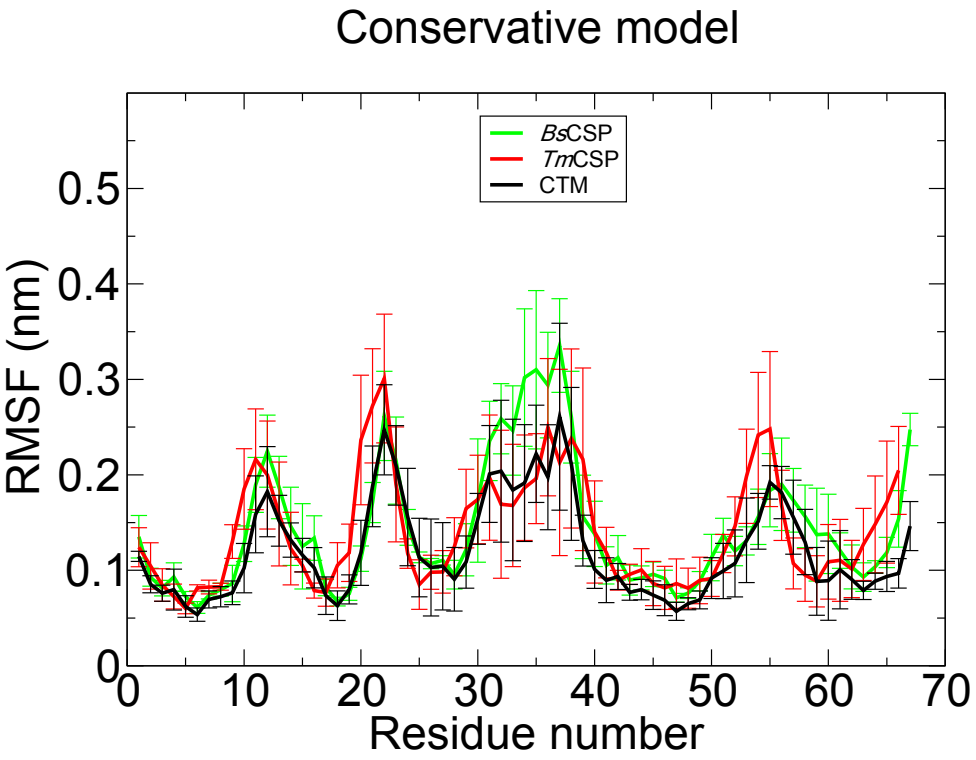

Figure S4: Comparison of the root mean square fluctuations (RMSF) of Cα atoms from each residue in the three cold-shock proteins, using the conservative model. Mean values from 5 independent simulations are shown and the error bars represent the standard deviation between runs. As with the original model, all three proteins show increased flexibility for loop regions between β-strands. There is slightly reduced flexibility in the long loop region for *Tm*CSP and the CTM compared to *Bs*CSP.

