



UNIVERSITY OF LEEDS

Comparing the thermal, chemical and mechanical stabilities of extremophilic cold shock protein variants

Michael Christopher Wilson

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology

Faculty of Biological Sciences

September 2016

The candidate confirms that the submitted work is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated overleaf. The candidate confirms the appropriate credit has been given within the thesis where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgement.

Acknowledgements

I would like to thank Dr Lorna Dougan and Dr David Brockwell for their amazing continued support and guidance throughout this project. I would like to thank Dr Toni Hoffman for taking the time and patience in demonstrating to me many of the protein expression techniques used in the project. My thanks go to Dr Kasia Tych who developed the variable temperature AFM and who always had the time to help with any AFM difficulties experienced. I am grateful to Dr James Ault and Dr Lars Kuhn for their help in conducting the MS and NMR experiments respectively.

All the members of the Brockwell/Radford and Dougan groups have been so welcoming and created a great community in which to work. I would like to specially thank Megan Hughes, Danielle Walsh, Ellen Kendrick, Julia Humes, Bob Schriffin and Rhys Thomas for their input and suggestions. I would like to thank Nasir Khan for keeping the lab stocked with relevant chemicals a good supply of biscuits as well as providing the relevant safety training. Lastly I would like to thank my family for their incredible support over the last 4 years.

Jointly authored publications

Section 3.5 contains work from the following jointly authored papers, below are the details of the publications and my respective contribution. Work not performed by myself is clearly detailed in the relevant sections.

- Tych K. M., Batchelor M., Hoffmann T., Wilson M. C., Paci E., Brockwell D. J. and Dougan L. (2016) Tuning protein mechanics through an ionic cluster graft from an extremophilic protein *Soft Matter*, **12**: 2688-2699

All of the data described within the section 'protein thermodynamic stability' was performed by me including data in Figure 2 a and b as well as chemical equilibrium curve values quoted in the section 'Thermodynamic and mechanical stability of CTM'. AFM measurements were performed by Tych K. M. and Hoffmann T. while MD simulations were conducted by Batchelor M.

- Tych K. M., Batchelor M., Hoffmann T., Wilson M. C., Hughes M. L., Paci E., Brockwell D. J., Dougan L. (2016) Differential Effects of Hydrophobic Core Packing Residues for Thermodynamic and Mechanical Stability of a Hyperthermophilic Protein. *Langmuir*. 32 (29), 7392–7402

All data within sections 'Thermodynamic Stability Curves of *TmCSP*, *TmCSP* L40A, and *TmCSP* V62A' was performed by me including data used in figure 3. The AFM experiments in this paper were performed by Tych K. M. and Hoffman T. MD simulations were performed by Batchelor M. Monte Carlo simulations were performed by Tych K. M. and Hughes M. L.

Abstract

Proteins conduct a vast array of chemical processes and achieve this through a balance of flexibility and function. Proteins from extremophilic organisms have evolved to adjust this balance to function in extreme conditions. Research into hot-adapted proteins has been encouraged through successes in improving enzyme thermostability, though cold-adapted proteins also offer substantial potential as they can function effectively at lower temperatures. Research into cold-adapted proteins remains limited despite cold climates comprising over 80% of Earth's biosphere.

This study compares the stabilities of cold shock protein (Csp) homologues from bacteria growing at vastly different temperatures. The cold-adapted Csps show reduced thermal denaturation mid-points (T_m values) compared to a temperate Csp with PB6 Csp displaying the lowest T_m of 42.8 °C contrasting with 52.3 °C shown by *BsCsp*. Thermostability projections using the Gibbs-Helmholtz equation show cold-adapted Csps display slightly reduced thermostabilities and remain folded over a narrower range of temperatures. Csp thermostabilities were found to be very similar at around 8 degrees below the optimal growth temperature of the bacteria each Csp was derived from, suggesting thermostabilities evolved relating to Csp operating temperatures. The roles of electrostatics and hydrophobic packing in stabilizing a hyperthermophilic Csp are evaluated using mutants that highlight how a single side chain length reduction dramatically decreases Csp thermostability. Kinetic studies showed that the Csps exhibit similar folding rate constants but different unfolding rate constants. Initial flexibility studies suggest that DNA binding increases Csp rigidity and triggers a conformational change in a psychrotrophic Csp.

Csp mechanical stabilities showed the same hierarchy as thermostabilities and similar sensitivity to temperature changes. Energy landscape projections constructed using Monte Carlo simulations showed the mechanical energy barrier height of Csp unfolding was temperature independent. At 5 °C the Csps showed

similar mechanical softness however the hyperthermophilic *TmCsp* shows greater mechanical softness at higher temperatures.

Ethical considerations

To select for those bacteria that have taken up specific plasmid DNA, antibiotic selection markers are used throughout this project. This means that the project generates bacteria with resistance to antibiotics so great care must be taken with their disposal in order to ensure that no contamination of the environment can take place. All bacterial liquid cultures were treated with the detergent Virkon in a 1% solution and soaked for at least 2 hours to ensure thorough disinfection. Laboratory coats and nitrile gloves were worn at all times when working with these bacterial cultures. These were always kept within the lab and the gloves sent for incineration after use. Plastic waste such as pipette tips, petri dishes and tubes are sent for autoclaving and incineration.

Contents

List of Abbreviations	12
1. Introduction	16
1.1. Protein stability	16
1.1.1 The protein stability-function trade-off	17
1.2 Types of extremophiles	19
1.2.1 Using extremophilic proteins to address the stability-function trade off	21
1.3 Methods of comparing protein stabilities	22
1.3.1 Mutational studies	22
1.3.2 Comparative studies	25
1.3.3 Comparative studies using extremophilic protein structures	27
1.3.4 Thermophilic protein adaptations	28
1.4 Psychrophilic proteins	29
1.4.1 Life at cold temperatures	29
1.4.2 Challenges of life at low temperatures	31
1.4.3 Adaptations of psychrophilic enzymes to cold conditions	35
1.4.4 The hydrophobic effect and cold-denaturation	37
1.4.5 Protein cold-adaption is an array of different approaches	39
1.4.6 Protein folding in cold-adapted proteins	40
1.5 Applications of cold-adapted proteins	41
1.6 Cold shock proteins	43
1.6.1 Cold shock protein structure and function	43
1.6.2 The role of <i>E. coli</i> CspA mRNA as a thermosensor	45
1.6.3 The role of cold shock proteins in regulating protein synthesis	46
1.6.4 Eukaryotic Csp homologues	49
1.7 Objectives	50
2. Materials and Methods	52
2.1 Materials	52
2.1.1 Cell culture medium and <i>E. coli</i> strains	53
2.2 Molecular biology methods	54
2.2.1 Polymerase chain reaction	54
2.2.2 Agarose gel electrophoresis	55
2.2.3 Restriction endonuclease digestion	56
2.2.4 Preparing LB agar plates	56

2.2.5 Preparing competent cells.....	57
2.2.6 DNA ligation into pET vectors.....	57
2.2.7 Transformation of pET vectors into <i>E. coli</i> cells	57
2.2.8 Plasmid DNA extraction.....	58
2.2.9 Gibson assembly reaction	58
2.2.10 Colony PCR	59
2.3 Expression of proteins in <i>E. coli</i>	60
2.3.1 DNA transformation into expression cells.....	60
2.3.2 Small scale expression trials	60
2.3.3 SDS-PAGE.....	61
2.3.4 Small scale Nickel-NTA purification.....	63
2.3.5 Large scale expression of His-tagged vectors.....	63
2.4 Purification of His-tagged proteins.....	64
2.4.1 Ni-NTA affinity purification of His-tagged proteins.....	64
2.4.2 DNA removal by ion-exchange chromatography	65
2.4.3 Size-exclusion chromatography	65
2.4.4 Dialysis and freeze drying.....	66
2.5 Production of ¹⁵ N labelled proteins for NMR	66
2.5.1 Making minimal media.....	66
2.5.2 Cleavage of PB6 Csp from MBP-PB6 Csp fusion protein	68
2.6 Multiple sequence alignment.....	68
2.7 Thermal and chemical stability assays	68
2.7.1 Sodium phosphate buffer.....	68
2.7.2 Circular dichroism.....	69
2.7.3 Using BCM to follow changes in tryptophan fluorescence	70
2.7.4 Thermal transitions followed by tryptophan fluorescence.....	70
2.7.5 Specific heat capacity determination	71
2.7.6 Gibbs-Helmholtz equation to generate thermostability dependence	71
2.7.7 Chemical denaturation curves.....	72
2.8 Stopped-flow unfolding and refolding kinetics.....	72
2.9 ¹ H- ¹⁵ N HSQC NMR spectra.....	74
2.10 Atomic force microscopy.....	74
2.10.1 Preparation of gold surface for AFM.....	75
2.10.2 Calibration of the cantilever spring constant.....	75

2.10.3 Force extension experiments.....	77
2.10.4 Trace selection	77
2.10.5 Variable temperature force extension experiments	78
2.10.6 Monte Carlo simulations.....	79
3 Biophysical characterisation of the monomers	82
3.1 Introduction	82
3.1.1 The Gibbs-Helmholtz equation and protein stability curves	82
3.1.2 Adaptations of the thermostability curves of thermophilic proteins.....	84
3.1.3 Evolution of thermostability curve parameters.....	86
3.1.4 Previous thermostability curve studies.....	87
3.1.5 Aims.....	88
3.2 Results.....	89
3.2.1 Selecting and purifying the cold shock protein monomers	89
3.2.2 Multiple sequence alignments of Csp sequences.....	91
3.2.3 Generation of over-expression plasmids for <i>BsCsp</i> and 3 cold-adapted Csps	96
3.2.4 Optimising expression of the monomers.....	97
3.3 Purification of monomeric Csps.....	100
3.3.1 Ni-NTA affinity chromatography.....	100
3.3.2 Ion-exchange chromatography.....	101
3.3.3 Size exclusion chromatography	102
3.3.4 Confirmation of protein identity by MS.....	103
3.4. Biophysical characterisation of the monomers	103
3.4.1 CD spectra to confirm folding	103
3.4.2 Thermal denaturation followed by change in circular dichroism.....	105
3.4.3 DNA contamination issues and using the barycentric median	107
3.4.4 Thermal denaturation values obtained by circular dichroism and fluorescence	109
3.4.5 Determination of the change in specific heat capacity upon unfolding	112
3.4.6 Comparing thermostability parameters between the Csps.....	115
3.4.7 Thermostability temperature dependence curves	116
3.4.8 Which parameters are important in the thermostability differences between the Csps?	119
3.4.9 Comparing the thermostabilities of proteins at the growth temperatures of their respective organisms	120
3.4.10 Chemical denaturation using GdnHCl and urea.....	123
3.4.11 Insights into the properties of extreme hot and extreme cold-adapted Csps .	126

3.5	Cold shock protein variants	127
3.5.1	Studies of a charged triple mutant of <i>BsCsp</i>	127
3.5.2	<i>Thermotoga maritima</i> hydrophobic deletion mutants	131
3.6	Discussion	135
4	Characterising cold shock protein flexibility by NMR and kinetics	137
4.1	Introduction.....	137
4.1.1	Flexibility comparisons between temperature adapted homologues	138
4.1.2	Controversy over the theory that thermophiles and hyperthermophiles are less dynamic	139
4.1.3	Global and local protein flexibility adaptations	140
4.1.4	Combining flexibility and thermostability properties of extremophilic enzymes	142
4.1.5	Dynamics of cold-shock proteins.....	142
4.1.6	Studies of protein folding and unfolding kinetics	143
4.1.7	Previous folding and unfolding studies on Csps.....	144
4.2	Results	144
4.2.1	Folding and unfolding of the Csp homologues.....	144
4.3	Expression of ¹⁵ N labelled protein for NMR studies.....	149
4.3.1	Over-expression and purification of ¹⁵ N MBP-PB6 Csp	150
4.3.2	Cleavage of PB6 Csp from MBP-PB6 Csp and purification of PB6 Csp	152
4.4.1	HSQC spectra of the Csp monomers	154
4.4.2	Exploring variations in dynamics with temperature for the Csps using HSQC spectra	157
4.4.3	Changes in the HSQC spectra of the Csps upon binding of DNA.....	161
4.4.4	Temperature sensitivity of the DNA bound forms	166
4.5	Conclusions.....	169
5	Comparing the mechanical stabilities of cold shock proteins using atomic force microscopy	171
5.1	Introduction.....	171
5.1.1	Probing the mechanical stabilities of proteins using AFM	172
5.1.2	Using polyproteins in AFM studies.....	175
5.1.3	Differences between unfolding proteins by chemical denaturants and applied forces.....	183
5.1.4	Previous AFM studies on Csps.....	184
5.1.5	Aims.....	186

5.2 Results.....	187
5.2.1 Production of the <i>TmCsp</i> homopolymer.....	187
5.2.2 Expression of the <i>TmCsp</i> homopolymer.....	190
5.2.3 Using Gibson assembly to generate (I27-PB6 Csp) ₃ -I27 polyprotein.....	191
5.2.4 Expression and purification of the (I27-PB6 Csp) ₃ -I27 polyprotein.....	195
5.2.5 Confirmation of generation of the (I27-PB6 Csp) ₃ -I27 polyprotein.....	198
5.3 AFM mechanical unfolding.....	200
5.3.1 Identification of a mechanical signature for PB6 Csp.....	201
5.3.2 Filtering of AFM traces.....	203
5.3.3 Obtaining and analysing (I27-PB6 Csp) ₃ -I27 mechanical unfolding data at 14 °C.....	204
5.4 Comparing the mechanical unfolding of the different Csps at 14 °C.....	206
5.5 Mechanical unfolding forces of Csps at different temperatures.....	208
5.6 Determination of mechanical unfolding energy barriers.....	213
5.6.1 Monte Carlo simulations.....	213
5.6.2 Interpreting Δx_u and k_u values obtained from the Monte Carlo simulations.....	214
5.6.3 Obtaining unfolding energy landscapes.....	217
5.7 Discussion.....	220
Appendix.....	228
References.....	238

List of Figures

Figure 1.1: Temperature dependence of $\Delta\Delta G_c$ and $T\Delta S$ with respect to temperature.	17
Figure 1.2: A comparison showing the similarity of adenylate kinase structures.	28
Figure 1.3: The changes in enzyme activity at different temperatures between a mesophilic α -amylase represented psychrophilic α -amylase from <i>Pseudoalteromonas haloplanktis</i>	32
Figure 1.4: Comparison between thermostability and activity of 3 alpha amylases isolated from organisms with live at vastly different temperatures.	33
Figure 1.5: The differences in protein thermostabilities of temperature-adapted α -amylases obtained from microcalorimetry studies.....	34
Figure 1.6: Comparison of the energy pathway for an enzyme controlled reaction in a mesophilic and psychrophilic enzyme.....	35
Figure 1.7: Folding energy landscapes hypothesised for psychrophiles and thermophiles.....	41

Figure 1.8: Representative ribbon structure (PDB code 2L15) of the CspA from <i>E. coli</i> showing the five β sheets wrapped around to form a β -barrel structure.....	44
Figure 1.9: <i>E. coli</i> CspA mRNA as a temperature sensor	47
Figure 2.1: Typical thermal calibrations scan with markers representing selection of the minima surrounding the first resonant peak	76
Figure 2.2: Example of a force against displacement calibration graph. The sensitivity is determined between the markers placed on the graph.....	76
Figure 3.1: Thermostability curve for a protein displaying the dependence of Gibbs free energy of unfolding $-\Delta G^{F-U}$ on temperature.	84
Figure 3.2: Three strategies employed by evolution to increase T_m in thermophilic proteins.	85
Figure 3.3: Evolution of mesophiles and hyperthermophiles evaluated by extrapolation to ancestral sequences	86
Figure 3.4: A comparison of the thermostabilities of α -amylase enzymes.....	87
Figure 3.5: Comparison of the temperature dependence of dihydrofolate reductases.....	88
Figure 3.6: Alignment of cold shock domain containing proteins from bivalves.	91
Figure 3.7: Multiple sequence alignments of different temperature adapted cold shock proteins.....	92
Figure 3.8: Csp alignment highlighting charged residues	94
Figure 3.9: Phylogenetic tree of a selection of bacterial Csp. Hot adapted Csp. shown in red and orange, temperate in green, cold adapted in blue and purple.....	95
Figure 3.10: 1.5% agarose gel showing addition of <i>Xho</i> I and <i>Mlu</i> I restriction sites to Csp sequences by PCR. The desired product is approximately 220 bp for each Csp.	96
Figure 3.11 Agarose gel displaying products of restriction digestion of pET3a plasmid.....	97
Figure 3.12: Expression profile of BL21 DE3 cells transformed with Csp sequences, induced with IPTG and incubated at 37 °C.....	98
Figure 3.13: Ni-NTA pull-down assays contrasting expression of cold shock proteins at 37 °C and 26 °C.....	99
Figure 3.14: Ni-NTA affinity purification of PB6 Csp.	100
Figure 3.15: Ion exchange separation of PB6 Csp and SDS-PAGE showing fractions eluted from a resource Q column.....	101
Figure 3.16: Size exclusion purification of PB6 Csp monomer using a Sup75 column.	102
Figure 3.17: Mass spectra of the purified Csp. with indicated masses of the largest peak top right in each picture.	103

Figure 3.18: CD spectra between 190-280 nm validating folded state of the Csp's and re-folding capability.....	104
Figure 3.19: An overlay of 5 repeats of thermal unfolding of <i>BsCsp</i> followed by circular dichroism at 222.6 nm in 63 mM sodium phosphate buffer.	105
Figure 3.20: Fluorescence spectra of <i>BsCsp</i> from 7 °C (dark green) to 82 °C (dark red) in 63 mM sodium phosphate buffer.....	109
Figure 3.21: Thermal unfolding of Csp's by CD (left) and tryptophan fluorescence followed by λ_{BCM} (right) in 63 mM sodium phosphate buffer.....	110
Figure 3.22: Thermal unfolding of Csp's with varying concentrations of GdnHCl measured by λ_{BCM} 320-380 nm.	114
Figure 3.23: Plots of ΔH against T_m values pairs obtained with increasing concentration of GdnHCl from right to left.....	114
Figure 3.24: Comparison of thermostability parameters between the Csp's. Values taken from fluorescence data	116
Figure 3.25: Thermostability profile for the Csp's in 63 mM sodium phosphate buffer derived from the Gibbs-Helmholtz equation.	117
Figure 3.26: Comparisons of Csp thermostabilities relative to the optimum growth temperatures of their host bacteria.....	121
Figure 3.27: Histogram showing the relative thermostabilities of the Csp's at the respective optimum growth temperatures of their host bacteria (left) and at 8 °C below that temperature (right)	121
Figure 3.28: α -amylases compared relative to the temperature at which their host organism grows optimally.	122
Figure 3.29: Equilibrium curve showing the unfolding of Csp in 63 mM sodium phosphate buffer in response to guanidine hydrochloride denatureant at 10 °C.....	124
Figure 3.30: Comparison of Gibbs-Helmholtz projection with chemical denaturation approaches.	125
Figure 3.31: Plot of thermostability of the Csp's against the optimum growth temperature of their host organism.....	126
Figure 3.32: Secondary structure and sequence comparisons of <i>BsCsp</i> WT, <i>BsCsp</i> CTM and <i>TmCsp</i>	128
Figure 3.33: Thermal unfolding of CTM compared with <i>BsCsp</i> and <i>TmCsp</i>	129
Figure 3.34: Thermostability plot for CTM	131
Figure 3.35: Location of the modifications in the hydrophobic deletion variants.	132

Figure 3.36 Comparison of the thermal and chemical stabilities of the hydrophobic deletion mutants <i>BsCsp</i> and <i>TmCsp</i> measured by tryptophan fluorescence	133
Figure 3.37: Thermostability projection of <i>TmCsp</i> V62A and L40A in 63 mM sodium phosphate buffer compared to <i>TmCsp</i> WT.....	134
Figure 4.1 PB6 Csp folding and unfolding by fluorescence stopped flow	145
Figure 4.2: Chevron plot of Csp folding and unfolding at 10 °C.....	146
Figure 4.3: Schematic showing the use of the MBP fusion protein to express PB6 Csp in minimal medium.....	150
Figure 4.4: 1.5% Agarose gel showing double digest of pMal C5X plasmid containing MBP fusion proteins.....	151
Figure 4.5: Expression of MBP-PB6 Csp in minimal medium.....	151
Figure 4.6: Elution profile displaying fraction eluted from Ni-NTA purification of MBP-PB6 fusion protein.....	152
Figure 4.7: Ion exchange separation of MBP-PB6 Csp fusion protein and SDS-PAGE showing fractions eluted from a resource Q column.....	152
Figure 4.8: TEV cleavage of MBP-PB6 Csp to separate PB6 Csp from MBP-PB6 Csp	153
Figure 4.9: Purification of PB6 Csp from a mix of PB6 Csp, MBP, MBP-PB6 Csp and TEV by size-exclusion chromatography.....	153
Figure 4.10: HSQC of PB6 Csp in 63 mM sodium phosphate buffer at 10 °C taken on a 750 MHz spectrometer.....	155
Figure 4.11: HSQC of <i>BsCsp</i> in 63 mM sodium phosphate buffer at 10 °C taken on a 750 MHz spectrometer.....	156
Figure 4.12: HSQC of <i>TmCsp</i> in 63 mM sodium phosphate buffer at 10 °C taken on a 600 MHz spectrometer	156
Figure 4.13: Temperature dependence of the HSQC spectra of PB6 Csp.....	158
Figure 4.14: Temperature dependence of the HSQC spectra of <i>BsCsp</i>	159
Figure 4.15: Temperature dependence of the HSQC spectra of <i>TmCsp</i> . Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer	160
Figure 4.16. HSQC spectra showing the changes to PB6 Csp upon binding of DNA. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer	163
Figure 4.17. HSQC spectra showing the changes to <i>BsCsp</i> upon binding of DNA. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer	164

Figure 4.18. HSQC spectra showing the changes to <i>TmCsp</i> upon binding of DNA. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer	165
Figure 4.19 Secondary structure of <i>BsCsp</i> highlighting region which differed in the DNA form with temperature.....	166
Figure 4.20: Temperature dependence of the HSQC spectra of DNA-bound PB6 Csp	167
Figure 4.21: Temperature dependence of the HSQC spectra of DNA-bound <i>BsCsp</i>	168
Figure 4.22: Temperature dependence of the HSQC spectra of DNA-bound <i>TmCsp</i>	169
Figure 5.1: Free energy profile showing the effect of force on the energy levels of the folded (F), transition state (TS) and unfolded (U) protein states in a two-state transition.....	172
Figure 5.2: Schematic of an AFM microscope being used for single molecule unfolding of a protein.	173
Figure 5.3: Schematic of the methodology of unfolding a polyprotein by AFM including analysis method.....	177
Figure 5.4: Unfolding data for (I27- <i>TmCsp</i>) ₃ -I27	185
Figure 5.5: The range of the thermostability profile surveyed by AFM experiments	187
Figure 5.6: Schematic representing the two-step molecular biology approach involved in the production of the <i>TmCsp</i> homopentamer <i>TmCsp</i> ₅	188
Figure 5.7: 2% Agarose gel displaying amplification of PCR templates of <i>TmCsp</i> cassettes for position 1 and 5 in duplicate.....	189
Figure 5.8: Agarose gel displaying restriction digest products used to generate <i>TmCsp</i> ₅ plasmid....	190
Figure 5.9: Nickel pull-down assay of I27- <i>TmCsp</i> ₅ -I27 expression with IPTG in <i>E. coli</i> BLR [DE3] pLysS cells.	191
Figure 5.10: Schematic showing production of an (I27-PB6 Csp) ₃ -I27 polyprotein DNA construct in a pET3a plasmid using the Gibson reaction	192
Figure 5.11: Agarose displaying amplification of PB6 Csp cassettes by PCR for the different sites within the polyprotein	192
Figure 5.12: Products of the amplification region of the Gibson reaction shown on a 1.5% agarose gel.....	193
Figure 5.13: Using a second stage to the Gibson reaction to replace the final I27 domain containing a mutant with a fresh I27 domain.....	194
Figure 5.14: Agarose gel of colony PCR products displaying the sizes of the DNA sections within the polyprotein region.....	194

Figure 5.15: Expression trial of the (I27-PB6 Csp) ₃ -I27 polyprotein in <i>E. coli</i> BLR [DE3] PLYS cells. .	195
Figure 5.16: SDS-PAGE of Ni-NTA separation of the proteins from large scale expression of (I27-PB6 Csp) ₃ -I27 polyprotein.	196
Figure 5.17: Ion exchange separation of (I27-PB6 Csp) ₃ -I27 and SDS-PAGE showing fractions eluted from a resource Q column.	197
Figure 5.18: Size exclusion purification of (I27-PB6 Csp) ₃ -I27 using a Sup75 column.....	198
Figure 5.19: CD spectra of (I27-PB6 Csp) ₃ -I27	199
Figure 5.20: Thermal denaturation of (I27-PB6 Csp) ₃ -I27	199
Figure 5.21: Mechanical unfolding of (I27-PB6 Csp) ₃ -I27.....	201
Figure 5.22: Representative AFM unfolding traces for (I27-PB6 Csp) ₃ -I27 performed at 14 °C in 63 mM sodium phosphate buffer at a series of pulling speeds.....	202
Figure 5.23: Histograms showing the normal distribution of F_U (upper) and interpeak distance (lower) for unfolding of (I27-PB6 Csp) ₃ -I27 in 63 mM sodium phosphate buffer	205
Figure 5.24: Scatterplots for (I27-Csp) ₃ -I27 unfolding in 63 mM sodium phosphate buffer at a pulling speed of 600 nm/s and temperature 14 °C.....	207
Figure 5.25: Median unfolding forces across triplicate data sets for the Csp's within (I27-Csp) ₃ -I27 constructs.....	208
Figure 5.26: Comparison of the unfolding force pulling speed dependence of I27 (orange tones) and PB6 Csp (blue tones) in the (I27-PB6 Csp) ₃ -I27 polyprotein construct performed at a series of different temperatures	210
Figure 5.27: Comparison of the unfolding force pulling speed dependence of the I27 marker protein and Csp for all 3 constructs at a range of temperatures. <i>TmCsp</i> and <i>BsCsp</i> at 5 °C and 23 °C performed by K. M. Tych.....	211
Figure 5.28: A comparison of the unfolding force temperature dependence of the Csp's with <i>TmCsp</i> (red), <i>BsCsp</i> (green) and PB6 Csp (blue) at 600 nm/s left and 2000 nm/s right	211
Figure 5.29: Expression of the median unfolding force of the Csp's at specified temperatures as a proportion of their unfolding force at 600 nm/s 5 °C.....	212
Figure 5.30: Comparison of the variation of Δx_u with temperature for the Csp's with <i>TmCsp</i> in red, <i>BsCsp</i> in green and PB6 Csp in blue. Error bars were determined from the bootstrapping method.	215
Figure 5.31: variation of k_u with temperature for the Csp's with PB6 Csp plotted in blue, <i>BsCsp</i> plotted in green and <i>TmCsp</i> red	216

Figure 5.32: Variation in mechanical energy barrier heights with temperature. <i>TmCsp</i> is shown in red, <i>BsCsp</i> in green, PB6 Csp in blue and I27 taken from (I27-PB6 Csp) ₃ -I27 construct shown in orange.	217
Figure 5.33: Comparison of the unfolding energy landscapes for the Csp	218
Figure 5.34: Comparison of the unfolding energy landscapes for the 3 different Csp at different temperatures	219
Figure 5.35: Comparison of the unfolding energy landscape of the Csp at the respective optimal growth temperatures of their host organisms with <i>TmCsp</i> red $T_{opt} = 80\text{ }^{\circ}\text{C}$, <i>BsCsp</i> green $T_{opt} = 37\text{ }^{\circ}\text{C}$ and PB6 Csp $T_{opt} = 22\text{ }^{\circ}\text{C}$	220

List of Tables

Table 2.1: Composition of bacterial growth media used.	53
Table 2.2: PCR primers utilised in PCR DNA amplifications.	54
Table 2.3: Composition of PCR mixtures performed in triplicate.	54
Table 2.4: Thermal cycling conditions for PCR amplification.	55
Table 2.5: Composition of a typical restriction digest mixture used in this investigation. Restriction digestion mixes were set up on ice.	56
Table 2.6: Primers used for the Gibson assembly reaction synthesis of (I27-PB6 Csp) ₃ -I27	59
Table 2.7: Plasmid and cell types used over the course of the study	60
Table 2.8: The compositions of the two sections of an SDS-PAGE gel. The volumes shown are sufficient for making 2 gels.	62
Table 2.9: Components for M9 minimal salts.	67
Table 2.10: Components for M9 minimal media.	67
Table 2.11: Parameters fixed in the Monte Carlo simulations for (I27-Csp) ₃ -I27	81
Table 3.1: The number of charged residues present in different Csp from hyperthermophilic top and extreme psychrophile bottom	94
Table 3.2: Percentage sequence identity between the different Csp homologues	96
Table 3.3: Comparison of stabilities between different Csp determined by CD and Tryptophan fluorescence in 63 mM sodium phosphate buffer.	111
Table 3.4: Thermostability parameters for the Csp.	115
Table 3.5: Maximum thermostability parameters for the Csp in 63 mM sodium phosphate buffer	118

Table 3.6: Parameters derived from fitting of equilibrium denaturation curves performed with guanidine hydrochloride at 10 °C and 23 °C. Errors indicated derived from fitting.	124
Table 3.7: Parameters derived from fitting of equilibrium denaturation curves performed with urea at 10 °C. Errors indicated derived from fitting.	125
Table 3.8: Parameters obtained from fitting of thermal denaturation of <i>TmCsp</i> , <i>BsCsp</i> and <i>BsCsp</i> CTM	129
Table 3.9: Parameters extracted from chemical denaturation of the CspS at 10 °C by GdnHCl.....	130
Table 3.10: Thermostability parameters for <i>TmCsp</i> , <i>BsCsp</i> and CTM in 63 mM sodium phosphate buffer. ΔC_p values were obtained from gradients on Figure S3.....	131
Table 3.11: Thermal unfolding parameters including V62A and L40A in 63 mM sodium phosphate buffer	133
Table 3.12: Values obtained from fits to equilibrium denaturation curves with GdnHCl at 10 °C in 63 mM sodium phosphate buffer	134
Table 3.13: Thermostability parameters of V62A and L40A mutants in 63 mM sodium phosphate buffer	135
Table 4.1: Stopped flow derived parameters at 10 °C in 63 mM sodium phosphate buffer.	147
Table 4.2: Stopped flow derived parameters derived at 10 °C in 63 mM sodium phosphate buffer .	148
Table 4.3: Comparison of ΔG^{F-U} obtained by stopped flow at 10 °C with equilibrium curve values obtained at 10 °C in 63 mM sodium phosphate buffer	148
Table 5.1: Comparison of T_m values for PB6 Csp and I27 as monomers and within (I27-PB6 Csp) ₃ -I27	199
Table 5.2: Unfolding forces and interpeak distances obtained at 14 °C for (I27-PB6 Csp) ₃ -I27.....	206
Table 5.3: Mechanical energy barrier properties derived from Monte-Carlo simulations and the Arrhenius equation	216

List of Abbreviations

$-(\Delta G^{F-U})$ - Change in Gibbs free energy from folded to unfolded state

ΔG^{F-TS} - Change in Gibbs free energy from folded to transition state

AFM - Atomic force microscopy

APS - Ammonium persulphate

BCM - Barycentric medium

BLAST - Basic Local Alignment Search Tool

bp - Base pairs

BSA - Bovine serum albumin

BsCsp - *Bacillus subtilis* Cold shock protein

BcCsp - *Bacillus cadalyticus* Cold shock protein

Csp - Cold shock protein

CD - Circular dichroism

CTM - Charged triple mutant

kDa - kiloDaltons

DNA - Deoxyribonucleic acid

dNTP - Deoxyribonucleotide triphosphate

DTT - Dithiothreitol

E. coli - *Escherichia coli*

EDTA - Ethylenediaminetetracetic acid

F_U - Unfolding force

FT - Flow through

HSQC - Heteronuclear single quantum coherence spectroscopy

IPTG - Isopropyl β -D-1-thiogalactopyranoside

LB - Lysogeny broth

LUCA - Last universal common ancestor

MBP - Maltose binding protein

MC - Monte Carlo

MS - Mass Spectrometry

Ni-NTA - Nickel Nitrilotriacetic acid

NMR - Nuclear magnetic resonance

OD₆₀₀ - Optical density at wavelength of 600 nm

P₂P - Peak to peak

PB6 Csp - *Psychrobacter* sp6 Cold shock protein

PCR - Polymerase chain reaction

PDB - Protein data bank

psi - Pounds per square inch

PsiCsp - *Psychromonas ingrahamii* Cold shock protein

RNA - Ribonucleic acid

rpm - Revolutions per minute

SDS-PAGE - Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SheCsp - *Shewanella livingstonensis* Cold shock protein

SMFS - Single Molecule Force Spectroscopy

SOB - Super optimal broth

SOC - Super optimal broth with catabolite repression

TAE - Tris-acetate-EDTA

TEMED - Tetramethylethylenediamine

TmCsp - *Thermotoga maritima* Cold shock protein

Tris - Tris(hydroxymethyl)aminomethane

U.V. - Ultra violet

WLC - Worm-like chain

1. Introduction

1.1. Protein stability

Proteins perform a vast array of different essential processes within organisms and to be able to perform these functions most proteins must be folded into precise 3-dimensional structures. Proteins are synthesised as a polypeptide chains and as they fold and acquire more native contacts their free energy decreases and the number of possible conformations the polypeptide chain can adopt decreases (1). Folding proceeds by the path of least resistance with intermediates converging towards an energy minimum to minimise frustration in the protein (2) There are however examples of proteins which do not require a specific 3D shape, termed disordered proteins which include linkers and some ribosomal proteins (3).

The stability of a protein depends on its amino acid sequence. Even a single amino acid mutation can convert a highly stable protein into one which is unstable or cannot fold correctly. Interactions stabilising protein tertiary and quaternary structures include hydrogen bonding, electrostatic interactions, van der Waals forces, disulphide bonds, dipole-dipole interactions and hydrophobic effects (4). The most numerous interaction types are hydrogen bonding and hydrophobic interactions. They are individually weak but provide strength through large numbers of interactions.

The stability of proteins is not only dependent on the polypeptide chain alone but also on the surrounding solvent. Unlike many other solvents, water molecules are small enough to enter cavities within proteins. There are also temperature dependent changes in the density of water molecules and the ordering of water molecules (5-6). This gives a unique pattern of energy changes which are specific to water. $\Delta\Delta G_c$ is the change in free energy between solvent states which in water is equivalent to enthalpy ΔH . $T\Delta S$ is the product of temperature and entropy. Both of

these factors are temperature dependent as shown in Figure 1.1 but have similar magnitudes in protein systems around physiological temperatures.

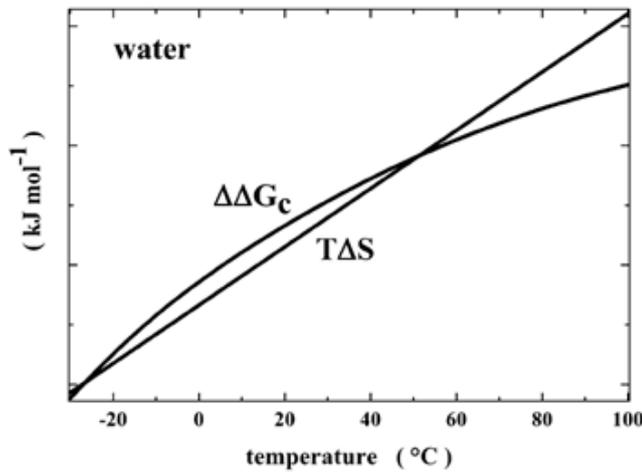


Figure 1.1: Temperature dependence of $\Delta\Delta G_c$ and $T\Delta S$ with respect to temperature. With water as a solvent $\Delta\Delta G_c$ is equivalent to change in enthalpy ΔH . T is the temperature in kelvin and ΔS represents change in entropy. Taken from (5)

The balance of enthalpy and the product of entropy and temperature are linked by the Gibbs free energy equation (equation 1.1). By subtracting the value of $T\Delta S$ from ΔH ($\Delta\Delta G_c$) in the above plot it can be seen that the protein only remains stable over a limited range of temperatures. As the energy level of the folded state of a protein is only marginally more stable than an array of unfolded states this balance can easily be tipped in favour of the unfolded state by changes in pH, temperature or the presence of chemical denaturants (7).

$$\text{Eqn 1.1: } \Delta G^{F-U}(T) = \Delta H^{F-U} - T\Delta S^{F-U}$$

1.1.1 The protein stability-function trade-off

The marginal stabilities of current proteins seems surprising as it may be expected that natural selection would lead to a trend of increasing thermostability. Proteins have not evolved the maximum thermostability that they could possibly achieve because protein evolution has had to balance thermostability with maintaining the

flexibility necessary for protein function. A degree of flexibility is required for the function of many proteins, from rearrangements in enzymes during catalysis to transporter proteins which may open up and close to structural proteins such as muscles and elastic tissue involved in responding to mechanical forces (8). With enzymes the theory of induced fit formulated by Koshland involves the shape of an enzyme adapting to the binding of a substrate and was postulated based on studies of the binding of glucose by hexokinase (9-10). The more recently devised theory for catalysis of conformational selection also relies on distinct molecular motions within the enzyme to allow the enzyme to be able to bind to substrate, transition state and products (11). A lower thermostability of proteins may also make proteins easier to degrade when the protein is no longer required although this also depends on kinetic effects.

The rates of chemical reactions can be increased by performing the reaction at higher temperatures however the thermostability of proteins can become a limiting factor as increasing thermal vibrations at high temperatures can lead to the loss of protein 3 dimensional structures. To be able to remain stable at higher temperatures a protein can be genetically modified to contain amino acids that form additional bonds such as salt bridges or a more tightly packed hydrophobic core (12-13). This can produce a protein which is not only more thermostable but also more robust against changes in factors such as pH. Such changes however can interfere with the flexibility of the protein constraining it and making it less able to exhibit the molecular motions required for its function (14-15).

Alternatively a protein can be engineered to exhibit more flexibility to try and improve the rate at which it performs its function by weakening or removing some of the interactions which can restrict the internal molecular movements. This can however lead to a destabilisation of the folded state relative to the variety of possible unfolded states making the protein unstable. This can lead to transient

unfolding or the protein may not be able to fold to a stable 3 dimensional structure at all (13)(15).

The balance between the limitations of thermostability and dynamics is often referred to as the stability-function trade-off and provokes some interesting prospects (16). The first is how the thermostability of a protein can be engineered to take advantage of the greater reaction rates at higher temperatures without adding interactions which themselves disrupt enzyme activity (17-18). If a commercial enzyme is modified to be able to function at higher temperatures it may not be advantageous the modifications decrease activity more than the activity gain from using a higher temperature. The other approach is to look at optimising the activity by increasing flexibility and allowing some sacrifice of thermostability (19). This method may allow enzyme-controlled reactions that currently require heating to be conducted effectively at moderate temperatures saving on heating costs (20). The extent to which thermostability can be reduced and the protein still maintained in a folded state may be limited. It is also not yet fully understood how effective increases in flexibility are in improving enzyme activity as protein dynamics can be dependent on many factors.

1.2 Types of extremophiles

Some of the issues brought up by the stability-function trade-off have already been addressed in nature through the evolution of organisms living at high and low temperatures termed extremophiles. Interest in extremophiles increased with the discovery of *Sulfolobus acidocaldarius* in 1965 by Thomas and Louise Brock within a sulphurous, acidic, geothermal pool in the Yellowstone National Park (21-22). It could be cultured to temperatures as high as 95 °C and this sparked a wider interest into the extreme conditions at which life can survive (23).

There are a variety of different types of extremophile such as thermophilic organisms which live above 40 °C and analogously psychrophilic and psychrotolerant organisms which live below 15 °C (24-25). Those usually living in the intermediate range of temperatures are classified as mesophiles. Salt-loving bacteria termed halophiles are capable of growing in conditions with NaCl concentrations as high as 5 M while most organisms cannot grow efficiently at a concentration above 0.5 M NaCl (26). Terms such as extreme, obligate or the prefix hyper are used to denote organisms adapted irreversibly to their respective extreme condition such that they cannot survive under regular conditions. Terms like facultative or moderate are used where an organism can tolerate the extreme condition but do not require it to survive. Other types of extremophiles include piezophiles which can grow at the high pressures above 40 Mpa found on ocean floors, acidophiles which grow at low pH levels and alkaliphiles which can grow in high pH conditions (27-29). Extremophilic organisms may be categorised into multiple classifications such as the thermoacidophile *Sulfolobus acidocaldarius*, although some combinations may be mutually exclusive. It is worth noting that all of the extremophilic terms are relative and are considered from the human perspective which may not be fully representative (30). This investigation will primarily focus on the temperature adapted extremophiles.

How extremophiles have evolved is still subject to debate. Some theories based around extrapolating modern protein and genetic sequences back to ancestral sequences suggest that the last universal common ancestor (LUCA) of all organisms was thermophilic. It may be that current mesophilic proteins were derived from thermophilic proteins which evolved to be more mesophilic due to a decrease in the Earth's temperature and extreme cold periods in Earth's history (31-34). It is unclear if current thermophilic proteins maintained thermostability through evolution or if they evolved to be mesophilic proteins then re-evolved to higher temperatures (35).

1.2.1 Using extremophilic proteins to address the stability-function trade off

The potential of the proteins of extremophilic organisms for biotechnological uses was highlighted by Thomas Brock through the discovery of another thermophile in Yellowstone National Park called *Thermus aquaticus* (36). This bacterium contains a particularly thermostable DNA polymerase commonly known as taq polymerase which has become routinely used in the polymerase chain reaction (PCR) as it can survive the high temperatures over repeated thermal cycles that is required in this procedure without loss of activity (37). More recently, thermostable enzymes have been discovered that maintain activity beyond the range over which life is seen with the thermostable enzyme amylopullulanase still retaining activity up to 142 °C. Thermostable enzymes have long shelf-lives and are less susceptible to changes in pH. (38).