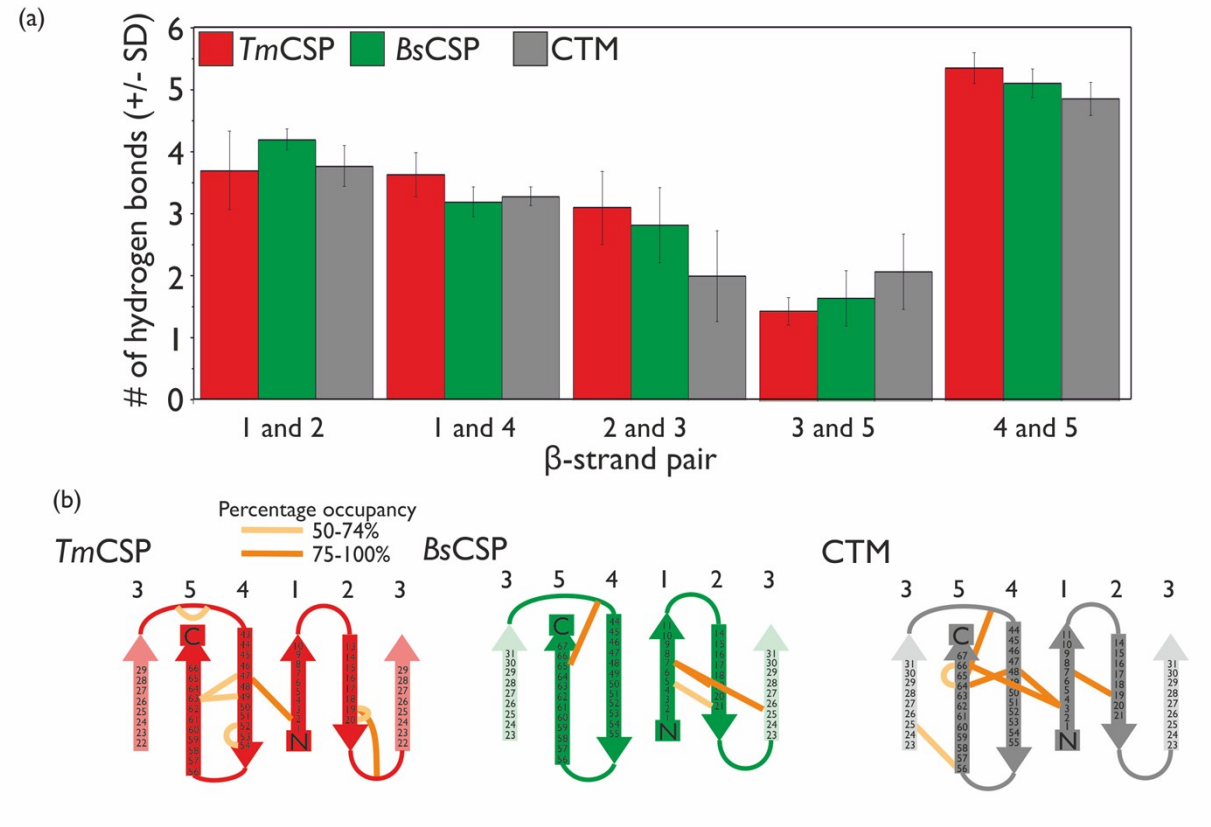

Figure S5: Analysis of MD simulations using the conservative model (defined above) of each protein. (a) The mean number of hydrogen bonds ± the standard deviation (SD, shown as error bars) over the course of five 200 ns simulations between each pair of β-strands in *Tm*CSP (red bars), *Bs*CSP (green 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. On average, the total numbers of salt bridges were: *Tm*CSP 8.0 ± 0.7, *Bs*CSP 5.0 ± 1.0, and CTM 9.1 ± 0.7. Overall, more salt bridges are observed when using the conservative model compared to the original (loose) model. Fewer salt bridges are observed in *Bs*CSP than *Tm*CSP and none are significantly occupied between β-strands 1 and 4 in *Bs*CSP. In CTM, the desired additional salt bridges between 1-4 and 4-5 are found as expected but some off-target salt bridges are also occupied *e.g.* a direct 1-5 salt bridge (R3-E66). There is also a small increase in the average numbers of inter-strand hydrogen bonds using the conservative model in comparison to the original – however the pattern and the similarity (within error) between the three CSPs remain.

Comparison of cold shock protein behaviour using an explicitly solvated MD model

We performed additional simulations using an explicitly solvated protein model. The CHARMM param36 force field was used. Starting structures were generated using VMD.[6] Proteins were solvated in a water box containing a 1.2 nm surround of water molecules (~4400) and NaCl ions were added at concentration of ~50 mM. NAMD was used to run the simulations.[7] A short heating protocol followed by 0.2 ns of equilibration preceded a single 200 ns simulation for each protein. Inclusion of an additional 100 ns of simulation (for *Tm*CSP) gave no change to any of the reported parameters, indicating that the modelling is well converged at least on this timescale. This model gave backbone atom RMSD values that plateaued at 0.25 nm (*Tm*CSP) and 0.15 nm (*Bs*CSP/CTM). This following describes the results from analyses equivalent to those performed for the original model.

n=1 simulations	*Bs*CSP	*Tm*CSP	CTM
Beta content (%)	53 ± 3	51 ± 3	53 ± 3
Helical content (%)	4 ± 1	4 ± 1	4 ± 2
Radius of gyration (nm)	1.110 ± 0.010	1.130 ± 0.011	1.113 ± 0.010

Table S3: Secondary structure and radius of gyration comparisons for the three cold shock proteins using the explicitly solvated model. Data are presented as the mean ± standard deviation within each 200 ns simulation.