Thermophilic proteins offer increased thermostability to achieve greater reaction rates while proteins derived from psychrophilic organisms may function with greater activity at lower temperatures. Both classes of protein have evolved effective strategies to balance stability and flexibility at their respective operating temperatures (39-40). Flexibility is particularly important in the active site. In hyperthermophiles the active site still needs to be stable at very high temperatures and yet still able to change to fit the substrate, transition state and products (42). A tightly bound substrate is favourable to stability but this makes conversion to the transition state bound form more difficult and can increase the energy barrier for the enzyme substrate complex to convert to the transition state (41). A study by *Scoichet et al.* into a series of lysozyme mutants optimised for increased thermostability and a second group optimised for increased activity showed that those maximised for activity gave poor stability and vice versa (13).

The modifications used in extremophilic proteins could provide insight that can be used to optimize the thermostability and flexibility of other commercially interesting proteins. To do this, extremophilic proteins need to be compared to

mesophilic proteins to identify what the differences are and what effects these have on protein thermostability and activity (43-44). Thermophilic proteins have improved stability relative to mesophilic proteins and some of the literature states that cold-adapted proteins may equivalently have more optimised dynamics (41). A better knowledge of how psychrophilic and thermophilic proteins differ from mesophilic proteins may allow optimisation of dynamics or thermostability in mesophilic proteins. Where these two types of adaptations do not exclude each other this could allow the production of proteins that display an improvement in both characteristics (45). As proteins feature in so many metabolic processes more knowledge of extremophilic proteins could help us to better understand the limits of protein thermostability at the boundaries of where life can exist.

1.3 Methods of comparing protein stabilities

1.3.1 Mutational studies

There are two separate ways to deduce the impact of structural features in increasing protein stability, flexibility or activity. The first is to perform mutational studies which involve altering the genetic DNA sequence of the protein to change one or more of the amino acids. This can involve insertion or deletion of sets of 3 bases while maintaining the correct reading frame of the sequence or the substitution of one or more bases. To examine the effect of a salt bridge interaction for example, the charged residues might be substituted for polar residues and the stability of the mutant compared with that of the wild type. Stability can be qualified using heat, chemical denaturants or mechanical forces if measuring mechanical stability. Heat and chemical denaturation are examples of thermodynamic effects while mechanical stability is dependent on kinetics. The role of hydrophobic residues to increase stability through core packing can be assessed by mutating larger hydrophobic residues to smaller ones such as alanine or glycine. Other aspects investigated by mutation studies could include active site mutations followed by assays of enzyme activity or removal of cysteine residues to measure the impact of disulphide bridges on stability.

Some studies have been performed which substitute residues in a protein for all 19 possible alternatives although this leads very quickly to a vast number of protein variants. To investigate P number of positions, with y number of variations at each position, gives N number of mutants to produce (equation 1.2). If y = 19 then this very quickly generates a large value of N as P increases so a more selective approach is preferable.

$$\text{Eqn 1.2: } N = y^P$$

Mutants may be strategically selected to discover the roles of residues using predictions based on knowledge of the protein structure and function. Alternatively they may be selected with a systematic approach such as replacing each aromatic residue in the protein with alanine, in turn. The main difficulty is that each separate mutant must be generated by site-directed mutagenesis and expressed. Some mutants may not express and large scale mutant studies can be laborious and time consuming. In such studies the substitutions which are made are generally of amino acids with similar properties. Substitutions of hydrophobic core residues for polar or charged residues or vice versa are uncommon as these often do not result in a protein which can fold correctly. As well as hydrophobicity and charge, the size of amino acids must also be considered to avoid creating steric clashes. In deletion mutants, loop size changes are often studied. Predicting the effects that individual mutations will have is a great challenge in molecular biology though programmes are being developed which are working towards this (46-47).

The cumulative effect of multiple mutations can also be assessed by this method (48). If for example 2 mutations both stabilize a protein, a double mutant may be made containing both mutations and this may stabilise the protein by more than the sum of the individual mutations. However combined mutations may also conflict with others such as causing steric hindrances and repulsions between like charges leading to a decrease in protein stability. Mutations may have undesirable

effects such that a mutation to improve thermostability of an enzyme may give the intended stability increase but may also reduce the activity of the protein. If this is for an enzyme to be used at low temperatures then some loss of stability may be acceptable so what properties are ideal depends on the practical use.

The main difficulty with trying to optimise thermostability is that the majority of mutations will decrease or abolish the thermostability of a protein and only a small number increase it. As more interactions become optimised, the proportions that further increase thermostability fall further as there are many more interactions that could be disrupted with each mutation. To be able to function at high temperatures thermophilic proteins have evolved complex combinations conferring thermostability and function that would be very complex to discover through mutational studies. For example, a variant of an enzyme which has been isolated from a hyperthermophilic organism represents a good starting point for rational design of an enzyme to be used at high temperatures. Better knowledge of what makes a protein exhibit greater thermostability or activity can be derived from these studies and hopefully applied to other proteins, however, patterns observed in one protein family may not always be applicable to other protein families.

One study that can be performed for temperature adapted homologues is to substitute equivalent residues from a thermophilic homologue into a mesophilic protein to see what effects these residues may have towards temperature adaption. This strategy however does not provide the optimisation of surrounding residues which can be important such that one salt bridge can be destabilising but a network of salt bridges highly stabilising. Temperature adapted proteins have many mutations to optimise function and some may involve general patterns based on many small optimising improvements that act co-operatively to produce an adaptive effect. This is particularly pertinent for psychrophilic protein where genetic modification to produce individual mutations or double and triple mutants cannot show these more subtle effects.

1.3.2 Comparative studies

An alternative approach to study differences between similar proteins is comparative studies which can allow vast numbers of changes to be seen quickly and enable trends and patterns to be picked out of data for further investigation. Homology searching allows the input of a DNA or protein sequence into a databank, producing an output of similar sequences from a vast array of different organisms. The effectiveness of such approaches have been dramatically improved by improvements in computing power and the recent generation of vast databases of genetic data and protein databanks (49-50). Current techniques allow millions of sequences to be rapidly compared and similarity scores ranked. To be able to get the most out of alignments of extremophilic protein sequences it necessary to have additional data about the living conditions of the organisms the proteins derive from.

Once many homologous sequences have been obtained, they can be aligned with each other and gaps added where amino acids are missing in some of the sequences (49). An alignment of temperature adapted proteins from within the same family can show what patterns of residues and motifs are important. It can also allow other trends such as the number and balance of charged residues to be observed or modifications in the length of loops.

Thermophilic homologues can often be easily identified by highly conserved stabilizing motifs which confer high sequence identity allowing easy identification in homology searches (12). There are very few combinations of mutations that provide sufficient thermostability for proteins at high temperatures which have evolved due to very strong selection pressure for greater stability imposed by extreme hot climates (40). The vast majority of mutations occurring in the sequence of proteins will decrease thermostability. When in conditions where there is a lower selection pressure for greater stability this would lead to a general trend of reducing thermostability. There may therefore be far more divergence in the pathways of

evolution of psychrophilic proteins. As there are also less conserved motifs in cold-adapted proteins they may be much harder to identify by homology searches.

There are some other issues around using databases for studies of extremophiles. As many proteins contain similar domains it is possible that a search for similar sequences could bring up some sequences which are not homologous. This can prove an issue where organisms have multiple different proteins sharing sections of conserved sequences. The composition of the databases can be highly weighted with large amounts of mesophilic and thermophilic sequence data but far more limited psychrophilic sequence data. It is thought that less than 1% of achieved protein sequences are psychrophilic (41). When comparing protein sequences it is important to consider that proteins may evolve to cope with a variety of extreme conditions and not necessarily just one factor. For instance, a comparison between a thermoacidophilic protein and a psychrophilic protein involves the comparison of multiple types of adaptation so cannot be considered as temperature adaptation alone (30). Identifying such cases requires good knowledge of the living conditions of the organisms investigated.

One of the greatest biological challenges is to be able to use primary sequence data to predict folded protein structures reliably and accurately (51-52). There is also the challenge of using sequence or structural data to predict the levels of thermostability or activity of a protein so then the effect of mutations could be estimated. Currently reliable values of thermostability parameters cannot be obtained from sequence data alone so separate proteins need to be expressed, purified and tested (53). While mesophilic proteins are generally easy to express, extremophilic proteins are more challenging. Once the homologues have been expressed, many different techniques can be applied to them and the effect of the adaptation ascertained very quickly. One issue with comparative studies can be where experiments are performed at different temperatures, in varying buffers or even using different techniques. Any changes in the method used to record data about

each protein produces method to method variation which can lead to inaccurate conclusions. Many comparative studies investigate mesophilic and thermophilic homologues but few include a psychrophilic homologue.

Comparative studies allow the effect of adaption to be seen across multiple organisms simultaneously and trends to be more easily discovered. Predictions can be made about which residue changes are important but comparative studies alone cannot distinctly show the effect and the contribution of individual changes to an adaptive effect.

1.3.3 Comparative studies using extremophilic protein structures

Relatively recently it has become possible to compare structural data between homologues though this is still highly restricted by the availability of structures. The impact of the differences between the sequences on overall protein stability can be inferred by structural comparisons. A combination of structural and bioinformatics studies have shown that only small numbers of amino acid substitutions occur between psychrophilic, mesophilic and thermophilic protein homologues (54). These are not generally distributed uniformly across the protein but instead concentrated in particular regions associated with protein functionality. Residues that have a structural role or bind ligands tend to be very highly conserved with substitutions often occurring in surrounding residues and loop regions. Catalytic residues are generally very highly conserved in extremophilic enzymes as they are required for protein function (55). Structural elements and overall topology are very highly conserved within protein families across homologues adapted towards different temperatures. This is shown by comparisons of the crystal structures of adenylate kinase proteins from *Bacillus globisporus*, *Bacillus subtilis* and *Bacillus stearothermophilus* in Figure 1.2 (56). The fact that only small numbers of amino acids differ between psychrophilic, mesophilic and thermophilic homologues, makes sequence comparisons very useful.

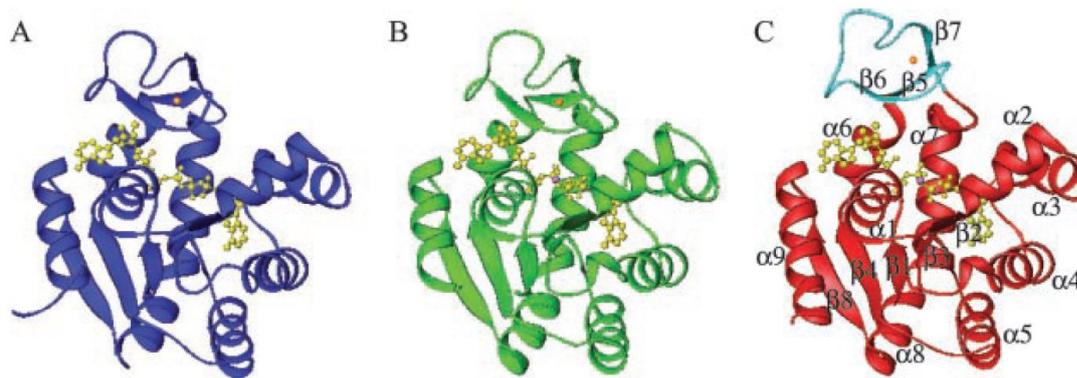


Figure 1.2: A comparison showing the similarity of adenylate kinase structures. The adenylate kinases were isolated from the psychrophile *Bacillus globisporus* (A, on the left), the mesophile *Bacillus subtilis* (B, in the centre) and the thermophile *Bacillus stearothermophilus* (C, on the right). The structures share amino acid identity of between 65% and 75%. Yellow spheres represent active site residues and the cyan residues the lid domain. Taken from (56).

A significant difficulty with this approach is that relatively few X-ray crystal structures (<50) of psychrophilic proteins have been obtained in comparison to other extremophilic proteins. This may be due to these proteins being more thermolabile or difficulties in producing crystals of defined protein conformations due to the increased flexibility of psychrophilic proteins (55)(57).

1.3.4 Thermophilic protein adaptations

The previous sections have highlighted some of the methodology of comparative studies and recent improvements. Using these approaches a significant amount has been discovered about thermophilic proteins and the adaptations they have evolved to maintain thermostability at high temperatures (16)(44). In 1975 *Perutz et al.* performed the first comparative study into the differences between homologues of the same protein family from organisms adapted to different temperatures (58). The study identified differences in the amino acid composition between a mesophilic and thermophilic homologue of ferredoxin with an elevation in numbers of charged residues giving increased numbers of ionic bonds in the thermophilic variant (59). Other studies have made comparisons with proteins from extreme

thermophiles such as *Thermotoga maritima* which lives in deep sea thermal vents (60). A study by Robinson-Rechavi and Godzik found that proteins from *Thermotoga maritima* have an increase in hydrophobic packing density compared to *E. coli* proteins (61). The amino acid composition of thermophilic proteins has been found to contain greater numbers of long chain and aromatic residues for improved hydrophobic packing and more structurally constraining amino acids such as proline (62) Since this investigation, a large number of studies have been performed comparing the stabilities of thermophilic and mesophilic protein homologues which have confirmed trends such as greater number of salt bridges and increased hydrophobic packing in thermophiles (63-64). A collection of these data are seen in reference (16). The adaptations of thermophilic proteins have become a focus for scientific study, as thermostable enzymes can allow reactions to be performed at higher temperatures which yield faster reaction rates. There is much however still to learn about how psychrophilic proteins are adapted and how this could be applied to improve protein function at low temperatures.

1.4 Psychrophilic proteins

1.4.1 Life at cold temperatures

A general definition for a psychrophile as stated by Morita in 1975 is an organism that has an optimum growth temperature of less than 15 °C and which cannot grow above 20 °C (24). Whilst psychrophiles are considered as extremophilic organisms the category actually encompasses a vast range of life, spanning the kingdoms of bacteria, archaea and eukaryotes (65). Much of the biosphere of the Earth is composed of cold climates with over 80% having a temperature consistently below 5 °C (66). This is because while we may primarily consider the land surface, most of Earth's surface is covered with deep oceans. It is believed that if life is found on the surface of Mars that it would fall into the psychrophilic category (25)(30). Organisms such as the Antarctic bacterium *Moritella profunda* have maximum growth rates at around 12 °C and grow as low as 2 °C (67). The lowest temperature for life to be sustainable was, until recently, thought to be around -20 °C in high salt

environments such as brine veins in sea ice (19)(70). In recent years microbes have been discovered living in Antarctic dry valleys at temperatures as low as -60°C (68). Ice crystals may still contain small amounts of liquid water between them which can support life (19). The production of biological antifreezes such as glycolipids to prevent ice crystal nucleation appears important for decreasing the freezing point of water in some sub-zero psychrophiles (69-71). Despite the prevalence of psychrophiles within the biosphere it seems rather surprising that comparatively far fewer studies have been performed on them in comparison to thermophilic organisms. The effects of high temperatures on organisms are easier to observe as the heat causes clearer physiological damage to cells and denaturation of proteins (72).

Psychrophiles are not just able to grow at low temperatures but are adapted irreversibly to them such that they cannot grow at mesophilic temperatures (19)(30). Organisms which can facultatively grow at low temperatures but retain the capacity to grow at higher temperatures are termed psychrotrophs or psychrotolerant (71). The boundaries used in the literature can be quite varied so it is not always obvious which sub-category is being referred to. A further complexity is that temperatures during the summer and winter in many psychrophilic climates may vary significantly. Organisms may be able to tolerate degrees of low temperature but there are differences between organisms which can survive low temperatures for a period of time and those which thrive at low temperatures (30). Psychrophiles can be further divided by their range of growth temperatures with stenopsychrophiles only able to grow within a very narrow temperature range while eurypsychrophiles can survive in a wider range of temperatures (71). Stenopsychrophiles tend to be more specifically adapted to occupy a specific niche having faster growth rates under their optimised conditions. Eurypsychrophiles however are often the dominant branch as they are more able to cope with temperature fluctuations (73).

1.4.2 Challenges of life at low temperatures

Psychrophilic organisms face a wide range of challenges such as increased solution viscosity, decreased diffusion rates slowing both nutrient uptake and excretion, decreased protein synthesis rates, stabilization of nucleic acid secondary structures and protein mis-folding (71)(74). Psychrophiles require significant adaptations to cope with these factors, for example the membranes of psychrophilic enzymes show an increased number of unsaturated, branched and shorter fatty acids chains to maintain flexibility (20)(67)(75). This study examines the adaptations of proteins to cope with issues regarding protein activity and protein folding changes in cold-adapted organisms.

One of the most significant problems psychrophiles face is the exponentially decreasing rates of reaction with lower temperature. This corresponds to the Boltzmann equation (Eqn 1.3) shown below (65). In the equation k_{cat} is the enzyme rate, k the transmission coefficient, k_B the Boltzmann constant, T the temperature in Kelvin, R is the universal gas constant, ΔG^{F-TS} is the activation energy and h is Plank's constant.

$$Eqn\ 1.3: k_{cat} = k \frac{k_B T}{h} e^{-\Delta G^{F-TS}/RT}$$

A decrease in temperature of 10 °C is generally thought to decrease reaction rate by between 2 and 4 times. Increasing enzyme expression to counter this effect would be very energetically expensive and so psychrophiles have instead evolved to increase enzyme activity and the theory generally proposed is that this occurs through increases in flexibility (54)(75)(77). At the same low temperature the specific activity of psychrophilic enzymes can be up to 10 times greater than their mesophilic homologues. However, when comparing the activity of both types of enzyme at their respective optimal operating temperatures, the activity of psychrophilic enzymes is still generally significantly below that of mesophilic enzymes as shown in Figure 1.3 (78).

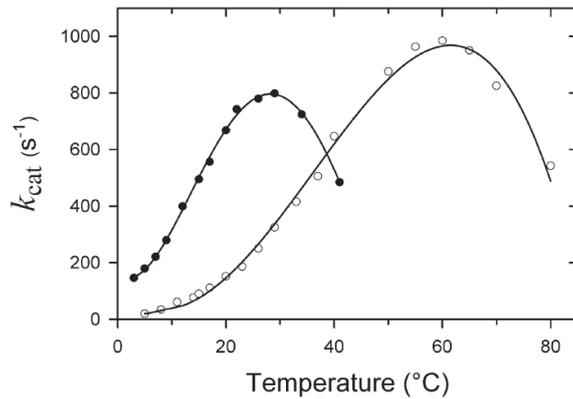


Figure 1.3: The changes in enzyme activity at different temperatures between a mesophilic α -amylase and a psychrophilic α -amylase from *Pseudoalteromonas haloplanktis*. The psychrophilic protein is represented by closed circles and the mesophilic variant by open circles. Adapted from (41)(78).

A comparison of the percentage activity and percentage folding for α -amylase enzymes is shown in Figure 1.4. As would be expected the order of temperature of maximum activity and the temperature at which the Csps remain stable both follow the same hierarchy psychrophilic homologue < mesophilic homologue < thermophilic homologue (41)(78). It is also interesting to see that for the thermophilic and mesophilic homologue the temperature of maximum activity is very near the thermal stability limit of the protein. With the psychrophilic variant however the temperature of maximum activity decreases substantially before the limits of thermal stability. It could be that in psychrophiles adaptation of the active site leaves it more heat labile than the rest of the protein structure which has been suggested by chemical denaturation studies, activity assays and molecular dynamics simulations (62)(79-80). Alternatively other temperature sensitive factors could be affecting enzyme reaction rates beyond unfolding of parts of the protein structure.

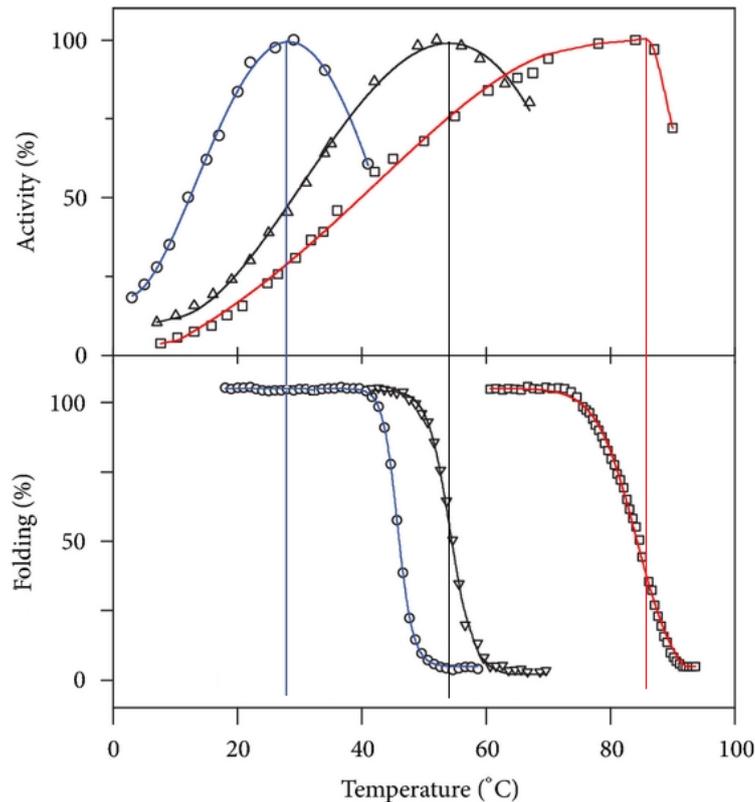


Figure 1.4: Comparison between thermostability and activity of 3 alpha amylases isolated from organisms which live at vastly different temperatures. Percentage activity (top) and percentage of folded protein (bottom) at different temperatures for α amylases isolated from the thermophile *Thermobifida fusca* in red, mesophile: pig pancreas in black, and the psychrophile *Pseudoalteromonas haloplanktis* in blue. Taken from (41).