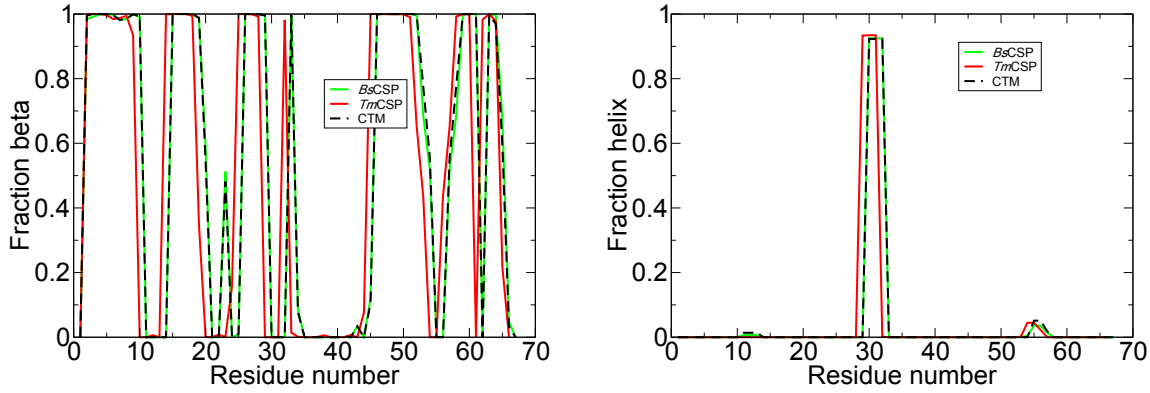

Figure S6: Comparison of the average secondary structure for each of the three cold-shock proteins as a function of position within the sequence, using the explicitly solvated model. All show near identical secondary structure, dominated by 5 β strands (left). The explicit solvent model shows significant and conserved helical content (right) just after β3 in the three proteins.

Again, as a reminder, the hydrophobic core of the cold shock proteins determined using just PDB structures:

*Bs*CSP
MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA
*Tm*CSP
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
CTM
MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

The hydrophobic core of the cold shock proteins determined through analysing 2000 snapshots from each explicitly solvated model simulation:

*Bs*CSP
MLEGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKEA
*Tm*CSP
M-RGKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
CTM
MLRGKVKWFNSEKGFGFI-EVEGQDDVFVHFSAIQGEGFKTLEEGQAVEFEIVEGNRGPQAANVKKEA

The core residues indicated in the PDB are largely maintained over the course of the simulation. The hydrophobic cores of all three proteins are very closely related (virtually identical). Two residues are more strongly shielded from solvent in *Tm*CSP than the equivalent residues in *Bs*CSP or CTM: I32 is 0.7% exposed in *Tm*CSP compared to ~5% exposure of I33 in *Bs*CSP and CTM; V46 is 0.04% exposed in *Tm*CSP compared to ~1% exposure of V47 in *Bs*CSP and CTM.