One way to illustrate how the relative thermostabilities of temperature adapted proteins vary is to plot temperature dependence curves (81) (see section 3). A greater height indicates a more thermostable protein at that temperature and where the values falls below the x-axis the protein would be unfolded at that temperature. Microcalorimetry studies performed on α -amylase enzymes have shown that the thermostability curve for the psychrophilic form exhibits reduced values of $-\Delta G^{F-U}$ as shown in Figure 1.5. The temperature of maximum activity varies significantly between the temperature adapted enzymes but the temperature of maximum protein thermostability varies to a much smaller extent (41). At temperatures around the freezing point of water some cold-adapted proteins

exhibit cold denaturation which could mean that their range of thermostability is narrower as well as shifted (76).

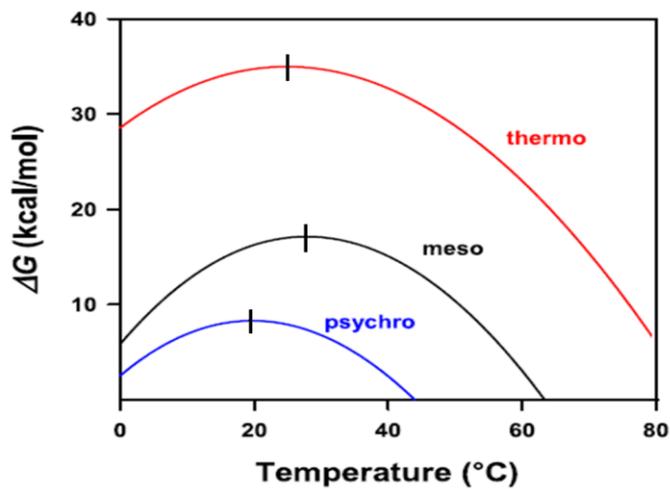


Figure 1.5: The differences in protein thermostabilities of temperature-adapted α -amylases obtained from microcalorimetry studies. The vertical lines mark the temperature of maximum thermostability. *Thermobifida fusca* (TFA) in red, mesophile: pig pancreas in black (PPA), and psychrophile *Pseudoalteromonas haloplanktis* (AHA) in blue. Adapted from (41).

While a reduced thermostability and greater low temperature activity is seen in the vast majority of studies into psychrophilic proteins there are still many unknowns. Lower thermostability may be a consequence of optimisation for greater flexibility to attempt to increase activity to offset the slow reaction rates present at low temperatures (71). Alternatively lower stability could simply be a product of less selective pressure for high thermostability with the enzyme activity improvements related to other factors than flexibility.

Lower thermostability may not be directly selected for in psychrophiles but may be either a consequence of genetic drift due to lack of selective pressure for greater thermostability or alternatively selective pressure to increase flexibility to attempt to offset the slow reaction rates present at low temperatures (71).

1.4.3 Adaptations of psychrophilic enzymes to cold conditions

It is thought that psychrophilic enzymes are able to maintain greater activity at low temperatures as they occupy a less stabilised native state with a reduced affinity for binding substrate (82). This reduces the energy well between free enzyme and substrate-bound enzyme forms hence decreasing the energy barrier between the enzyme-substrate complex and the transition state (Figure 1.6). The reduction in stabilization of the enzyme substrate complex is often achieved in psychrophiles by an opening up of the active site region. This may also serve to broaden the substrate specificity of psychrophilic enzymes.

In psychrophilic proteins, residues around the active site are often substituted for residues with shorter side chains therefore opening up the active site to easier entry of substrate and release of product (56)(82). Other distant regions that control changes to the active site shape or access to the active site are often prone to modification in extremophilic enzymes (57). Local, rather than global, changes in thermostability allow for increased protein movement in functional regions whilst having less effect on the overall thermostability of the protein structure (80).

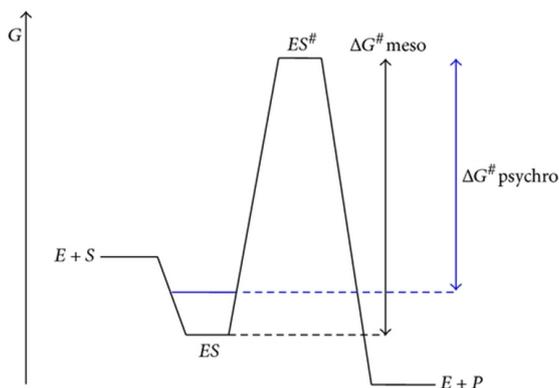


Figure 1.6: Comparison of the energy pathway for an enzyme controlled reaction in a mesophilic and psychrophilic enzyme. The reaction pathway of a mesophile is shown in black and the adjustment of weaker substrate binding evolved in a psychrophilic enzyme shown in blue. $ES^{\#}$ represents the transition state and $\Delta G^{\#}$ the energy barrier from bound enzyme to the transition state. Taken from (82).

Many of the interactions within proteins are strengthened at low temperatures making the protein become more rigid and conformational changes more difficult. To maintain function, psychrophilic proteins are believed to have evolved increased flexibility relative to mesophilic proteins through changes in amino acid composition, loss of some electrostatic interactions and a weakening of hydrophobic packing (74)(83-84). Increases in molecular dynamics may be crucial in allowing proteins to function effectively at low temperatures (85). Psychrophilic proteins are often less stable than their mesophilic homologues due to a lack of selective pressure for thermostability and selective pressure for greater activity (76). Increases in protein flexibility at the expense of thermostability do however have limitations as too much loss of thermostability could lead to unstable proteins which unfold too readily which would be selected against. Whilst the literature often presents thermostability and flexibility as inversely related they cannot be thought of as strictly antagonistic as some mutations increase both properties (45)(71).

The relative strength of interactions and the relative contributions of each specific type of interaction to overall protein thermostability vary with temperature even without changes in protein sequence (84). At lower temperatures the conformational entropy of molecules is reduced leading to strengthening of Van der Waals forces, ionic bonds and hydrogen bonds. This effect has the potential to prevent interchange between conformations restricting the protein movement required for functionality (19)(71). For this reason a decrease in ionic interactions and the replacement of charged amino acids for polar ones is often seen in psychrophilic proteins (76)(84). The specific charged residues which are lost may depend on the role of interactions with external solvent and their locations are usually more dispersed across the protein surface than in mesophiles or thermophiles (71).

Some psychrophilic proteins show an increase in overall protein negative charge compared with mesophilic homologues which may favour increased interaction with solvent (19)(74). Large alterations in surface charges by substitution of charged residues have been reported around ligand binding sites in some psychrophilic proteins including the NADH binding site of malate dehydrogenase (86). This could represent a way of using charge to help direct a co-factor or substrate to a binding site. Interactions with stabilizing ions such as metals can be weakened at low temperatures by orders of magnitude (54).

Psychrophilic proteins also show other changes in amino acid composition. The levels of the conformationally restraining amino acids, proline and the multiple hydrogen bond forming arginine are decreased in psychrophilic proteins (19)(71). A greater proportion of the amino acid substitutions occur in flexible loop regions and psychrophilic protein sequences frequently contain clusters of multiple flexible glycine residues which act as hinge regions (16)(71)(87). The nature of these changes to protein amino acid composition appears to be the opposite of those adaptations seen in thermophilic proteins (39)(43). Some psychrophiles show upregulation of many low molecular weight compounds such as carbohydrates, glycine and carnitine which are believed to prevent cold stress damage by helping to prevent protein denaturation (67).

1.4.4 The hydrophobic effect and cold-denaturation

Proteins generally have an amphipathic nature with hydrophobic amino acids in their interior packing tightly to avoid solvent contact and hydrophilic amino acids on the protein surface. When a protein unfolds this disrupts the arrangement of water molecules increasing the entropy of the system. Unfolding is however unfavoured as the denatured state would expose more hydrophobic residues to solvent (88-89). The shielding of hydrophobic groups from water is termed the hydrophobic effect (57). The hydrophobic effect is strongest at temperatures around room

temperature weakening both above and below this level. At low temperatures the structure of water becomes more ordered, forming hexagonal structures and the arrangement of water molecules at the solvent-protein interface changes (88)(90). The favourable entropic effect of dispersal of water molecules then becomes greater than the enthalpic penalty of the hydration of hydrophobic groups by solvent, pushing the equilibrium towards the unfolded state (43)(89). This unfolding at low temperature is termed cold denaturation and the restrictions it applies to the temperature range over which proteins can function at may limit life at low temperatures (25)(71).

Proteins contain a core of hydrophobic residues and an exterior of hydrophilic residues. Psychrophilic proteins however may contain more exposed hydrophobic groups and interior polar groups than mesophilic or thermophilic proteins. Reduced levels of hydrophobic packing may allow easier access of water molecules into protein clefts (89). Increased numbers of hydrophobic amino acids on the protein surface generates increased entropic destabilization with solvent molecules reducing the energetic penalty for solvent hydration (57). The hydration of groups with solvent may contribute towards psychrophilic protein flexibility and help counter the energetic problems caused by the formation of strong hydrogen bond networks in water at low temperatures (20). Weakening of the hydrophobic packing may however, make psychrophilic proteins more susceptible to cold unfolding. This may restrict the limits of life at cold temperatures and meaning further shifts in the functional temperatures of proteins to lower temperatures may not be possible (6)(19)(88)(91). The unfolded state produced by cold-denaturation of a protein may be very different from that produced by thermally unfolding of a protein as the denaturation occurs due to different effects (88).

1.4.5 Protein cold-adaption is an array of different approaches

Although these trends have been observed across many psychrophilic protein studies, not all psychrophilic proteins will exhibit each the traits discussed (19)(92). The picture becomes even more complex in multidomain proteins where domain interfaces may vary (93-94). Overall cold adaption of proteins appears not to be achieved by a single strategy but an array of different approaches depending on the protein concerned (41)(71). This has made it very difficult to engineer psychrophilic properties into mesophilic proteins by techniques such as rational design (19).

All of these described adaptations are proposed to boost conformational entropy, generally reducing protein thermostability and increasing the flexibility of psychrophilic proteins (85). The effects are not strictly antagonistic so not all mutations that increase flexibility also decrease thermostability. Cold-temperature adaptation of protein occurs, although not always in ways that would be predicted from thermophile enzymes (64). A study by *Xu et al.* attempted to improve the activity of a thermophilic enzyme by mutating each residue to its equivalent residue present in a psychrophilic variant (45). As expected, while many of the mutations improved activity most also decreased thermostability however a few of the mutations increased thermostability. This may be because psychrophiles rely on a smaller number of higher specific interactions to maintain thermostability while thermophiles use a more general stabilization of large numbers of interactions throughout the protein. Some of the mutations increased both thermostability and activity which is desirable for genetically engineered enzymes (45).

As thermostability mechanisms depend on each other in a co-operative way it is often difficult to distinguish the contribution of the different factors such as electrostatics to changes in protein thermostability (92). For example, formation of individual salt bridges can interfere with the interactions between protein and solvent, so can destabilise a protein. The contribution of salt bridges to protein thermostability is often dependent on the surrounding residues and the formation

of electrostatic networks. The presence of lone polar groups can reduce protein thermostability, however the formation of hydrogen bonds between polar groups can help stabilise proteins (95). Electrostatic optimisation also involves removal of repulsive matching charges as well as introduction of favourable opposite charge interactions. A study of 100 adenylate kinase mutants by *Howell et al.* has shown that optimisation of hydrophobic packing could achieve a T_m as high as 68 °C from a wild type T_m of 49.6 °C but that to achieve T_m values above this required electrostatic optimisation (53). With thermophiles, the evolutionary aim is to increase thermostability and residues mediating this can be elucidated through mutational studies. The effect of modifications can be followed by changes in the enthalpy of protein unfolding (ΔH) and Gibbs free energy of unfolding ΔG^{F-U} . The evolutionary aims of protein cold-adaption are believed to focus on increasing flexibility so it is more difficult to determine how much individual or sets of mutations contribute towards achieving that aim. This may be a reason why comparisons between homologue have tended to focus on the differences between mesophilic and thermophilic proteins.

1.4.6 Protein folding in cold-adapted proteins

With slower movement of molecules and an increase in the cell medium viscosity, protein folding represents a more complex challenge at reduced temperatures. Larger proteins generally unfold via sequential unfolding of structural domains of varying stabilities. Psychrophilic proteins may show far simpler folding landscapes with fewer intermediates, helping to prevent slow folding or mis-folding at low temperatures (74). Thermal unfolding studies of α -amylases have shown psychrophilic homologues to exhibit two-state kinetics with a sharp transition between folded and unfolded states whereas mesophilic and thermophilic forms show multiple intermediate transitions (39)(91). These simpler folding landscapes could make unfolding a more reversible process in psychrophilic proteins (19). The amount of time required to shift between configurations is significantly increased at low temperatures so protein folding in psychrophiles requires specific chaperones

(96). It has been observed that expressing psychrophilic chaperones in *E. coli* allows growth to lower temperatures suggesting that protein folding is a limiting factor in low temperature growth (74).

Psychrophilic proteins may exist in a greater number of interchangeable conformational isomers than mesophilic proteins, with multiple states of very similar energy levels separated by small energy barriers (Figure 1.7) (41). Their folding landscape has a shallower energy minimum so it has been suggested that psychrophilic proteins are more susceptible to undesirable transient unfolding (97).

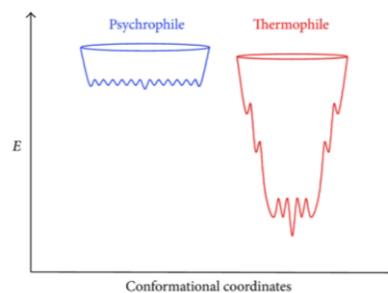


Figure 1.7: Folding energy landscapes hypothesised for psychrophiles and thermophiles. Taken from (16)

1.5 Applications of cold-adapted proteins

Understanding cold adaptation strategies at the molecular level has wide ranging biotechnological applications. Low temperatures are often used to inhibit bacterial growth in food preservation methods like refrigeration. The mechanisms by which pathogens such as *Listeria monocytogenes* adapt to these cold conditions could represent possible antimicrobial targets (98). With regard to farming, the knowledge gained from studying cold adaption could be important for modifying crops to protect them from cold conditions, especially in regard to nitrogen-fixing legumes (99).

In the biotechnology industry a major objective is to increase enzymatic reaction rates which currently is often achieved using high temperatures and thermostable enzymes. This however involves high energy consumption and high costs. Alternatively insights obtained into psychrophilic enzyme adaptations could be used to modify current enzymes to increase their activity at low temperatures. This would reduce energy costs, increase the ease of setting up reaction vessels and be more environmentally friendly (20)(100). Production of washing powder containing lipases and proteases which function at low temperatures represents a potential market for this strategy (79)(101). With regard to environmental issues, psychrophilic enzymes may be able to aid in the breakdown of pollutants in seawater such as hydrocarbons and phenols found in oil spillages, though results from this approach remain mixed so far (102). This may prove vital in protecting sensitive habitats within the Arctic and Antarctic if exploitation of oil resources is to occur. Another environmental use may be to optimise low-temperature cellulase enzymes to break down plant matter for biofuel generation (103). Being able to perform these reactions without the energy costs and practicalities of using high temperature reaction vessels may improve the economic viability of biofuel production (104).

Cold adapted enzymes could also help in the bio-catalysis of heat-labile substrates such as the process of removing lactose from milk or where heating would be undesirable such as chitinase enzymes to tenderise meat or xylanase to ripen cheeses (16)(105-106). This could also help in organic synthesis where degradation, side reactions or chirality issues may occur at higher temperatures (107). Psychrophiles possess an array of chaperone proteins to assist in protein folding at low temperatures which could be exploited to improve success rates in expressing proteins that regularly show mis-folding (74).

The lower thermostability of psychrophilic enzymes is usually disadvantageous however, one of the most commonly used psychrophilic enzymes, alkaline

phosphatase uses this as an advantage. The enzyme alkaline phosphatase removes the 5' phosphates from double stranded DNA to prevent cleaved DNA re-circularising (108). The presence of an active form of this enzyme can interfere with subsequent reactions but the low thermostability of psychrophilic alkaline phosphatase allows it to be denatured without damaging other products in the reaction mixture. Alkaline phosphatase from the Antarctic bacterium TAB5 is currently commercially expressed in *E. coli* cells (109).

Despite the great potential of psychrophilic proteins in industrial biotechnology, not enough is yet understood about how the thermostability and activity of proteins can be adapted to low temperature for effective rational design approaches to be easily conducted (19)(102). Enzymes generated through mutational studies often have issues with undesirable side reactions while use of extremophilic enzymes generally avoids this issue (101)

1.6 Cold shock proteins

1.6.1 Cold shock protein structure and function

At low temperatures transcription and translation rates in protein synthesis can be limiting factors for growth. One class of proteins with a crucial role in the adaptation of protein synthesis to temperature reduction is a family of highly conserved proteins called the cold shock proteins (Csps). They are small β -barrel proteins (58 aa to 73 aa) of the oligonucleotide/oligosaccharide-binding class of protein fold architectures which have been discovered across all kingdoms of life (67)(72)(110-112). As the name suggests Csps are substantially upregulated in response to a drop in environmental temperature. Low temperatures strengthen hydrogen bonds and the secondary structure stabilisation of nucleic acid bases becomes more problematic. Csps have been proposed to act as chaperones, binding to single stranded DNA and RNA molecules preventing formation of secondary structures which could inhibit the initiation of translation (98)(112). The structure of these 7.5

kDa proteins is comprised of five highly conserved antiparallel β -strands with some CspA also containing a single additional α -helix (67)(110-112). The structure of CspA from *E.coli* is shown in Figure 1.8 (113). These proteins contain two oligonucleotide binding motifs on strands $\beta 2$ and $\beta 3$ respectively which include aromatic residues thought to intercalate DNA or RNA (111)(113). Mutation of these residues to non-aromatic residues knocks-out both DNA-binding and Csp function (67).



Figure 1.8: Representative ribbon structure (PDB code 2L15) of the CspA from *E. coli* showing the five β -sheets wrapped around to form a β -barrel structure. Adapted from (113)

In some bacteria including *E. coli* and *B. subtilis*, CspA have been found to be the class of proteins displaying the greatest increase in expression in response to a short-term drop in temperature. Cellular expression of CspA can reach millimolar concentrations when *E. coli* is cooled from 37 °C to 10 °C comprising over 10% of total expressed protein (67). Some CspA are present at lower levels without a temperature decrease and could be involved in many other stress responses such as CspD in *E. coli* which is induced by a lack of nutrients (114-115). CspA only bind to single stranded nucleic acid chains however multiple CspA may bind DNA strands as short as 7 nucleotides long (116). CspA bind DNA containing a large proportion of thymidine residues with greater affinity as thymidine residues give better interaction with the aromatic residues in the binding site (116). The highest affinity DNA binding sequence was found to be a sequence of 6-7 thymine bases (117-118).

The first Csp to be discovered was the *E. coli* protein CspA and since then eight further families of cold shock proteins have been identified in *E. coli* designated

from CspB to CspI (98). Homologues of the *E. coli* Csp's have been discovered in psychrophiles, mesophiles, halophiles and thermophiles including the hyperthermophilic bacterium *Thermotoga maritima* (67). The prevalence of Csp proteins across all kingdoms of life suggests an early common ancestor. The Csp family of proteins represent a good model system for studying protein temperature adaptation as they are well studied, highly conserved and are present in mesophilic and extremophilic organisms across all kingdoms of life (67)(110-111).

1.6.2 The role of *E. coli* CspA mRNA as a thermosensor

Regulation of Csp expression is proposed to occur post-transcriptionally, mainly at the mRNA level. The mRNA of *E. coli* CspA exhibits two different secondary structures depending on temperature which have significantly different thermostabilities (119). These differences have been attributed to an unusually long 5' untranslated region present in Csp mRNAs (67)(120). The form of CspA mRNA present at normal growth temperatures in *E. coli* is highly unstable with a short half-life so does not build up in the cell (119). When the environmental temperature is decreased produced CspA mRNA adopts an alternate structure with a much longer half-life which allows far greater levels of translation to occur from each mRNA molecule before it is degraded (121). This means that significant production of the CspA protein only occurs at reduced temperatures allowing this system to serve as a thermosensor to detect temperature changes (120). The more stable low temperature CspA mRNA form can convert into the high temperature form with an increase in temperature but the high temperature form cannot convert into the low temperature form. This regulates the rate at which a cold-shock response can be initiated (67). *Jiang et al* investigated this system by over-expressing the 5' untranslated region of CspA (121). They found that this causes Csp expression to become constitutive rather than transient and gave a prolonged period of slower bacterial growth due to a prolonged cold-shock response.