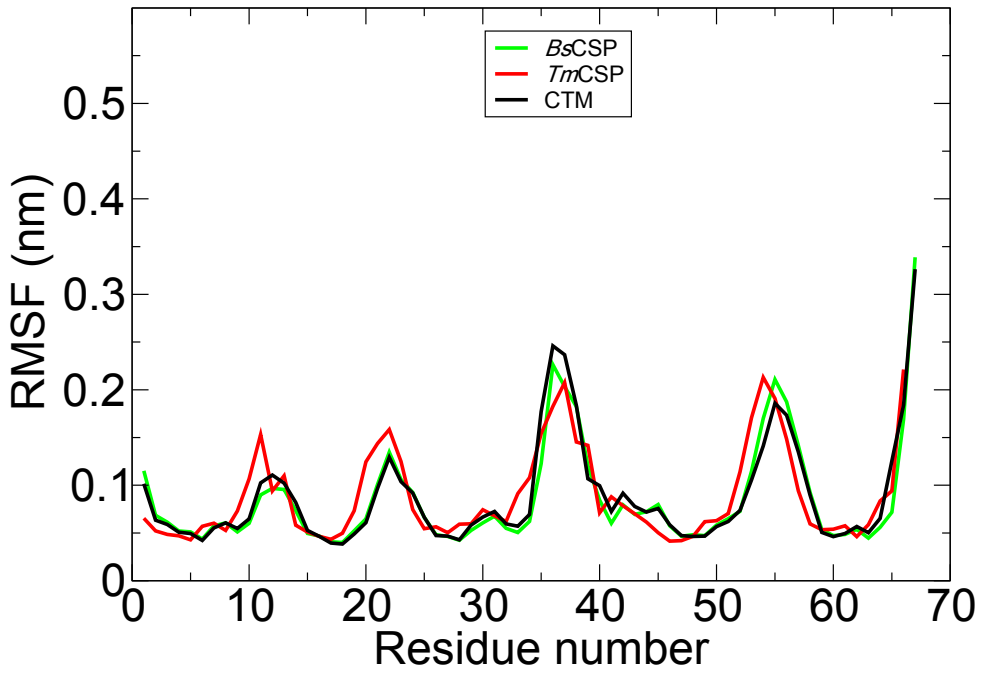

Figure S7: Comparison of the root mean square fluctuations (RMSF) of Cα atoms from each residue in the three cold-shock proteins, using the explicitly solvated model. As with the original model, all three proteins show increased flexibility for loop regions between β-strands. The flexibility is similar for *Bs*CSP, *Tm*CSP and the CTM.

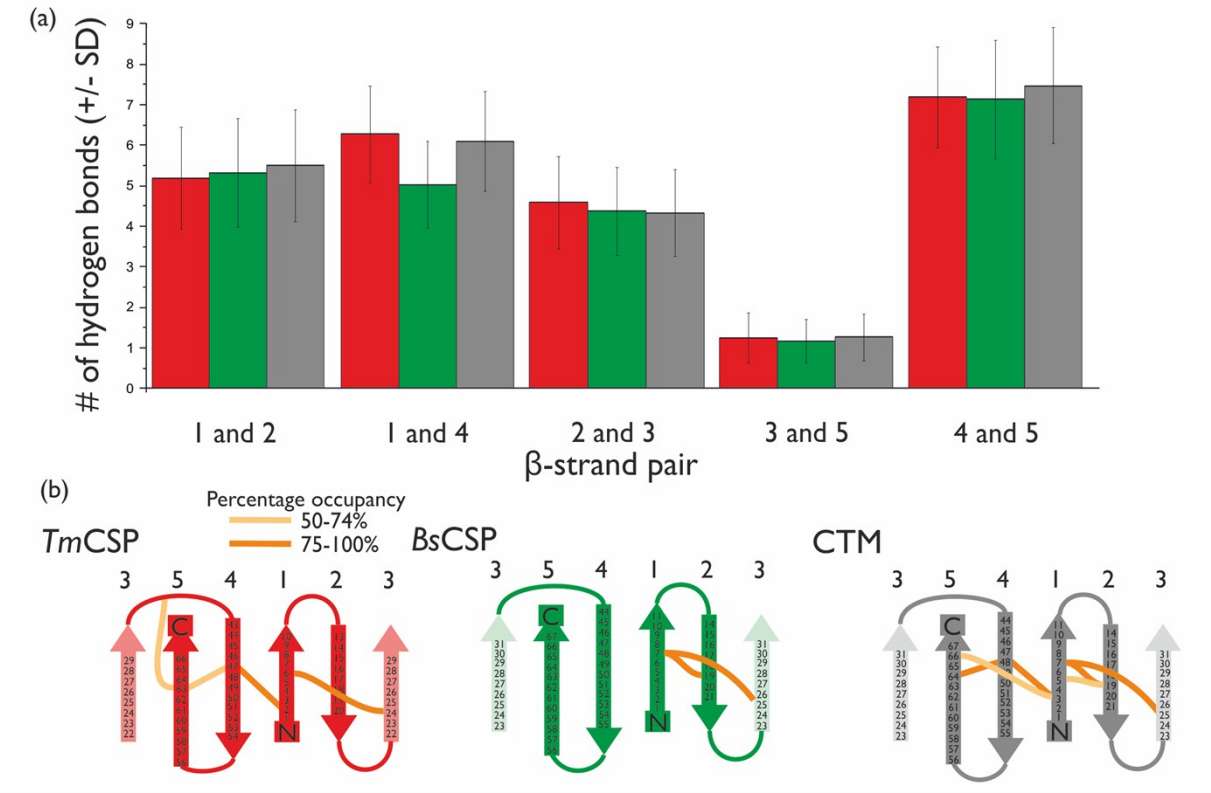

Figure S8: Analysis of MD simulations using the explicitly solvated model of each protein. (a) The mean number of hydrogen bonds ± the standard deviation (SD, shown as error bars) within the 200 ns simulations between each pair of β-strands in *Tm*CSP (red bars), *Bs*CSP (green bars) and CTM (grey bars). (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 the 200 ns simulations. On average ± SD, the total numbers of salt bridges were: *Tm*CSP 4.3 ± 1.1, *Bs*CSP 4.3 ± 1.2, and CTM 6.4 ± 1.4.