1.6.3 The role of cold shock proteins in regulating protein synthesis

CspA from *E. coli* has been shown to regulate transcription of its own mRNA and hence could form a feedback loop regulating CspA expression (122). It has also been shown that CspA mRNA can also affect expression of other Csps (123). Csps are proposed to activate the promoters of cold adaption genes and may prevent the formation of unfavourable secondary structures in the mRNA of cold adaption genes hence regulating transcription rates (67)(115). Binding of Csps to mRNA could also block the degradation of mRNA by cellular RNases increasing expression of cold-adaption genes (78)(124). By these processes the CspA mRNA and Csp can both detect the change in temperature and induce a change in gene expression (see Figure 1.9).

A study by *Hofweber et al.* suggested that Csps may downregulate protein expression to reduce energy expenditure on protein production in times of stress (125). They found a general inhibition of *E. coli* protein expression in vitro with the extent of downregulation proportional to the Csp concentration present (125). This means Csps can serve as a metabolic brake for the cell in response to low temperature slowing the growth rate of *E. coli* (126). They also showed downregulation of translation of an mRNA encoding chloramphenicol acetyltransferase implying that the inhibitory effect could work on both transcription and translation providing a double method of inhibition (125). As Csps only bind ssDNA Csps would only bind DNA currently being transcribed. It is unclear how specific Csp binding is towards the promoter regions of certain genes or how much the protein downregulation is the product of random Csp binding affecting all genes. There are common patterns in regulation elements for cold adaption genes but connections between these and Csp binding has not been found so far (121).

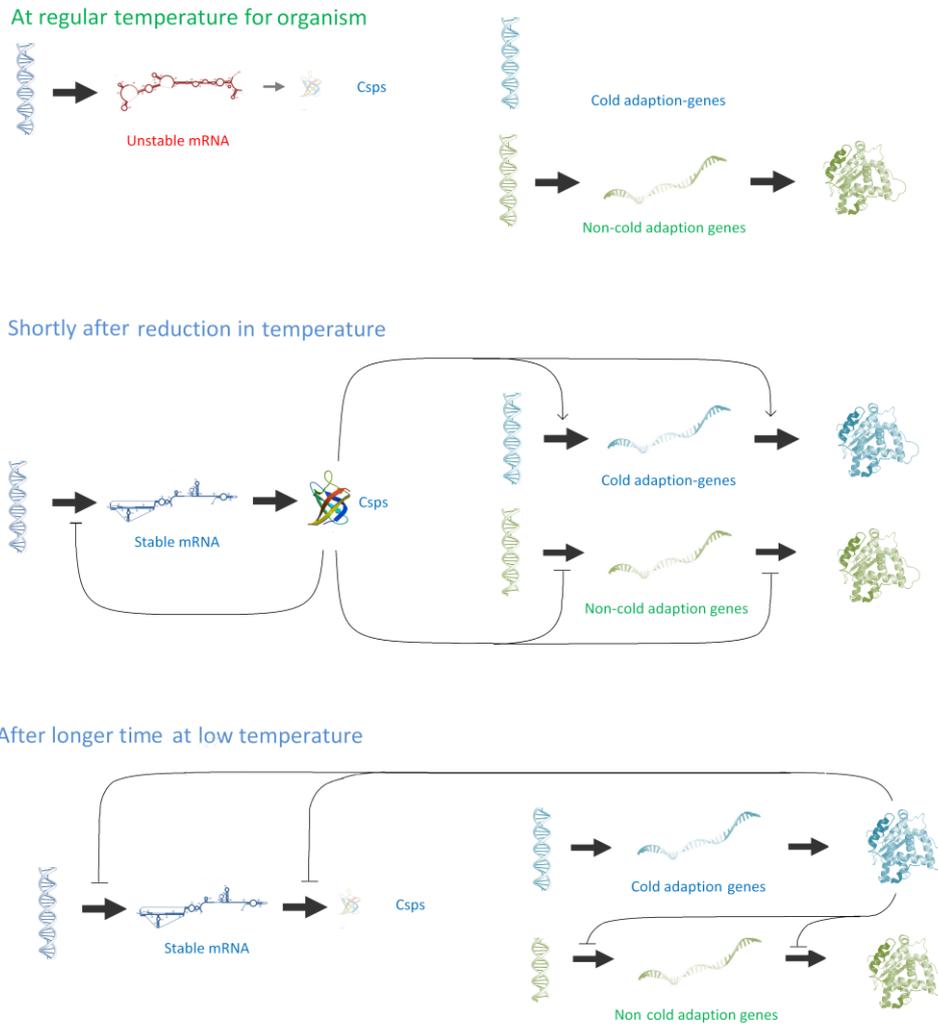


Figure 1.9: *E. coli* CspA mRNA as a temperature sensor. Top: Lack of Csp expression at moderate temperatures due to unstable mRNA conformation. Middle: Increase in Csp expression shortly after temperature is reduced and its proposed downstream effects on cold adaption genes, general genes not involved in cold adaption and Csp expression. Bottom: The cold shock response may be short lived as downstream effects of acclimatisation protein expression may lead to repression of Csp expression in longer periods of decreased temperature. Aspects adapted from (113) and (119).

Levels of DNA coiling have been identified as important in the regulation of gene activity and temperature sensitive alterations in DNA coiling may represent another method of cold adaption. Some studies suggest this to be a direct effect and others that it is mediated through upregulation of DNA gyrase enzymes (96). Temperature sensitive DNA supercoiling may act as a thermosensor and changes in supercoiling could activate some gene promoters (123).

Most cold adaptation genes are expressed at lower levels at moderate temperatures and upregulated at low temperatures (115). 26 genes are upregulated in response to cold shock in *E. coli* reduced from 25 °C to 10 °C (124). Nucleotide binding proteins in general show an increased prevalence within psychrophiles due to the stabilisation of mRNAs at low temperatures (74). In the psychrophilic bacterium *Psychrobacter cryopegella* RNA, protein and DNA synthesis rates were found to decrease exponentially with decreasing temperature (73). Many enzymes involved in energy metabolism were repressed at low temperatures however those involved in regulating the later stages of electron transport chain and filtering intermediates into tri-carboxylic acid cycle were upregulated to increase metabolic energy generation (127).

Csps are only transiently expressed with peak Csp expression at around one hour after the drop in temperature (121). Over a longer period of sustained low temperature expression of a second protein class termed cold acclimatisation proteins is upregulated and Csp concentrations decrease. These often exhibit the ability to alter DNA winding states such as helicase and gyrase functions (25). The exact nature of the divide between the two classes though is ambiguous (67). Some cold shock proteins from mesophiles have psychrophilic cold acclimatisation protein homologues (123). There are 20 characterised cold acclimatisation proteins found in *Rhizobia* and these have 40 - 60% homology to *E. coli* Csps. Cold-acclimatisation proteins have been suggested to mediate the decrease in Csp expression after a longer period of colds though factors which affect the way the ribosome reads Csp mRNAs have also been suggested to mediate this process (122)(128).

The function of Csps shows some redundancy so that deletion of some can be compensated for by the presence of others (56). Deletion of multiple Csps genes however can lead to failure of the cold shock response (122). It has proved possible to initiate a regular cold shock response in *E. coli* by substituting Csps from different organisms indicating a universality of Csp function (125-126).

1.6.4 Eukaryotic Csp homologues

Bacterial Csps incorporate only a single cold shock domain but in eukaryotes they are usually seen coupled to regulatory domains within proteins (110). The Csps of gram negative bacteria have been found to show high levels of sequence identity of around 43% to eukaryotic Y-box domains including a set of proteins in the major histocompatibility complex II (129). Eukaryotic Csps are able to travel between the nucleus and cytoplasm and nuclear localisation signals have been detected in some eukaryotic Csps. Over 6000 proteins have currently been discovered which contain the cold shock domain motif (130). The high homology observed shows that the cold shock domain has been highly conserved for the binding of nucleotides during evolution and is seen across archaea, eukaryotic and prokaryotic organisms suggesting an early common ancestor (67).

The role of Csps in regulating gene expression gives cold-shock and Y-box proteins a potential therapeutic application in the prevention of various diseases. A study by *Matsumoto et al.* showed over-expression of CspA could decrease the spread of cancer through NR1-SM cells within the lymph system of mice through inhibition of protein expression (131). A second study applied a contrasting approach using small interfering RNA to block expression of Csp genes which increased the susceptibility of cancer cells to temperature changes. This was found to disproportionately affect faster growing tumour cells (132). Csps have also been linked to regulation of the expression of growth factors and the phosphorylation state of Y-box domains linked to potential roles in tumour suppression (133). The reduction in protein expression mediated by Csps has also been suggested to have a protective role against aggregation mediated neurodegenerative diseases in the neurons of mice and may help preserve the connections between neurons (134-135).

1.7 Objectives

This investigation aims to study how the structures of psychrophilic proteins have evolved to address the changes in the relative strengths of intermolecular interactions that occur at low temperatures. Few studies have been performed into psychrophilic proteins so this research has the potential to provide fascinating new insights. As the differences between proteins adapted to different temperatures involve relatively small numbers of changes then comparing difference in sequences and thermostabilities could reveal the strategies that cold-adapted organisms use to balance protein dynamics and thermostability.

This investigation focuses on members of the cold shock protein (Csp) family. First the sequences of different members of the Csp are compared to see to what extent they differ and what patterns are present in those adapted to high and low temperatures. The relative stabilities of the different Csps are assessed to observe the extent to which the stabilities of the Csps to thermal denaturation, chemical denaturants and mechanical forces vary between those Csps expressed in organisms living at different temperatures. A difficulty with previous comparative studies has been variation in the techniques and buffers used between homologues but all homologues in this study are investigated using the same techniques and buffers. As extremophilic proteins are optimised to different temperatures it is investigated how the stabilities vary with temperature and what the thermostability levels are at the optimal temperature of the host organisms.

No previous studies have compared the mechanical stability of cold-adapted proteins or evaluated comparisons of different temperature adapted variants but this may be achievable using the single molecule technique of AFM. Such a procedure requires the generation of polyproteins. A clear mechanical signature has been previously observed for a hyperthermophilic Csp variant so it will be confirmed if a similar signature is present for the psychrotrophic variant (136-137). To understand more about the energy barriers Monte Carlo simulations are applied

to the AFM data. Stopped flow experiments are conducted to determine the differences in folding and unfolding rates between the Csp's to examine differences in their energy landscapes. There is also a section detailing a preliminary characterisation of how Csp dynamics varies with temperature.

These studies will help identify the extent to which cold-adapted proteins maintain their native states but still maintain their function. A reasonable amount of literature has been published about how extreme hot-adapted (hyperthermophilic) proteins are adapted. This study tries to identify more about the methods that extreme cold-adapted proteins may apply to achieve thermostability and activity at very low temperatures which is an area with little published data.

2. Materials and Methods

2.1 Materials

Deionised water was used in all protocols filtered to 18 M Ω using a Purite system. Agar was purchased from Melford, UK. LB medium in powder form and BugBuster master mix were purchased from Merck, Germany. Antibiotic stocks of carbenicillin and chloramphenicol were purchased from Sigma Life Sciences, USA. Ethidium bromide, bromophenol blue, magnesium chloride, calcium chloride, DNase, RNase, ammonium acetate, ammonium persulphate and TEMED were purchased from Sigma Life Sciences, USA. IPTG, DTT and auto induction media powder were purchased from Formedium, UK. Agarose and imidazole were purchased from Acros Organics, Belgium. Vercon powder, Tris, agar, sodium phosphate, sodium chloride, potassium phosphate, acetic acid, glycerol and SDS were purchased from Fisher Scientific, USA. 30% (w/v) bisacrylamide was purchased from Severn Biotech. Protease inhibitor pills were purchased from thermo scientific, USA. Protein markers were purchased from Invitrogen, UK or Bio-Rad, USA. Nickel sepharose resin was purchased from GE healthcare, Sweden. Instant blue stain was purchased from Expedeon, UK.

All restriction enzymes, alkaline phosphatase, sticky-end ligase master mix, Gibson assembly master-mix, DNA polymerase and associated buffers were purchased from New England Biolabs, USA. The DNA markers for agarose gels, 6 \times loading buffer, dNTPs and GoTaq polymerase were purchased from Promega, USA. All primers and synthesised cold shock protein genes were purchased from Eurofins. Plasmid pET3a was obtained from Dr Toni Hoffmann containing *Xho*I and *Mlu*I restriction sites for insertion of desired Csp sequences. pET3a plasmids containing (I27-*TmCsp*)₃-I27 were obtained from Dr Toni Hoffmann for the polyprotein methods. All pET3a plasmids contained an N-terminal (His)₆ affinity tag for purification, an IPTG inducible lac operator and a T7 viral promoter sequence. In addition all polyprotein sequences contain two cysteine at the C-terminus for AFM studies.

Cuvettes were purchased from Sigma Life Sciences, USA. Snakeskin tubing was purchased from Thermo Scientific, USA. Filters were purchased from Millipore, UK. Mini-Prep, Midi-Prep and Gel extraction kits were purchased from Qiagen, Germany or Promega, USA. AFM cantilevers were purchased from Bruker, USA.

2.1.1 Cell culture medium and *E. coli* strains

All media types used were sterilised using a 1 hour autoclave cycle with heating to a temperature of 120 °C for 20 minutes at 15 psi. Sterile techniques were used in transfers of media. Lysogeny broth was obtained as a powder but all other media were made up from individual components. For minimal media see section 2.5

Table 2.1: Composition of bacterial growth media used. Remaining volume equates to addition of deionised water. These numbers presented are per litre but amounts added were scaled to required volume of media. See specific methods for inclusion or absence of antibiotic selection markers.

Media	Component	Amount g/l
lysogeny broth (LB)	bactotryptone	10
	Yeast extract	5
	NaCl	10
2×TY broth	bactotryptone	16
	Yeast extract	10
	NaCl	5
Super optimal broth with catabolite repression (SOC)	bactotryptone	20
	Yeast extract	5
	NaCl	10
	KCl	0.186
	MgSO ₄	2.467

Cloning Strains - *E. coli* XL1-Blue, SURE2 and DH5α

Expression strains - *E. coli* JM83, BL21(DE3) pLysS and BLR (DE3) pLysS

2.2 Molecular biology methods

2.2.1 Polymerase chain reaction

Primers were designed using the Invitrogen OligoPerfect design tool.

Table 2.2: PCR primers utilised in PCR DNA amplifications. Gibson assembly primers indicated separately

Csp primers	Length (bp)	DNA sequence 5' to 3'
<i>Psycrobacter</i> PB6 forward	27	GATCCTCGAGCAGCGATAAAGTGAAG
<i>Psycrobacter</i> PB6 reverse	28	GCATACGCGTTTAAATCGCTTCAATCTG
<i>Shewanella</i> forward	28	GATCCTCGAGCAACAAAACCCGGCCT
<i>Shewanella</i> reverse	28	GCATACGCGTTTACACCGCCACCACGTT
<i>Thermotoga</i> cassette 1 forward	22	ATCACCATCACCATCACTCGAG
<i>Thermotoga</i> cassette 1 reverse	38	GTGATGGTGACTAGTGTTTCCACCACTTTCACATGCGC
<i>Thermotoga</i> cassette 5 forward	39	CATCACCATGGGCCCGCCGCGCAAAGTGAAATGGTTTG
<i>Thermotoga</i> cassette 5 reverse	46	ATGGTGATGGACGCGTTTAAACAACATTCCACCACTTTCACATGCGCC

PCR mixtures were set up on ice as detailed in Table 5.2 using Phusion or GoTaq DNA polymerase and subjected to thermal cycling in a Biorad thermocycler as described in Table 5.3. PCR mixtures were performed in triplicate to generate sufficient DNA for further study. The temperature for the annealing step was typically 5 °C below the T_m value for the primers

Table 2.3: Composition of PCR mixtures performed in triplicate.

Component	Volume (µl)	Final concentration
Sterile water	41.5	-
5 × Colourless buffer	10	1×
dNTP mix (dATP,dCTP,dTTP,dGTP)	1	0.2 mM each dNTP
Upstream primer	1	0.2 µM
Downstream primer	1	0.2 µM
Template strand	0.1	-
DNA polymerase	0.4	1.25 µM

Table 2.4: Thermal cycling conditions for PCR amplification.

Stage	Temperature (°C)	Time	Cycles
Denaturation	98	2 minutes	1
Denaturation	98	15 seconds	} 28
Annealing	58-62	30 seconds	
Extension	72	20 seconds	
Final Extension	72	10 minutes	1
Hold	4	-	1

2.2.2 Agarose gel electrophoresis

A 50× stock solution of TAE buffer was prepared with 121 g Tris, 28.55 ml/l glacial acetic acid, 50 ml 0.5 M EDTA (pH 8.0). Agarose was dissolved into 100 ml 1× TAE buffer in a 250 ml conical flask by heat. A 1% (w/v) agarose gel was used for analysis of DNA samples over 1 kbp in size and 2% (w/v) used where the analysed DNA was smaller than 1 kbp. The solution was allowed to cool to around 50 °C and poured into a gel tray containing a comb and sides sealed by metal inserts and 6 µl of 10 mg/ml ethidium bromide mixed in to the solution thoroughly with a pipette until the red colour was no longer visible. The set gels were placed in an electrophoresis tank and flooded with 1x TAE buffer so the buffer covered the gel.

DNA samples were prepared for loading by the addition of 6× gel loading buffer (0.25% (w/v) 0.25% (w/v) xylene cyanol, 40% (w/v) sucrose in H₂O) in a ratio of 5:1 DNA to loading buffer and loaded into wells. 1 kbp and 100 bp markers were loaded in end lanes to allow DNA sizes to be quantified. The gel was run for 45 minutes at 90 V to allow for sufficient DNA size separation. Imaging of agarose gels was performed using a Syngene U.V. transilluminator and the relevant illuminated bands excised using a scalpel. Extracted gel slices were added to pre-weighed 2 ml Eppendorf tube to deduce the mass of the gel slice. The DNA contained within each slice was extracted into 40 µl of tris buffer using a Qiagen gel extraction kit (250) following manufacturer's instructions. The concentration of DNA was determined spectroscopically by A₂₆₀ value. A 4 µl sample of the extracted DNA was diluted 1:11

with 44 μ l tris buffer in a 1 cm path length U.V.-transparent plastic cuvette. The cuvette was placed in a U.V.-Vis spectroscope and absorbance readings at 230 nm, 280 nm and 320nm used to establish a baseline. The A_{260} value with baseline subtracted is used to determine the concentration of DNA with an A_{260} value of 1 equal to 50 g/mL of double stranded DNA.

2.2.3 Restriction endonuclease digestion

Table 2.5: Composition of a typical restriction digest mixture used in this investigation. Restriction digestion mixes were set up on ice.

Component	Volume used (μ l)
DNA (PCR or plasmid from mini/midi-prep)	36
10 \times restriction enzyme buffer NEB4	4.3 (DNA volume/9)
100 \times Bovine serum albumin solution (10 mg/ml)	0.43 (DNA volume/90)
Upstream restriction enzyme (10,000-50,000 units/ml)	1
Downstream restriction enzyme (10,000-50,000 units/ml)	1

Samples were maintained on ice while components were added and after gentle mixing, transferred to a 37 $^{\circ}$ C shaking incubator at 200 rpm for 2 hours. If *Apal* was one of the restriction enzymes the first hour of incubation was performed at 25 $^{\circ}$ C before transfer to 37 $^{\circ}$ C for the second hour. Where the digested DNA was to serve as a vector for ligation of a DNA insert a 1:9 ratio of phosphatase buffer was added to the reaction mixture along with 1 μ l of Antarctic phosphatase (5000 units/ml, NEB). The mixture was incubated for a further 30 minutes at 37 $^{\circ}$ C before enzyme inactivation utilising a heating block at 65 $^{\circ}$ C for 5 minutes. The digested DNA was then analysed using agarose gel electrophoresis or purified from the enzymes using a Qiagen gel extraction kit. All DNA was stored at -20 $^{\circ}$ C in a freezer.

2.2.4 Preparing LB agar plates

A solution was prepared containing 25 g LB medium and 15 g LB agar diluted into 1l of water. The solution was autoclaved at 120 $^{\circ}$ C for 20 minutes at 15 psi and

allowed to cool to around 40 °C before addition of 100 mg/ml carbenicillin to a final concentration of 100 µg/ml. For higher stringency plates chloramphenicol (25 µg/ml) and 1% glucose solution were also added. The solution was poured into petri dishes under sterile conditions and plates left for 2 hours for the LB agar to fully solidify.

2.2.5 Preparing competent cells

Relevant bacterial strains were grown in 100 ml LB without antibiotic to an optical density (OD₆₀₀) of 0.4. Cells were pelleted for 10 minutes in a benchtop centrifuge at 4,000 rpm at 4 °C, the supernatant removed and the pellet resuspended in 10 ml 100 mM pre-chilled CaCl₂. After 10 minutes on ice the cells were again pelleted and resuspended in 100 mM sterile CaCl₂ with 30% glycerol. The cells were separated into 50 ml aliquots and stored at -80 °C.

2.2.6 DNA ligation into pET vectors

DNA ligation was performed using a molar ratio of 3-5:1 of insert to plasmid in a total volume of 5 µl. This was mixed with an equal volume of NEB instant sticky-end ligase master mix. As a negative control, the same plasmid was used but the insert was substituted with an equivalent volume of sterile water.