The salt bridges observed are in line with those found with implicit solvation. However, unlike the observation in implicit solvent models, the overall numbers of salt bridges observed in *Bs*CSP are similar to *Tm*CSP. The important feature is the location and % occupancy of individual salt bridges – no salt bridges are significantly occupied that bridge between strands 1-4 and 4-5 in *Bs*CSP. In CTM, the desired additional salt bridges between 1-4 and 4-5 are found as expected, along with a concomitant rise in total average numbers of salt bridges. As observed using the conservative model, off-target salt bridges are also occupied in CTM, for example, the direct 1-5 salt bridge (R3-E66). There is a general increase in the average numbers of inter-strand hydrogen bonds using the explicitly solvated model in comparison with implicit solvent models, though the pattern and the similarity (within error) between the three CSPs remains. The additional hydrogen bond observed on average between strands 1-4 for *Tm*CSP/CTM compared to *Bs*CSP is specifically due the side-chain interaction between R2 and E47 (*Tm*CSP) or R3 and E48 (CTM). These residue side-chains are intimately associated (maintain a shirt distance between) such that they count not only as highly occupied salt bridges but also as hydrogen bonds using our definitions.

Mechanical unfolding pathway of the cold shock proteins determined using molecular dynamics

Due to inherent computational limitations, these unfolding simulations occur on a timescale many orders of magnitude faster than achieved experimentally. Consequently, the accuracy of these simulations reduces as unfolding proceeds further away from the native state, since the structure does not have time in which to relax/re-orient to lower energy conformers. Such effects are minimized for proteins that unfold in a single step, all-or-none manner resulting in excellent agreement between simulated and experimental approaches for mechanically brittle proteins such as I27.[6] To validate the atomistically-detailed unfolding pathway of CSPs provided by MD simulations we first considered the extent of unfolding prior to irreversible unfolding as measured experimentally. The similarity in the measured difference in contour length (ΔL_C) from the SMFS data for CSPs suggests that the mechanical clamp is broadly in the same position for each protein domain. This is not surprising given that the secondary structure arrangement and hydrogen bonding between β-strands is the same (Fig. 1(a) and 3(a)). If the unfolding of CSPs occurs in an all-or-none manner then the measured ΔL_C should be similar to that predicted based on the native conformation of the protein and its size (in terms of the number of amino acids). For *Tm*CSP/*Bs*CSP ΔL_C can be estimated to be 23.5/23.9 nm as an unfolded 66/67 residue polypeptide has an L_C value of 25.1/25.5 nm (assuming each residue contributes 0.38 nm[7] and the initial distance between N- and C-termini in the CSP proteins is ~1.6 nm). This value is in excellent agreement with the 23.5 nm value measured for each CSP, which indicates that the transition states for mechanical unfolding are close to the native state; a result that is similar to that found for thermal unfolding.[8] This indicates that the initial events along the simulated unfolding pathway are of relevance, where modelling accuracy is most likely to be good. Based on this, the overall results from pulling simulations were two-fold. Firstly, the initial event occurs at much lower force for CSPs than for I27, in line with the expectation from AFM. However, at this high pulling speed and using 20 simulations we were unable to discriminate between CSPs based on the magnitude in force of the initial peak. Secondly, whilst I27 unfolding simulations previously performed with the same model[9] all occur in a similar manner – with the A'–G strand disruption being the most significant event – different unfolding pathways were observed for the CSPs. For all CSPs, rupture of β-strands 1-4 or β-strands 4-5 or near simultaneous rupture of both underlie the initial peak in force and subsequent lengthening of the protein. Notably the proportions of simulations that followed each pathway differed, with *Tm*CSP more heavily favouring initial breakage of β-strands 1-4. A very similar pattern of results was obtained from MD simulations of unfolding using the conservative implicit solvent model.

We compared RMSD values from equilibrium simulations of the *Bs*CSP and CTM proteins for the backbone atoms that we approximated to be in the mechanical clamp region of the protein, between β-strands 1-4 (residues 2–6/44–49) and strands 4-5 (residues 48–54/57–64). The RMSD was the same for the β-strands 1-4, with 0.9 Å for CTM and 0.8 Å for *Bs*CSP. For β-strands 4-5 CTM had a slightly reduced RMSD (1.1 Å) compared with that of *Bs*CSP (1.5 Å). Using the conservative implicit solvent model: β1-β4 clamp region RMSD values were 0.7 Å for *Bs*CSP and 0.5 Å for CTM. β4-β5 clamp region RMSD values were 1.2 Å for *Bs*CSP and 1.1 Å for CTM. Using the explicit solvent model: β1-β4 clamp region RMSD values were 0.4 Å for both *Bs*CSP and CTM. β4-β5 clamp region RMSD values were 0.9 Å for *Bs*CSP and 1.0 Å for CTM.