2.2.7 Transformation of pET vectors into *E. coli* cells

Competent cells were taken from freezer storage at -80 °C, thawed on ice and divided into 50 µl aliquots in 1.5 ml microcentrifuge tubes. For repetitive sequences SURE2 cells were used with addition of 1.7 µl of 1.22 M β-mercaptoethanol to each aliquot to improve transformation efficiency. 1.5 µl of ligation mix was added to each aliquot and swirled while an equivalent 1.5 µl of water was added to a separate aliquot of cells as a negative control. Samples were incubated on ice for 30

minutes prior to a heat-shock in a 42 °C water bath for 30 or 45 seconds depending on the strain of *E. coli* being used. Tubes were returned to ice for 2 minutes before addition of 750 µl of LB medium. The tubes were placed in a shaking incubator for 1 hour at 37 °C. For each sample 100 µl was spread on to pre-heated LB agar plates supplemented with 100 µg/ml carbenicillin. The remaining solution was spun in a benchtop centrifuge for 10 minutes at 1000 rpm and 500 µl removed to further increase cell concentration and a further 100 µl spread. Plates were wrapped with parafilm and incubated overnight at 37 °C. Successful transformation of the plasmid DNA resulted in observable colonies on the LB agar plate and none or far fewer on the negative control.

2.2.8 Plasmid DNA extraction

Three colonies were picked using a pipette tip and added to 10 ml LB media supplemented with 100 µg/ml carbenicillin and grown overnight at 37 °C in a 200 rpm shaking incubator. Plasmid DNA was extracted using a Qiagen QIAprep or Promega Wizard SV mini-prep kit using manufacturer's instructions. The concentration of extracted DNA was ascertained using U.V.-Vis spectroscopy of a 1 in 12 dilution of 4 µl of the sample into water. For sequencing a 15 µl DNA sample was sent off to Beckman Coulter Genomics with the T7 viral promoter sequence used for the primer. Returned sequences were analysed using sequence manipulation suite bioinformatics tools. If greater quantities of DNA were required for experiments, 100 ml overnight cultures and a Qiagen or Promega midi prep kit were used as per manufacturer's instructions.

2.2.9 Gibson assembly reaction

A linearized pET3a plasmid containing His₆-(I27) and I27 domains for positions 3, 5 and 7 were obtained from Dr Toni Hoffmann for the polyprotein construction. Insert DNA cassettes were extended with primers by PCR (section 2.2.1) to add unique flanking sequences to provide each cassette with specificity for its location within

the Gibson construct. 25 ng of pET3a-I27 vector was combined with an approximately 5 fold molar excess of each of the 3 Csp and 3 I27 inserts in a total volume of 10 μ l. This was combined with a 1:1 ratio of Gibson master mix.

Table 2.6: Primers used for the Gibson assembly reaction synthesis of (I27-PB6 Csp)₃-I27

Csp primers	Length (bp)	DNA sequence
Position 2 PB6 forward	42	CTGAGCGTGGGCGCGACCATTAGCGATAAAGTGGAAGGCACC
Position 2 PB6 reverse	42	GCTCGCCAGACCAATAACGGTAATCGCTTCAATCTGTTCCGC
Position 4 PB6 forward	42	GCGCTGAGCGGCACCATTGTGAGCGATAAAGTGGAAGGCACC
Position 4 PB6 reverse	42	CGCCAGGCTACCGGTAATAACAATCGCTTCAATCTGTTCCGC
Position 6 PB6 forward	42	AGCGCGCTGGGCATTGTGACCAGCGATAAAGTGGAAGGCACC
Position 6 PB6 reverse	42	CAGGCTCACACCTGCGGTAATAATCGCTTCAATCTGTTCCGC

The Gibson reaction was conducted at 50 °C for 1 hour. This mix was used to transform 50 μ l of NEB cells (section 2.2.7) with super optimal broth with catabolite repression (SOC) medium substituted for LB to improve cell yield. The cells were plated out on to LB agar plates supplemented with carbenicillin and incubated overnight.

2.2.10 Colony PCR

Sterile pipette tips were used to pick each of 24 colonies from the LB agar plates. Each tip was swirled in 100 μ l of distilled water and a reference streak made out on to an LB agar plate with 100 μ g/ml carbenicillin selection antibiotic. The water solutions were vortexed to aid resuspension of cells, heated for 4 minutes in a heating block and centrifuged at 13,300 rpm in a benchtop centrifuge for 12 minutes. A PCR master mix was set up containing forward and reverse primers corresponding to the upstream and downstream extents of the insertion region, dNTPs, Q5 polymerase and Q5 polymerase buffer. 13.4 μ l of each colony resuspension was added to 6.6 μ l of the master mix and samples put through 28

PCR cycles (see section 2.2.1). The resulting solutions were combined with loading buffer and run on an agarose gel (see section 2.2.2). Plasmid DNA of the colonies containing inserts of the correct size was extracted and purified (section 2.28) and sent to Beckman Coulter Genomics for sequencing.

2.3 Expression of proteins in *E. coli*

2.3.1 DNA transformation into expression cells

BL21 [DE3] pLysS, BLR [DE3] pLysS or JM83 cells were taken from storage in a -80 °C freezer and thawed on ice. 1 µl of DNA was added to each sample with an equivalent volume of water used for the negative control. Transformation was conducted as per section 2.2.7 with 10 µl and 100 µl samples used to plate on to LB agar plates supplemented with carbenicillin antibiotic.

Table 2.7: Plasmid and cell types used over the course of the study

Protein	Use	Plasmid	<i>E. coli</i> Cell strain
Csp monomers	Molecular biology	pEX into pET3a	DH5α
	Expression	pET3a	BL21 (DE3) pLysS
Polyproteins	Molecular biology	pET3a	SURE-2
(I27-PB6 Csp) ₃ -I27	Gibson assembly	pET3a	DH5α (NEB)
	Expression	pET3a	BLR (DE3) pLysS
MBP-PB6 Csp	Molecular biology	pMAL C5X	XL1-Blue
	Expression	pMAL C5X	BL21 (DE3) pLysS

2.3.2 Small scale expression trials

Single colonies were picked from the transformed plates and added to 10 ml of LB broth containing 100 µg/ml carbenicillin for overnight incubation with 200 rpm shaking at 37 °C. After 18 hours the OD₆₀₀ of a 1:4 dilution of the cell culture was recorded in a U.V.-Vis spectrophotometer against a zero reference of LB alone. This

was used to calculate the volume of culture required to be added to a solution containing 30 ml LB and 100 µg/ml carbenicillin to give an OD₆₀₀ of 0.1. Cells were grown at 26 °C until an OD₆₀₀ of 0.6-0.8 and IPTG added to a final concentration of 1 mM to induce protein expression (138-139). The OD₆₀₀ of a 1ml sample was recorded each hour for 3 hours and for one overnight sample. After measure of each OD₆₀₀ value samples were centrifuged at 13,000 rpm in a benchtop centrifuge for 1 minute. The supernatant was removed and the pellet stored in a -20 °C freezer.

Where the protein content of the whole cell lysate was to be analysed by SDS-PAGE, samples were resuspended with lysis buffer (20 mM Tris, 300 mM NaCl, 0.025 (w/v) sodium azide, 0.15% (v/v) triton X100, 5 mM imidazole pH8). A volume of 20 µl lysis buffer was added to the pre-induction sample and the volume for the other samples scaled in proportion to their respective OD₆₀₀ values relative to the pre-induction value (OD_{600pre}) as per equation (2.1) to account for the different cell densities. Each sample was vortex mixed every 5 minutes for 20 minutes.

$$\text{Eqn 2.1: } V_{\text{lysis}} = \left(\frac{\text{OD}_{600}}{\text{OD}_{600\text{pre}}} \right) \times 20 \mu\text{l}$$

2.3.3 SDS-PAGE

SDS-PAGE gels were used to confirm protein overexpression and to determine the purity of protein samples after purification steps. A plastic gasket was added around 3 edges of an ATTO glass plate and a second glass plate pushed up against the gasket with the aid of clips to secure the positioning. Unpolymerised gel solutions were made as detailed in Table 2.8 with TEMED the final component to be added. Upon addition of TEMED to the resolving gel the tube was inverted repeatedly to thoroughly mix the components and the solution poured between the glass plates up to three quarters of the glass plate height. The remaining volume was filled by pipetting of stacking gel over the resolving gel and a comb added to generate wells. Boundaries between the gel types are maintained by the greater density of glycerol

in the resolving gel. Gels were left for around 5 minutes to set. For overnight storage the glass plates were wrapped with tissue, soaked with deionised water and sealed in a zip-lock bag at 4 °C.

Table 2.8: The compositions of the two sections of an SDS-PAGE gel. The volumes shown are sufficient for making 2 gels.

Component	Resolving gel volume (ml)	Stacking gel volume (ml)
30% (w/v) Acrylamide; 0.8% (w/v) bis-acrylamide	5.1	0.67
SDS-buffer: 3 M Tris.HCl, 0.3% (w/v) SDS pH 8.45	5.0	1.24
Water	3.22	3.05
80% (v/v) Glycerol	1.6	0.0
10% (w/v) Ammonium persulphate	0.16	0.08
Tetramethylethylenediamine (TEMED)	0.015	0.008

The gels were run using a Tris-Tricine buffer with the reservoir between the gels filled with cathode buffer (100 mM Tris.HCl, 100 mM Tricine, 0.1% (w/v) SDS, pH 8.3) and the outer reservoir filled with anode buffer (400 mM Tris.HCl, pH 8.8). Prior to loading samples were diluted 3:1 with loading dye (50 mM Tris.HCl pH 6.8, 100 mM DTT, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol, heated using a 100 °C heating block for 3 minutes and centrifuged for 2 minutes at 13,300 rpm to remove insoluble components. 10-15 µl of sample was loaded into each well flanked by lanes containing 3 µl and 10 µl of protein markers as a size comparison standard. Gels were run at a current of 30 mA initially as the samples migrated through the stacking gel before being raised to 60 mA until the solvent front reached the bottom of the gel. Gels were removed from the glass plates and immersed into a solution of Coomassie blue (Generon, USA) or Instantblue (Expedeon, USA) stain for at least 30 minutes. The stain was replaced with water for 10 minutes before imaging using a white light transilluminator or scanner.

2.3.4 Small scale Nickel-NTA purification

Cell pellets generated by the small scale protein expression (section 2.3.2) were resuspended in 1 ml BugBuster master mix for 20 minutes (or until no pellet was visible) with vortexing every 5 minutes. The sample was spun at 13,300 rpm for 10 minutes to remove insoluble components. 30 μ l of Nickel NTA-sepharose in 20% ethanol (GE Healthcare, Sweden) was added to a 1.5 ml tube and mixed with 500 μ l wash buffer (20 mM Tris, 300 mM NaCl, 0.025% (w/v) sodium azide). This was centrifuged at 4,000 rpm for 1 minute and the supernatant removed. The supernatant from the cleared cell lysate was added to the pelleted nickel beads, mixed and left for 15 minutes. The culture was centrifuged at 4,000 rpm for 1 minute and the supernatant removed. The pellet was resuspended with 300 μ l wash buffer, centrifuged for 2 minutes at 4,000 rpm in a benchtop centrifuge and the supernatant removed. This wash step was repeated 6 times to elute proteins which had not bound specifically to the beads. To elute the remaining protein from the beads 100 μ l of elution buffer (20 mM Tris, 300 mM NaCl, 0.0025 (w/v) sodium azide, 250 mM imidazole pH8) was added, the bead pellet resuspended, centrifuged for 2 minutes at 6,000 rpm and supernatant kept. This was carried out twice to ensure all bound protein was eluted. The resulting samples including flow through, elute and sample wash steps were run on SDS-PAGE by the method shown above.

2.3.5 Large scale expression of His-tagged vectors

A single colony of expression cells transformed with the plasmid was picked from an LB agar plate and used to inoculate 150 ml of LB supplemented with 100 μ g/ml carbenicillin. The culture was incubated overnight at 37 °C with shaking at 200 rpm. The following morning this was used to inoculate a series of 10 preheated sterilised 2 l flasks containing 500 ml LB and 100 μ g/ μ l carbenicillin to an OD₆₀₀ of 0.07 - 0.10. Cultures were grown at 26 °C with shaking at 2000 rpm until an OD₆₀₀ of 0.6 - 0.8 was reached at which point induction was performed with IPTG at a final concentration of 1mM. Cultures were grown for a further 18 hours before harvesting in a Beckman Coulter Avanti centrifuge with a JLA 8.1 rotor. Cells were

spun at 5,000 rpm for 30 minutes at 4 °C and pellets collected for storage at -20 °C. All glass flasks were disinfected with Vircon and centrifuge tubes with Trigene. For improved yield LB can be substituted with 2TY media (Table 2.1).

2.4 Purification of His-tagged proteins

2.4.1 Ni-NTA affinity purification of His-tagged proteins

The stages described in section 2.4 were all performed within the space of 5 days to avoid protein degradation. Cell pellets were resuspended in lysis buffer (20 mM Tris, 300 mM NaCl, 0.025 (w/v) sodium azide, 0.15% (v/v) triton X100, 5 mM imidazole pH8) to a volume of 40 ml. PMSF and benzamidine were added to final concentrations of 1 mM and 2 mM respectively along with an EDTA-free protease inhibitor tablet. A flake of RNase and DNase was added to the cell suspension, followed by homogenisation using a glass homogeniser. Cells were broken open using a cell disruptor at a pressure of 25 kpsi. Large insoluble components were removed through a 30 min centrifugation at 15,000 rpm in a Beckman Coulter Avanti centrifuge with a JLA 25.50 rotor and the cleared lysate passed through a 0.45 µm filter.

The first purification step was conducted using a 5 ml Histrap FF Nickel-NTA affinity column (GE Healthcare, Sweden) and an AKTA prime system. All buffers used in this protocol were filtered and degassed using a vacuum pump. The lines and the column were run with water and then the relevant buffer. The Ni-NTA column was washed with water and then 30 ml lysis buffer (20 mM Tris, 300 mM NaCl, 0.025% (w/v) sodium azide, 0.15% (v/v) triton X100, 5 mM imidazole pH 8). The sample was loaded onto the column via an inlet line with additional lysis buffer added during the loading process to give a loaded volume of 50 ml. Following sample loading 60 ml of wash buffer (20 mM Tris, 300 mM NaCl, 0.025% (w/v) sodium azide, 10 mM imidazole pH 8.0) buffer was run through the column. Protein remaining bound to the Ni-NTA column was eluted with using a stepped gradient from 0 - 100% of

elution buffer (20 mM Tris, 300 mM NaCl, 0.025% (w/v) sodium azide, 250 mM imidazole pH 8.0) in 25% increments with 25 ml volume for each step. Fractions of 3 ml were collected over the gradient and the absorbance at 280 nm of the eluted solution recorded automatically throughout to select fractions to be pooled for further analysis.

2.4.2 DNA removal by ion-exchange chromatography

A Viva Spin concentrator with 3 kDa membrane filter was centrifuged at 4000 rpm in a bucket centrifuge to reduce the sample volume to 2 ml. This was then combined with 18 ml low salt buffer (20 mM sodium phosphate, 50 mM NaCl pH8) and the process repeated. The sample was reduced to a volume of 3.5 ml which was loaded via a 5 ml loop on to Q-Trap column (GE Healthcare, Sweden). 40 ml of low-salt buffer was passed through the column with fractions collected and after 40 ml the NaCl gradient raised linearly from 50 mM to 1 M NaCl over 160 ml by mixing of low salt (20 mM sodium phosphate, 50 mM NaCl pH 8) and high salt (20 mM sodium phosphate, 1 M NaCl pH 8) buffers. A final 40 ml of high salt buffer was applied to the column to elute any DNA remaining bound to the column. Collected fractions were analysed by SDS-PAGE as described in section 2.3.3 and relevant fractions pooled.

2.4.3 Size-exclusion chromatography

A Sephadex analytical 75 column with a bed volume of 24 ml (GE Healthcare, Sweden) was equilibrated with water then buffer (25 mM Tris.HCl, 2 mM DTT, 300 mM NaCl pH8). The volume of pooled fractions was reduced to 3.5 ml using a fresh concentrating Viva spin column. Protein samples of 3.5 ml were loaded on to the column via a 5 ml loop. 3 ml fractions were collected and SDS-PAGE used to identify relevant fractions.

2.4.4 Dialysis and freeze drying

Those fractions found to contain the desired protein were pooled for dialysis. Samples were not left more than a day at this stage to prevent protein precipitation. A length of snakeskin dialysis tubing was soaked in deionised water and a knot tied in one end. The sample solution was loaded into the tubing with a further knot tied at the other end and a clip applied at either end. Samples were dialysed for two periods of 2 - 3 hours at 4 °C with stirring using 5 litres of distilled water with 2.5 mM ammonium bicarbonate solution. This was followed by a further two 2 - 3 hr periods of dialysis in ultrapure (milli-Q) water. The U.V. absorbance of samples was recorded to calculate protein concentration via the Beer-Lambert law where A is recorded absorbance, l is cuvette path length and ϵ is the molar extinction coefficient of the protein

$$\text{Eqn 2.2: } A = \epsilon Cl$$

Samples were aliquoted and snap froze in a mixture of dry ice and ethanol before being lyophilised using a Heto Power p23000 freeze drier (Thermo scientific) over 2-3 days. For the first purification of each specific protein, 0.5 mg of protein was dialysed into 20 mM sodium acetate and mass determination performed using ESI-MS (electrospray ionisation mass spectrometry) by Dr James Ault to verify production of the correct protein.

2.5 Production of ¹⁵N labelled proteins for NMR

2.5.1 Making minimal media

The initial step of making minimal media was production of M9 salts as detailed in Table 2.9 (139). M9 salts contain the only source of nitrogen in the media. Where the protein sample was intended for NMR studies ¹⁵N labelled ammonium chloride was used. After mixing M9 salts were autoclaved at 121 °C for 20 minutes at 15 psi.

Table 2.9: Components for M9 minimal salts. Remaining volume made up with deionised water. Minimal media salts were autoclaved at 121 °C for 20 minutes at 15psi after mixing. Ammonium chloride included labelled nitrogen ¹⁵N. 1 Litre of M9 salts is sufficient for 5 litres of minimal medium.

Component	Amount added per litre M9 salts
Na ₂ HPO ₄ ·7H ₂ O	64 g
KH ₂ PO ₄	15 g
NaCl	2.5 g
¹⁵ NH ₄ Cl	5 g

100 ml of M9 salts was added to each of 10, 2L flasks. 20 ml of MgSO₄ and CaCl₂ were resuspended to a 1M stock and filter sterilised before being added to the M9 media (Table 2.9). The solutions were made up to 500 ml with deionised water and autoclaved at 121 °C for 20 minutes at 15 psi. Glucose and carbenicillin were added after autoclaving as detailed in Table 2.10. To aid solubilisation of glucose it was initially resuspended to a 20 % (w/v) solution in 100 ml of deionised water preheated to 40 °C.

Table 2.10: Components for M9 minimal media. Stock solutions of MgSO₄ and CaCl₂ were filter sterilised using a 0.45 µl filter.

Stock solution added	Final concentration	
5 x M9 salts	100 ml/500 ml (1×)	Solution sterilised autoclaved after these added
1 M MgSO ₄	2 mM	
1 M CaCl ₂	0.1 mM	
20% Glucose	4% (w/v)	Added Post-autoclave sterilisation
100mg/ml Carbenicillin	100 µg/ml	

E. coli BL21 (DE3) PLysS cells were transformed with the desired plasmid and colonies inoculated into 15 ml LB media with 100 µg/ml carbenicillin and incubated

overnight with shaking at 37 °C. 3 ml of this culture was seeded into two 200 ml M9 minimal medium cultures and grown overnight under the same conditions (138). This culture was used to inoculate a series of 2 litre glass flasks each containing 500 ml M9 media with antibiotic selection marker (Table 2.1). The culture was grown to an OD₆₀₀ of 0.7 before being induced with IPTG to a final concentration of 1 mM. Cultures were harvested the following morning by centrifugation at 5000 rpm using a Beckman Coulter Avanti centrifuge with a JLA 8.1 rotor. Any protein kept in solution for more than 1 week had 0.02% (w/v) sodium azide added for storage.

2.5.2 Cleavage of PB6 Csp from MBP-PB6 Csp fusion protein

The MBP-PB6 fusion protein was expressed and purified (as per section 2.4). The purified protein was concentrated down to 4 ml in 25 mM Tris.HCl, 2 mM DTT, 300 mM NaCl pH 8 buffer and then was incubated with an excess (0.1 mg) of TEV protease for 20 min at room temperature and overnight at 4 °C. Complete digestion was confirmed with an SDS-PAGE gel. The solution was loaded into a Sup75 size-exclusion column as per section 2.4.3 and the identity of protein peaks analysed with SDS-PAGE before pooling, dialysis and freeze drying.

2.6 Multiple sequence alignment

Sequences were aligned in the programme Jalview (140) and alignment performed using clustalW alignment scores. The colour coding was then performed by BLOSSUM20 score.

2.7 Thermal and chemical stability assays

2.7.1 Sodium phosphate buffer

To generate 63 mM pH7.4 sodium phosphate buffer, stock solutions of sodium monophosphate (0.2 M) and sodium diphosphate (0.2 M) were mixed in a ratio of 1.9 monophosphate to 8.1 diphosphate as determined by the Henderson-Hasselbach equation (equation 2.3) where HA is the conjugate acid and A⁻ the

conjugate base. The pH was checked using a pH meter and the solution diluted with deionised water in a 1 : 2.17 ratio to reach the correct concentration.

$$\text{Eqn 2.3: } \text{pH} = \text{pKa} + \frac{\text{Log [A-]}}{[\text{HA}]}$$

2.7.2 Circular dichroism

Lyophilised protein samples were resuspended in 63 mM sodium phosphate buffer pH 7.4, filtered through a 0.45 μm filter then concentration adjusted by U.V.-absorbance to 0.3 mg/ml. CD spectra were recorded using a Chirascan CD instrument (Applied Photophysics, UK) with 300 ml protein sample in a 10 mm pathlength cuvette (141). Folding status was assessed by measurement of ellipticity spectra between 190 nm and 280 nm at 4 $^{\circ}\text{C}$ with a 1 nm bandwidth at a rate of 4 s/nm averaged over 3 repeats. Unfolding and refolding of protein samples was assessed by heating samples to 85 $^{\circ}\text{C}$ for 15 minutes before cooling back to 4 $^{\circ}\text{C}$ with spectra accumulated at both temperatures after 2 minutes equilibration.

Thermal transitions were followed at 5 separate wavelengths: 213 nm, 219 nm, 222.6 nm, 226 nm and 230 nm using a 3 nm bandwidth. The samples were heated using a 3 $^{\circ}\text{C}$ stepped gradient with five minutes equilibration at each temperature before ellipticity readings were taken, equating to an average temperature increase rate of 0.4 $^{\circ}\text{C}/\text{min}$. Temperatures were recorded using an internal probe placed into one cuvette. Ellipticity against temperature curves were plotted in Igor Pro Version 6.34 (Wavemetrics, USA) and fitted using equation 2.4 where T is temperature (in K) a_F is the folded signal intensity, b_F the pre-transition gradient, a_U the unfolded signal intensity and b_U the post-transition baseline gradient. ΔH is the enthalpy change upon protein unfolding (in kJ) and R the ideal gas constant with a value of 0.00831 kJ/K/mol.

$$\text{Eqn 2.4: } f(T) = \frac{(a_F + b_F T) e^{-\frac{\Delta H}{R}(\frac{1}{T_m} - \frac{1}{T})} + (a_U + b_U T)}{1 + e^{-\frac{\Delta H}{R}(\frac{1}{T_m} - \frac{1}{T})}}$$

2.7.3 Using BCM to follow changes in tryptophan fluorescence

Chemically and thermally induced unfolding transitions (sections 2.7.4 and 2.7.7) were followed using a PTI spectrofluorimeter (Photon Technology International, UK) with a Peltier temperature control and LPS-100 lamp. Fluorescence spectra were measured in a 1 cm pathlength quartz cuvette using an excitation wavelength of 280 nm. Unfolding transitions were followed by measuring the change in the barycentric median (BCM). The λ_{BCM} 'centre of mass' of each spectrum between 320 nm and 380 nm was calculated using equation 2.5 where $I(\lambda)$ is the fluorescence value at a respective wavelength. Each intensity is multiplied by the respective wavelength and the sum of these values divided by the sum of total intensities.

$$Eqn. 2.5 \lambda_{BCM} = \frac{\sum \lambda \cdot I(\lambda)}{\sum I(\lambda)}$$

The BCM values for each spectrum were plotted against their respective temperature or denaturant concentration and the unfolding transition followed by an increase in λ_{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 λ_{BCM} signal does not vary linearly with the fraction of folded protein. This is corrected using the relative quantum yields of the folded state relative to unfolded state (Q) as detailed in (142) and equation 2.6.

$$Eqn. 2.6 Q = \frac{\sum I(\lambda)_F}{\sum I(\lambda)_U}$$

2.7.4 Thermal transitions followed by tryptophan fluorescence

0.1 mg/mL protein samples were resuspended in 63 mM sodium phosphate buffer pH 7.4. These were equilibrated in a water-bath at 7 °C for a least 1 hour. Initial fluorescence spectra were recorded at a rate of 1 nm/s using an excitation wavelength of 280 nm and emission wavelengths of 320 - 380 nm where readings

were averaged over 2 repeat spectra. Samples were heated using a 3 °C stepped gradient with four minutes equilibration before each fluorescence spectrum was recorded, equating to an average temperature increase rate of 0.4 °C/min. Thermal unfolding curves of $\lambda_{\text{BCM320-380}}$ against temperature were fitted to an integrated Van't Hoff equation (equation 2.7). Q was defined in equation 2.6 and other parameters defined in section 2.4

$$\text{Eqn 2.7: } f(T) = \frac{(a_F + b_F T) \left(\frac{1}{Q}\right) e^{-\frac{\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T}\right)} + (a_U + b_U T)}{1 + \left(\frac{1}{Q}\right) e^{-\frac{\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T}\right)}}$$

2.7.5 Specific heat capacity determination

A series of 0.1 mg/mL protein samples were generated in 63 mM sodium phosphate buffer pH 7.4 at a range of GdnHCl concentrations below the mid-point ($[D]_{1/2}$) of protein unfolding. Each one was put through a thermal denaturation (section 2.7.4) and fitted with equation 2.7. The equation yields values for T_m and ΔH of the unfolding transition for each concentration of denaturant. These data pairs were used to create a ΔH against T_m plot. A weighted linear fit to these ΔH versus T_m data was used to determine the change in specific heat capacity between folded and unfolded states, ΔC_p .

2.7.6 Gibbs-Helmholtz equation to generate thermostability dependence

The values of ΔC_p and the average values of ΔH and T_m across four separate thermal denaturations in the absence of GdnHCl were inserted into the Gibbs-Helmholtz equation (equation 2.8) using values of T in 3 °C intervals, to produce a thermal stability curve (81)(143).

$$\text{Eqn 2.8: } \Delta G = \Delta H \left(1 - \frac{T}{T_m}\right) - \Delta C_p [(T_m - T) + T \ln\left(\frac{T}{T_m}\right)]$$

2.7.7 Chemical denaturation curves

Two stock solutions containing 63 mM phosphate buffer and 63 mM phosphate buffer with 6 M guanidine hydrochloride or urea were combined to produce stock solutions between 0 and 4-5 M at 1 M increments. A 1 mg/ml sample protein solution was diluted 1:9 into each solution. These stocks were then combined to produce a series of equally spaced denaturant concentrations each containing 0.1 mg/ml protein. Solutions were equilibrated overnight at 10 °C and measurements of each sample were recorded using an excitation wavelength of 280 nm and emission of 320-380 nm at a rate of 1 nm/s. For each spectrum the λ_{BCM} was calculated as per (equation 2.5). Chemical equilibrium curves were fitted to a two-state unfolding model in Igor Pro version 6.34 (Wavemetrics USA) using equation 2.9 (142-146) 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 denaturant concentration in the pre-transitional and post-transitional baselines. Q is again the relative quantum yield of the folded to unfolded state and R is the ideal gas constant and m is the dependence of ΔG^{F-U} on the mid-point of protein denaturation.

$$\text{Eqn 2.9: } f(D) = \frac{(a_F + b_F[D])\left(\frac{1}{Q}\right)e^{\left(\frac{-\Delta G + m[D]}{RT}\right)} + (a_U + b_U[D])}{1 + \left(\frac{1}{Q}\right)e^{\left(\frac{-\Delta G + m[D]}{RT}\right)}}.$$

2.8 Stopped-flow unfolding and refolding kinetics

Protein folding and unfolding was followed by the use of a stopped flow instrument (Applied-Photophysics, UK) cooled to 10 °C using a Neslab water bath. Unfolding experiments were performed by combining solutions of 30 μM Csp dissolved in pH 7.4 63 mM phosphate buffer separately in a 1:9 ratio with a series of phosphate buffer solutions containing different guanidine hydrochloride concentrations. The concentrations of guanidine hydrochloride used were increased in 0.125 M intervals between 1.75 M and 4 M GdnHCl. A wider range of guanidine hydrochloride concentrations (2.25 - 6 M with 0.25 M intervals) was used for $Tm\text{Csp}$. For refolding kinetics a solution of 30 μM unfolded Csp in 4.5 M GdnHCl (or 6M for $Tm\text{Csp}$) was combined in a 1:9 ratio with a series of separate low [GdnHCl] solutions (0 - 2 M).

Changes in fluorescence were followed at 280 nm. A minimum of 6 unfolding or refolding traces were obtained for each GdnHCl concentration after exclusion of erroneous traces which were then averaged. A previously calculated dead-time of 2 ms was assumed for all traces and was added to each time point. The average curves for each GdnHCl concentration were fitted to exponential curves in Origin 2015 using a Levenberg–Marquardt algorithm shown in equation 2.10 for folding and 2.11 for unfolding. y_0 is the fluorescence signal at time $t = 0$, A is the amplitude of the transition and t is time in seconds.

$$\text{Eqn 2.10: } f(f) = y_0 + (A_1)e^{kt}$$

$$\text{Eqn 2.11: } f(u) = y_0 + (A_1)e^{-kt}$$

The output rate constant k values were plotted against the concentration of denaturant $[D]$ to generate a chevron plot (148). The equations 2.12 and 2.13 were combined with equation 2.14 to generate equation 2.15. k_f represents the rate constant of protein folding and k_u the rate constant of protein unfolding.

$$\text{Eqn 2.12: } k_f = k_f^{H_2O} \times e^{[D](m_u)/RT}$$

$$\text{Eqn 2.13: } k_u = k_u^{H_2O} \times e^{-[D](m_f)/RT}$$

$$\text{Eqn 2.14: } k_{\text{obs}} = k_f + k_u$$

$$\text{Eqn 2.15: } k_{\text{obs}} = k_f^{H_2O} \times e^{[D](m_f)/RT} + k_u^{H_2O} \times e^{-[D](m_f)/RT}$$

Values for ΔG^{F-U} for each Csp were obtained from equation 2.16.

$$\text{Eqn 2.16: } \Delta G^{F-U} = -RT \ln \frac{k_f}{k_u}$$

As the unfolding of Csp is considered two-state, a value for m is obtained for each tested protein from the sum of the gradients for the folding and unfolding dependences in the chevron plot.

$$\text{Eqn 2.17: } m = m_f + m_u$$

The uncertainties in m were determined by a root mean squared approach in equations 2.18 and 2.19 where erm_u is error in m -value of unfolding and erm_f error in m -value of folding

$$Eqn\ 2.18: \text{Error in } m = \sqrt{erm_f^2 + erm_u^2}$$

$$Eqn\ 2.19: \text{Error in } m = \sqrt{\left(\frac{RT}{k_f} erk_f\right)^2 + \left(\frac{RT}{k_u} erk_u\right)^2}$$

The β Tanford value (β_T) was calculated using equation 2.20 shown below (147).

$$Eqn\ 2.20\ \beta_T = \frac{m_f}{m_{f-u}}$$

2.9 ^1H - ^{15}N HSQC NMR spectra

^1H - ^{15}N HSQC spectra were acquired with Varian Inova spectrometers with a cryogenic probe and proton resonance frequency of 600 MHz or 750 MHz. Spectra were acquired of samples containing 0.2 mM Csp in 0.5 ml 63 mM sodium phosphate buffer with 5% D_2O and 0.5% DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) as an internal shift standard. Locking and shimming were then performed and the 90° pulse calibrated. The ^1H - ^{15}N HSQC spectra were recorded over a time scale of 20 minutes. Spectra were taken at 5 $^\circ\text{C}$, 10 $^\circ\text{C}$, 23 $^\circ\text{C}$, 29 $^\circ\text{C}$ and 37 $^\circ\text{C}$ for each Csp. For readings denoted as including DNA a 1.1 molar ratio of dT7 was added to the samples. Spectra were processed using Bruker Topspin, NMR pipe and CCPN analysis software.

2.10 Atomic force microscopy

AFM unfolding experiments described in the next sections were performed on a custom built AFM instrument. For each experiment a fresh cantilever was used and new calibration performed. The same fluid cell was used throughout all experiments.

2.10.1 Preparation of gold surface for AFM

Gold-coated silicon was obtained from (Platypus technologies, USA). Spots of epoxy glue were added to the surface and a glass cover slip applied after each glue spot was added. The gold coated silicon was maintained at 120 °C for 2 - 3 hours. A fresh slide was removed from the surface for each experiment.

2.10.2 Calibration of the cantilever spring constant

A fresh probe containing 6 Silicon nitride cantilever tips with a thickness of 55 µm and cantilever lengths between 85 nm and 310 nm were fitted into a fluid cell. An O ring was inserted to prevent buffer evaporation and the fluid cell inserted into the AFM head. A laser was targeted on to the tip of the second largest cantilever with fine adjustment performed to maximise the amount of laser light detected on the photodiode.

Lyophilised polyprotein samples of 0.1 mg were resuspended in 63 mM buffer to a concentration of 0.4 mg/ml, filtered and 40 µl applied to a gold cover slip surface for 30 minutes to allow covalent binding of the sulfhydryl groups in the C-terminal cysteine residues of the polyprotein. Excess buffer was removed and replaced with fresh buffer to remove unbound polyprotein.

Calibration is performed using the thermal noise method first devised by *Hutter et al.* (149-151). Thermal calibration was performed by the production of a power spectrum measuring deflections of the cantilever within sodium phosphate buffer (Figure 2.1). A minimum of 5 repeats were performed and then the minima either side of the first large resonant peak were selected to allow an estimation of the area under the curve representing the voltage signal obtained from thermal fluctuations.

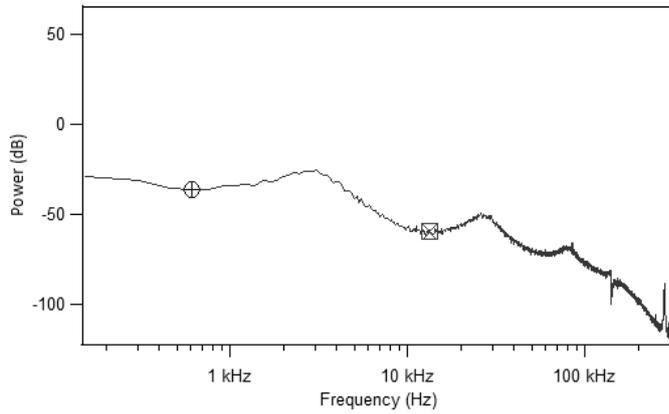


Figure 2.1: Typical thermal calibration scan with circle and square markers used to define the boundary of the first resonant peak.

The surface was brought into closer contact with the cantilever until a change in sum value was observed and then the distance from the surface optimised to give a clear approach and baseline (Figure 2.2). The potential difference was set to zero to centre the laser within the photodetector and corrected back to zero if it drifted beyond ± 1 . The slope of detected voltage (V) against displacement (z) gives the sensitivity of conversion of the cantilever deflection to a voltage (150-151).

$$\text{Eqn 2.21: } S = V/z$$

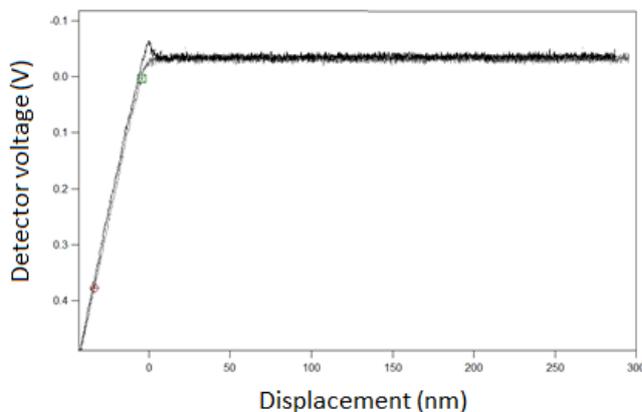


Figure 2.2: Example of a force against displacement calibration graph. The sensitivity is determined between the markers placed on the graph.

The thermal voltage signal was divided by the voltage-displacement conversion factor S to give the thermal displacement z . The value of the spring constant k_c is then calculated from equation 2.22. The spring constant of the cantilever were all found to be within a consistent range of 32 (± 2.5) pN.

$$\text{Eqn 2.22: } \frac{1}{2} k_B T = \frac{1}{2} k_c \langle z^2 \rangle$$

2.10.3 Force extension experiments

Unfolding experiments were performed at a series of different pulling velocities of 330 nm/s, 600 nm/s, 1100 nm/s, 2000 nm/s for polyproteins containing PB6 Csp. Pulling velocities of 100 nm/s, 200 nm/s, 600 nm/s and 2000 nm/s were used for *TmCsp* and *BsCsp* containing polyproteins. Force-extension traces were recorded in the programme Igor Pro. Force-displacement curves were recorded with a pushing force of 1500 pN.

2.10.4 Trace selection

Only traces that had a minimum of one cold shock protein unfolding event and at least two I27 events were considered for analysis and those containing large non-specific binding peaks, curved baselines or a larger number of domain unfolding peaks beyond the number present in a single polyprotein were rejected. Traces fitting these selection criteria were baselined and fit to a worm-like chain model by equation 2.23 where persistence length p is the stiffness of the polypeptide chain that approximates to the length of an amino acid, L_c is the contour length at rupture and x is the displacement between the tip and the surface (152).

$$\text{Eqn 2.23 : } F(x) : \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} \right]$$

The change in persistence length upon domain unfolding (ΔL_c) has been previously characterised to be 28 nm for I27 events and 23 nm for *TmCsp* events with a

persistence length of 0.35-0.4 nm. Events for which the WLC fits substantially differed from these values or traces containing forces in excess of 300 pN were rejected. All traces has baseline subtracted before force values were taken. For each unfolding event, the unfolding force (F_U) and peak to peak distance (p2p) was measured for each peak within the trace and the event assigned as either I27 or Csp unfolding based on the contour length and p2p distance. Where noise gave a higher unfolding force than the WLC fitted peak, the top of the peak was selected rather than the noise. The I27 and Csp events were separated and the respective pairs of F_U and p2p value for each unfolding event plotted to generate an unfolding force vs P2P distance scatterplot distribution. The median unfolding force and P2P were then recorded for I27 and Csp. The force and peak to peak distances were grouped into bins of 10 pN and 1 nm respectively and the frequency of events in each bin generated into histograms. A Gaussian curve was fitted to each to confirm a normal distribution of unfolding forces and p2p separation. Datasets with non-Gaussian distributions were discounted.

Three data sets which contained a minimum of 15 unfolding events for both I27 and Csp were obtained for each pulling velocity. Each experimental dataset was performed on a separate day with a fresh cantilever and calibration. Additional repeats were performed where either insufficient numbers of events had been obtained or where the median unfolding forces differed by more than 10 pN between triplicate datasets. The median unfolding forces and p2p distances of the 3 data sets at each pulling velocity were averaged. Average unfolding forces were plotted against the natural logarithm of their respective pulling velocity to generate a force pulling velocity dependence. These were then fitted with a linear fit.

2.10.5 Variable temperature force extension experiments

For the variable temperature experiments, a pump was used with tubing placed in an ice bucket and the flow rate adjusted to acquire the desired temperature (5 °C, 10 °C or 14 °C) and without an ice bucket for 32 °C. The pump was kept flowing

throughout the experiment. As before three data sets were taken at each of the 4 pulling velocities with a minimum of 15 Csp unfolding events in each.

2.10.6 Monte Carlo simulations

The unfolding energy landscape is modelled as a two-state process as conforming to the Bell model where each protein domain is either folded or unfolded. In the model A is attempt frequency, k_B the Boltzmann constant and $G(0)$ the free energy barrier (Equation 24).

$$\text{Eqn 24: } k_u(F) = A e^{-\left(\frac{G(0)-F\Delta x_u}{k_B T}\right)}$$

The model assumes the folded and unfolded states are separated by a single energy barrier controlled by the unfolding rate constant k_u and the mechanical distance from the folded state to the transition state (TS) along the reaction co-ordinate Δx_u . In a heteropolyprotein chain however, the values of Δx_u and k_u for I27 and for Csp are not independent of those for the alternate domain type.

Monte Carlo simulations are used to predict the unfolding forces in 1000 simulated construct extensions at an individual pulling speed using different values of Δx_u and k_u for each type of domain. A random number of domains is selected between 3 and 7 for each probability simulation. The values of temperature, persistence length, time step, linker length and the increase in distance upon domain unfolding are fixed in the simulations (Table 2.11). The values for the spring constant were averaged across the 12 experimental repeats for each polyprotein at each temperature. The simulations assume that domains do not refold under the short timescale and forces used. Simulations were conducted with a script in Igor Pro at pulling velocities of 330 nm/s and 2000 nm/s.

The simulations begin with all protein domains designated as folded and the extension per time step calculated (153). The probability of a domain unfolding is calculated at a time step (therefore at a force, F) and then compared to a number generated by a random number generator between 0 and 1. It is crucial that a small time step is selected to avoid probability values significantly over 1 occurring (153). If the probability exceeds the random number then one folded domain is deemed to have unfolded and the length of the chain extended by the difference in folded and unfolded chain lengths. If the probability is smaller than the random number calculations is performed again at the next time point when the force will be increased. The unfolding data proceeds with a value of N of one less each time a domain unfolds until no domains remain folding. The unfolding forces derived from repeats of the MC simulations with identical values of Δx_u and k_u generally fell within ± 1 pN when using simulated data for unfolding of 1000 traces.