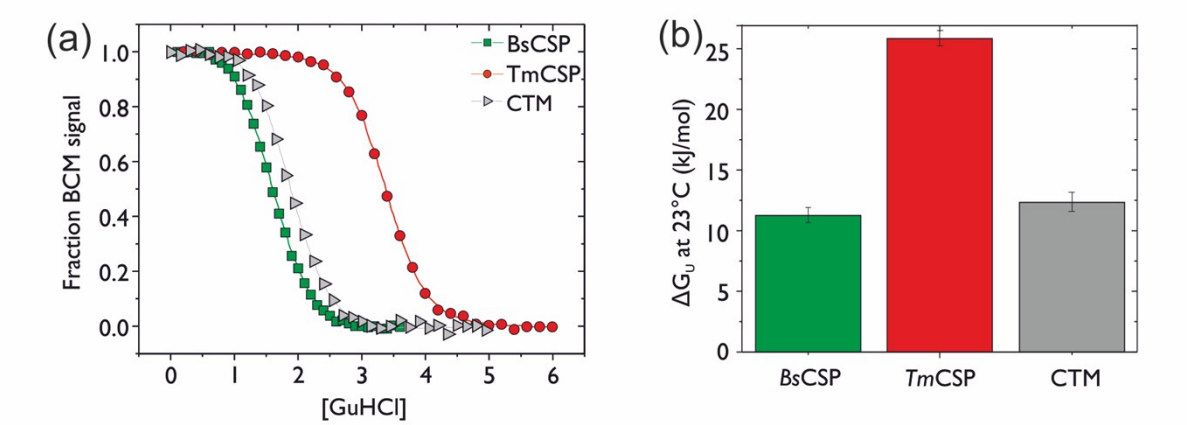

Figure S9. (a) Chemical denaturation experiments monitored by fluorescence emission spectroscopy measure the thermodynamic stability of *Bs*CSP (green), *Tm*CSP (red) and CTM (grey) at 23 °C. (b) The mesophilic protein *Bs*CSP (green) is thermodynamically less stable than the hyperthermophilic protein *Tm*CSP (red) at 23 °C. The thermodynamic stability of CTM protein at 23 °C (grey) lies between that of *Bs*CSP and *Tm*CSP.

Protein	ΔG_U (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	[D]½ (M)
*Bs*CSP	11.27 (± 0.65)	7.56 (± 0.34)	1.49
*Tm*CSP	25.86 (± 0.63)	7.84 (± 0.19)	3.28
CTM	12.34 (± 0.78)	7.69 (± 0.36)	1.60

Table S4. Summary of free energy parameters obtained from chemical denaturation equilibrium experiments at room temperature for *Bs*CSP and *Tm*CSP, where ΔG_U is the Gibbs free energy of unfolding, m is the m-value and [D]½ is the Guanidine hydrochloride concentration at which 50% of the protein is folded.

Mechanical unfolding experiments using Single Molecule Force Spectroscopy

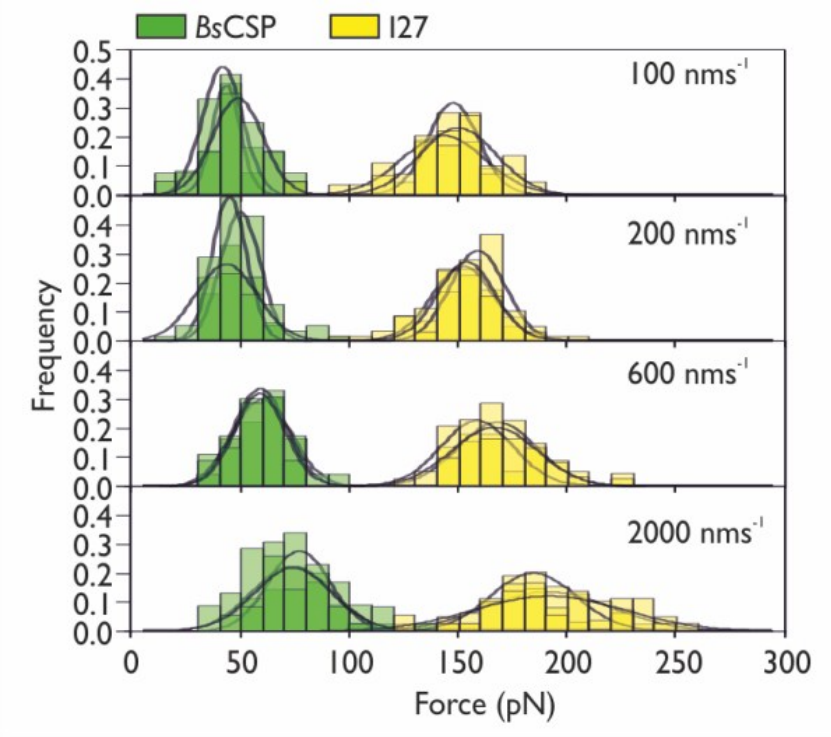

Figure S10: Unfolding force histograms for experiments conducted in triplicate at pulling speeds of 100, 200, 600 and 2000 nm s^{-1} for the (I27-*Bs*CSP)$_3$-I27 polyprotein. *Bs*CSP events are shown in green and those for I27 in yellow, grouped by pulling velocity. The histograms show a clear separation in the distributions of the forces resulting from the mechanical unfolding of I27 and the *Bs*CSP. Gaussian fits to histograms for each data set are used to obtain a measure of the unfolding forces.

Speed [nms^{-1}]	# *Bs*CSP	# I27	Median unfolding force *Bs*CSP [pN] (±SD)	Average [pN] (±SD)	Median unfolding force I27 [pN] (±SD)	Average [pN] (±SD)
100	13	21	42 (± 14)		145 (± 20)	
	12	27	41 (± 9)	43 (± 1)	143 (± 19)	146 (± 2)
	20	29	46 (± 12)		150 (± 16)	
200	30	56	50 (± 8)		151 (± 18)	
	18	35	44 (± 8)	46 (± 3)	158 (± 13)	154 (± 4)
	55	57	45 (±17)		154 (± 18)	
600	18	22	59 (± 13)		170 (± 18)	
	23	38	59 (± 11)	59 (± 1)	162 (± 20)	166 (± 4)
	45	60	60 (± 14)		165 (± 14)	
2000	54	77	76 (± 16)		186 (± 20)	
	35	38	78 (± 15)	77 (± 1)	188 (± 28)	187 (± 1)
	35	36	78 (± 20)		187 (± 1)	

Table S5. Summary of mechanical unfolding data for (I27-*Bs*CSP)$_3$-I27

Speed [nms^{-1}]	# CTM	# I27	Median unfolding force CTM [pN] (±SD)	Average [pN] (±SD)	Median unfolding force I27 [pN] (±SD)	Average [pN] (±SD)
100	13	26	45 (± 15)		141 (± 19)	
	12	28	36 (± 8)	39 (± 4)	142 (± 18)	144 (± 4)
	33	58	36 (± 18)		150 (± 18)	
200	14	29	45 (±10)		155 (± 19)	
	16	44	35 (± 20)	39 (± 5)	158 (± 19)	156 (± 2)
	18	30	37 (±12)		155 (± 17)	
600	23	33	54 (± 21)		175 (± 24)	
	50	74	53 (± 18)	51 (± 4)	179 (± 18)	176 (± 2)
	23	44	45 (± 17)		174 (± 28)	
2000	35	50	68 (± 16)		188 (± 30)	
	42	32	65 (± 18)	64 (±3)	191 (± 33)	187 (± 4)
	27	42	60 (± 17)		181 (± 24)	

Table S6. Summary of mechanical unfolding data for (I27-CTM)$_3$-I27

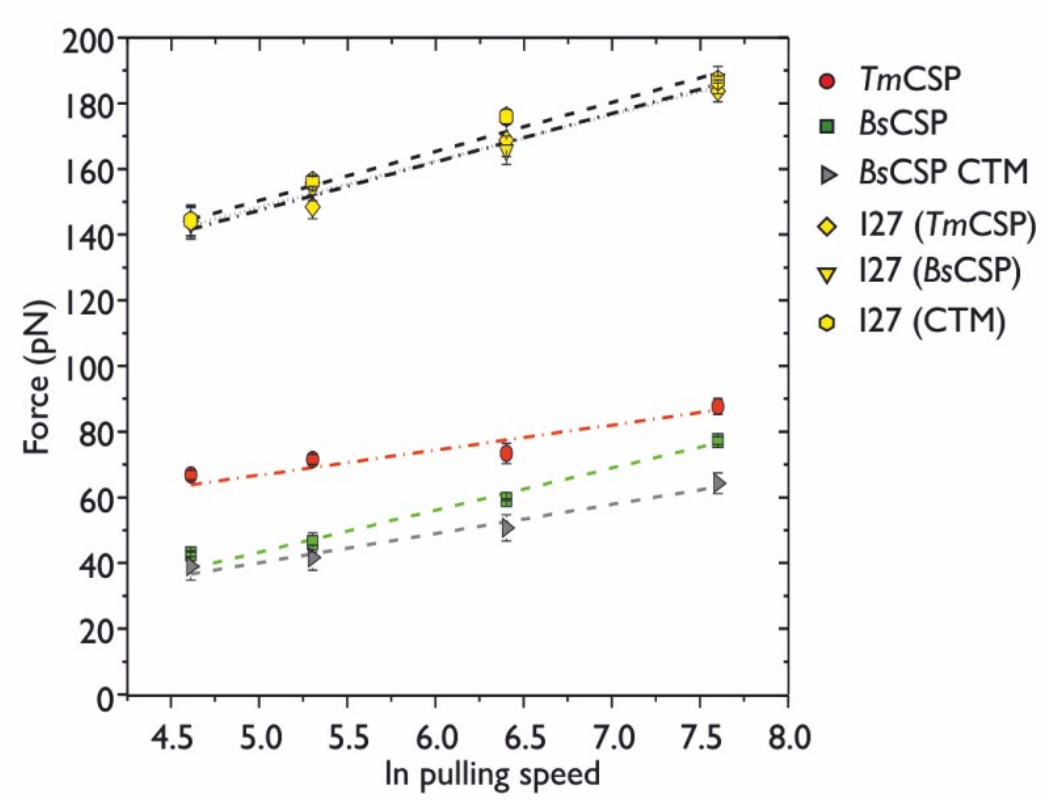

Figure S11: Pulling speed dependence of the unfolding force for all three polyprotein constructs, showing the averages of the median unfolding forces from three separate experiments at each pulling speed for each protein domain. The dashed lines show the Monte Carlo simulation fits to the data (parameters given in Table 1 in the main paper).

Protein	Δx_U (nm)	k_U (s^{-1})	ΔG^*_U (kJ/mol)
*Bs*CSP	0.45 (± 0.01)	0.2133 (± 0.0219)	38
*Tm*CSP	0.70 (± 0.08)	0.0026 (± 0.0018)	49
CTM	0.62 (± 0.07)	0.1163 (± 0.0654)	40
I27 (*Bs*CSP)	0.32 (± 0.01)	0.0006 (± 0.0002)	53
I27 (*Tm*CSP)	0.32 (± 0.04)	0.0015 (± 0.0012)	50
I27 (CTM)	0.28 (± 0.03)	0.0030 (± 0.0020)	49

Table S7. Summary of mechanical free energy parameters for *Tm*CSP, *Bs*CSP, CTM and I27

REFERENCES

(1) Haberthur, U.; Caflisch, A. *Journal of Computational Chemistry* **2008**, *29*, 701.

(2) Kremer, W.; Schuler, B.; Harrieder, S.; Geyer, M.; Gronwald, W.; Welker, C.; Jaenicke, R.; Kalbitzer, H. R. *European Journal of Biochemistry* **2001**, *268*, 2527.

(3) Schindelin, H.; Marahiel, M. A.; Heinemann, U. *Nature* **1993**, *364*, 164.

(4) Kalimeri, M.; Rahaman, O.; Melchionna, S.; Sterpone, F. *J Phys Chem B* **2013**, *117*, 13775.

(5) Lee, B.; Richards, F. M. *Journal of Molecular Biology* **1971**, *55*, 379.

(6) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph Model* **1996**, *14*, 33.

(7) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781.

(8) Lu, H.; Schulten, K. *Biophysical Journal* **2000**, *79*, 51.

(9) Carrion-Vazquez, M.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 11288.

(10) Motono, C.; Gromiha, M. M.; Kumar, S. *Proteins* **2008**, *71*, 655.

(11) Wolny, M.; Batchelor, M.; Knight, P. J.; Paci, E.; Dougan, L.; Peckham, M. *Journal of Biological Chemistry* **2014**, *289*, 27825.