The median unfolding forces determined using the simulations were then compared to the experimentally derived values. The values of Δx_u and k_u for Csp and I27 were then adjusted to obtain an improved correlation with experimental data until the force values of I27 and Csp aligned within ± 1 pN for both domains at both pulling velocities. To obtain upper and lower bounds for each parameter the standard deviation in unfolding forces between the 3 experimental datasets were added or subtracted from the median unfolding force values at 330 nm/s and 2000 nm/s to produce the sharpest and shallowest force pulling speed dependences within experimental uncertainties. These were again fitted with the Monte Carlo simulations to obtain the maximum and minimum values of Δx_u and k_u for each domain. The 3 values of median, maximum and minimum Δx_u and k_u were averaged to obtain a more reliable estimate of these parameters.

Table 2.11: Parameters fixed in the Monte Carlo simulations for (I27-Csp)₃-I27

Parameter	Value
Distance between N and C-terminal of folded Csp	1.4 nm
Distance between N and C-terminal of folded I27	4.1 nm
Persistence length	0.35 nm
Linker length	2.4 nm
Spring constant k_c	31-33 pN/nm (averaged over 12 experiments)
Temperature	5, 10, 14, 23 or 32 °C
Time step	0.0001s
Pulling velocity	100 nm/s or 330 nm/s and 2000 nm/s
Number of molecules to be pulled	1000

A k_u value is a measure of the probability of a protein domain unfolding. It was used to calculate the value of the activation energy of unfolding ΔG^{F-TS} using a rearranged Arrhenius equation where A is assigned a value of 1×10^6 (Equation 25).

$$Eqn\ 25: k_u = A e^{\left(\frac{-\Delta G^{F-TS}}{K_B T}\right)}$$

3 Biophysical characterisation of the monomers

3.1 Introduction

Extremophilic proteins can provide many insights into the mechanisms by which proteins are able to be dynamic whilst maintaining thermostability. Families of protein homologues represent ideal candidates for studying the impact of temperature adaption as they exhibit high levels of conservation. A common problem with such comparisons is that the data are often derived from separate studies which can introduce elements of variation because of the different techniques and conditions employed. A significant number of studies have been performed regarding the differences between mesophilic and thermophilic protein homologues but there are very few that include cold-adapted variants. A summary of some this is shown in (16). This section focuses on a comparison of the relative thermostabilities of the family of cold shock proteins (Csps) which represent a model system due to their small size and the presence of Csps in organisms living at a wide range of temperatures. Three cold-adapted Csps were selected to study the differences in their thermostabilities compared with that of an extreme hot-adapted (hyperthermophilic) Csp and temperate (mesophilic) Csp. The comparisons are made across all the Csps at the ambient temperatures of 10 °C and 23 °C as well as at the respective temperatures their host organisms live at. This study aims to ascertain how and to what extent the thermostabilities of Csps differ between those evolved to function at high temperatures and those which are evolved to function at lower temperatures. All of the selected Csp sequences were cloned into *E. coli* cells, expressed and purified.

3.1.1 The Gibbs-Helmholtz equation and protein stability curves

Proteins in general are only marginally stable as thermostability is a balance between two factors, enthalpy (ΔH) and the product of entropy and respective temperature ($T\Delta S$). Both of these factors are temperature dependent and are related by equation 3.3. Only over a relative short range of temperatures is the outcome of this sum negative and the protein folded. A representation of this can

be given by thermostability curves. When Gibbs free energy is expressed in terms of unfolding ΔG^{F-U} a more negative ΔG^{F-U} value indicates a greater thermostability at this temperature and a value of $-\Delta G^{F-U}$ below the x-axis (i.e. positive) means the protein will unfold at that temperature.

It is possible to directly measure the temperature dependence of protein stabilities using values extracted from fits to equilibrium curves performed at different temperatures however this is very time consuming. An alternative approach is to derive a prediction of the full thermostability temperature dependence profile of a protein by using the Gibbs-Helmholtz equation (41). This is derived from the separate terms of the Gibbs free energy equation as shown below where T_m is the mid-point temperature of thermal unfolding and ΔC_p the change in specific heat capacity between folded and unfolded states. ΔS_g and ΔH_g are the values of entropy and unfolding enthalpy respectively at T_m .

$$\text{Eqn 3.1: } \Delta H(T) = \Delta H_g + \Delta C_p(T - T_m)$$

$$\text{Eqn 3.2: } \Delta S(T) = \Delta S_g + \Delta C_p \ln\left(\frac{T}{T_m}\right)$$

$$\text{Eqn 3.3: } \Delta G(T) = \Delta H - T\Delta S$$

$$\text{Eqn 3.4: } \Delta G(T) = \Delta H_g\left(1 - \frac{T}{T_m}\right) + \Delta C_p\left(T_m - T + T \ln\left(\frac{T}{T_m}\right)\right)$$

The Gibbs-Helmholtz equation (equation 3.4) gives a full projection of the thermostability of proteins at different temperatures and can allow a comparison of the different effects of temperature adaption across a range of temperatures (Figure 3.1) (154). Three main factors can alter the shape and height of the thermostability curve. The maximum thermostability is mainly controlled by the value of enthalpy of unfolding (ΔH). The width of the curve is mostly determined by the change in specific heat capacity between folded and unfolded states (ΔC_p) and the right intercept set by (T_m).

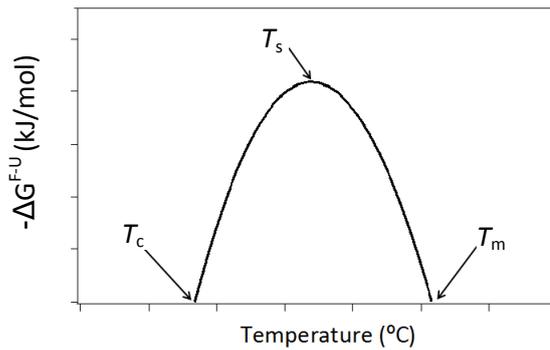


Figure 3.1: Thermostability curve for a protein displaying the dependence of Gibbs free energy of unfolding $-\Delta G^{F-U}$ on temperature. Greater thermostability at a temperature is indicated by a greater $-\Delta G^{F-U}$ value while values below $-\Delta G^{F-U} = 0$ on the graph indicate a temperature at which the protein would be unfolded. The intercept of the curve with x-axis at the right-hand side of the curve is the mid-point of thermal denaturation T_m , the equivalent point on the left-hand side is the mid-point of cold denaturation T_c . T_s is the temperature at which the highest $-\Delta G^{F-U}$ is observed. Dotted line indicates position of 0°C for a typical protein.

3.1.2 Adaptations of the thermostability curves of thermophilic proteins

Thermophilic proteins are commonly seen to be stable over a wider range of both higher and lower temperatures than their mesophilic homologues (16). One thing which appears to vary between different homologue comparisons is the extent to which the temperature of maximum thermostability is shifted to higher temperatures in thermophiles and lower temperatures in cold-adapted proteins. Some comparisons show a substantial shift while others show minimal differences. A lack of change in the temperature of maximum thermostability could be due to the thermostability profile shape being governed by the same hydrophobic effect.

All thermophilic proteins are characterised by an increase in the maximum temperature at which they remain stable T_m . In terms of thermostability curves this can be achieved by a number of methods as summarised in Figure 3.2 and (155-156). One strategy to increase T_m which is applied by proteins such as thermophilic RNaseH homologues is an increase in unfolding enthalpy which expands the whole of the thermostability curve profile (157). This is mediated by stronger hydrophobic packing, increased electrostatic interactions and compaction of cavities to exclude

water molecules (158). An alternative strategy employed by thermophilic proteins including cellulose and RNaseH is a reduction in the change in specific heat capacity ΔC_p (159-161). A ΔC_p reduction actually reduces the maximum ΔG^{F-U} but also broadens the thermostability profile to higher and lower temperatures making the proteins less sensitive to temperature changes (161). Homologues of the same protein family do not generally show large ΔC_p differences due to high levels of conserved structure (160)(162). The final method is to shift the whole thermostability curve to higher temperatures through increase in temperature of maximum stability T_s which can be achieved by entropic changes which reduce the entropy gain upon protein unfolding. Shifts of the thermostability curve to higher temperatures with no change in unfolding enthalpy or ΔC_p , is very rare. Thermophilic proteins can increase T_m by evolving through one or more of the above approaches and an ideal thermophilic protein has a high ΔH , high T_s and low ΔC_p (155-156).

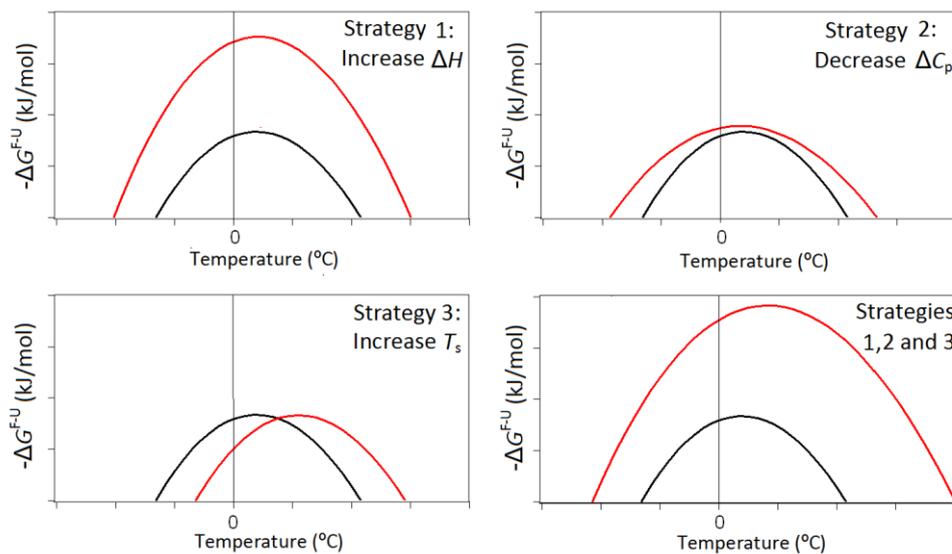


Figure 3.2: Three strategies employed by evolution to increase T_m in thermophilic proteins. Black indicates the standard protein and red after the adaption has been made. Strategy 1 shows a 50% increase in ΔH with no shift in ΔC_p or the temperature of optimum thermostability T_s . Strategy 2 involves stretching of the thermostability curve by a decrease in ΔC_p with no change in T_s or ΔH . Strategy 3 involves a full shift of the thermostability curve to higher temperatures with no other changes. The graph in the bottom right shows how the strategies can be combined to produce a greater effect. Any hot-adapted protein may use 1, 2 or all 3 of the strategies but not every comparison of a temperate and hot-adapted homologue will exhibit all 3 strategies.

3.1.3 Evolution of thermostability curve parameters

Studies have used phylogenetic trees and current DNA sequences to attempt to ascertain what the sequences of ancestral proteins may have been and how they evolved to be what they are today. These studies have yielded interesting conclusions. The evolution of RNaseH to increase thermostability was proposed to proceed by systematic drift where at each stage the parameters ΔH , ΔC_p and T_s fluctuate and are slowly optimised to an ideal solution (34). At each individual step the overall thermostability is increased but the change in each individual parameter may be favourable or unfavourable. This leads to gradual changes building up greater thermostability, however the history of evolution towards mesophilic and colder temperatures of RNaseH was proposed to have proceeded by a more stepwise manner with large sudden drops in thermostability interspaced by long periods of little change (Figure 3.3).

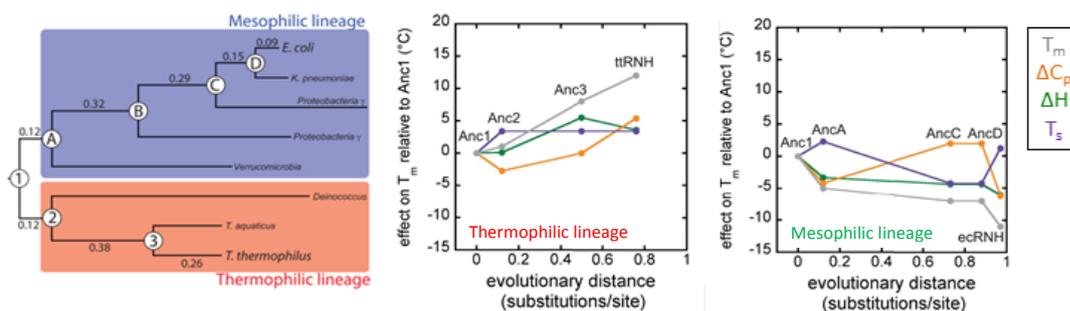


Figure 3.3: Evolution of mesophilic and hyperthermophilic proteins evaluated by extrapolation to ancestral sequences. Left: Proposed phylogenetic tree based on extrapolation of current RNaseH sequences to identify predicted ancestral sequences. middle: How changes in ΔC_p (orange), ΔH (green) and T_s (purple) build up over evolutionary time to produce a gradual increase in T_m in the evolution of a hyperthermophilic RNase sequence. Right: Loss of thermostability occurs in a stepwise manner over evolutionary time with different combinations of favourable and unfavourable thermodynamic changes. Adapted from (34).

In various thermostability parameters, opposite trends to those discussed above are expected in psychrophilic proteins so this would predict a reduction in unfolding enthalpy, narrowing of the thermostability curve and a shift of the curves to lower temperatures. The picture becomes complicated as psychrophilic proteins still need

to maintain thermostability so too much destabilisation could lead to transient unfolding. This means it cannot be assumed that the adjustments to the thermostability curves occurring for cold-adapted proteins are the opposite of those for a thermophilic protein.

3.1.4 Previous thermostability curve studies

A study which does include a psychrophilic homologue is a comparison of homologues of the enzyme α -amylase (16)(163) (Figure 3.4). The study showed that a thermophilic α -amylase from *Thermobifida fusca* (TFA) exhibited a greater thermostability across all temperatures than mesophilic α -amylases from *Sus scrofa* (PPA), *Drosophila melanogaster* (DMA). The psychrophilic α -amylase from *Pseudoalteromonas haloplanktis* (AHA) displayed the lowest thermostability across most of the temperature range and remains folded over the smallest range of temperatures (16)(163). The maximum thermostability ΔG_{\max} of the psychrophilic α -amylase reduced compared to its homologues and the temperature at which the maximum thermostability T_s occurs is also reduced. The shapes of the curves were generally similar suggesting only small ΔC_p differences.

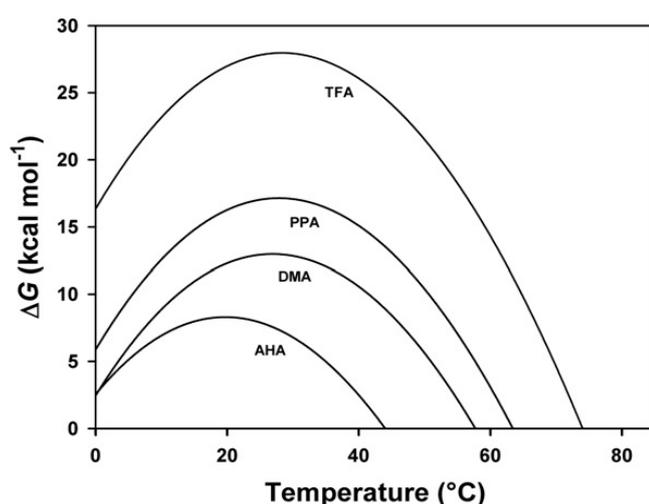


Figure 3.4: A comparison of the thermostabilities of α -amylase enzymes including homologues from the thermophile *Thermobifida fusca* (TFA), mesophiles: pig pancreas (PPA), *Drosophila melanogaster* (DMA) and psychrophile *Pseudoalteromonas haloplanktis* (AHA). Taken from (41)

A study of temperature adapted di-hydrofolatereductases showed how an extreme psychrophilic di-hydrofolatereductase (DHFR) from the bacteria *Moritella profunda* which has an optimal growth temperature of less than 2 °C has evolved very different approach (Figure 3.5) (164). The psychrophilic DHFR shows a reduced change in heat capacity ΔC_p relative to variants from mesophilic *E. coli* and thermophilic *Thermotoga maritima*. A broadening of the thermostability curve from a reduction in ΔC_p is typically associated with being a strategy used by thermophilic proteins to expand the thermostability curve to higher temperatures yet in this case a psychrophilic enzyme uses this same approach to expand the thermostability curve to lower temperatures to combat the problem of cold denaturation.

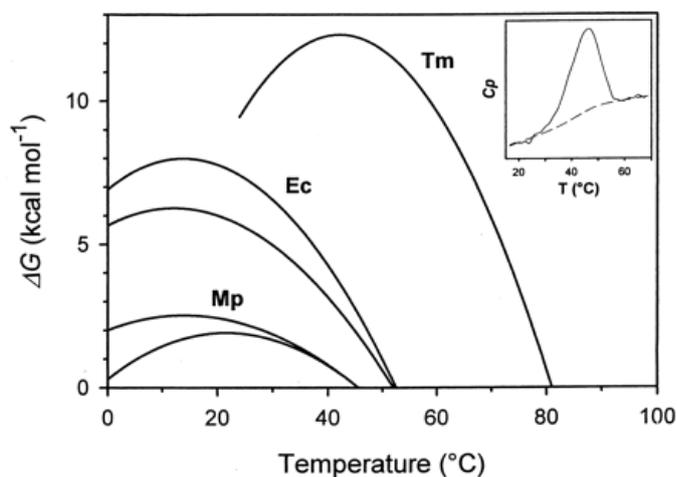


Figure 3.5: Comparison of the temperature dependence of dihydrofolate reductases from the hyperthermophile *Thermotoga Maritima* (Tm), mesophile *E. coli* (Ec) and psychrophile *Moritella profunda* (Mp). Taken from (164)

3.1.5 Aims

The initial phase of this investigation focuses on selecting a family of protein homologues with variants living at different temperatures. The sequences of members of the selected cold shock protein (Csp) family are aligned and analysed to observe the extent of sequence divergence between the different sequences from bacteria living at different temperatures and to allow identification of patterns between the homologues. The Csp's were then expressed in *E. coli* and purified with

detail given on optimising these procedures (sections 2.3-2.4). Thermostability parameters are determined allow full stability curves to be generated using the Gibbs-Helmholtz equation and the relative thermostabilities compared at a range of temperatures. Three cold-adapted Csp variants are considered to see how their thermostability curves adjust from the mesophilic variant and how they maintain thermostability at low temperatures. The thermostabilities relative to the optimal growth temperatures of their respective organisms are also considered. The final part of the section then examines 3 Csp mutants in an attempt to gain more insight into the factors behind the very high thermostability of the hyperthermophilic CspB from *Thermotoga Maritima*. The mutants include a *BsCsp* variant where additional electrostatic interactions are added from the hyperthermophilic *TmCsp* to increase its thermostability and 2 *TmCsp* mutants with shorter side chains to reduce hydrophobic packing. This should provide more detail on how the thermodynamics of the Csps vary in Csps evolved to function in extreme hot and cold conditions.

3.2 Results

3.2.1 Selecting and purifying the cold shock protein monomers

The overall aim of this study was to investigate differences in the thermal, thermodynamic, kinetic and mechanical properties of protein homologues isolated from psychrophilic, psychrophilic, mesophilic and hyperthermophilic organisms. The protein domains selected for study were to ideally be relatively small (below 100 amino acids) and have a stable folded structure. Proteins have not only evolved specific traits with regard to temperature as many organisms are also evolved to extremes of pH (acidophiles, alkaliphiles), high pressure (piezophiles) and high salt conditions (halophiles). Each of these conditions provides their own selective pressures on the organisms living in them. An organism can be adapted to multiple extremes such as a thermoacidophile or a psychrohalophile. These types of adaptations must be carefully considered when selecting different temperature adapted variants as each of these have their own set of adaptations so this could interfere in comparisons. Terms such as psychrotrophic, mesophilic and

thermophilic are often applied to categorize proteins based on the temperature their source organisms live at, however in reality thermal adaptation is a spectrum with hot and cold-adapted proteins adapted to varying extents.

A selection of protein families containing homologues across organisms adapted to vastly different conditions were selected as potential candidates for comparative studies including I27, twitchin, ubiquitin, lysozyme and cold shock proteins. The muscle proteins I27 and twitchin were compared in species of bivalves. These organisms contain many muscle proteins with similar amino acid sequences so it proved difficult to confirm which of the sequences were from homologues of the same protein family. The changes that mediate thermophilicity and psychrophilicity are very small so it is often difficult to distinguish between them and general genetic variation. Lysozyme was considered as a candidate protein family as it is an enzyme so differing activity levels could be tracked to confirm function. Part of the investigation was to include a mechanical characterisation and as previous studies into the mechanical stability of lysozyme showed it to be relatively mechanically weak, lysozyme was discounted (165).

The family which was selected is the cold shock protein family which comprise small β -barrel proteins that bind to DNA and RNA (section 1.6). The small size of Csp allows comparisons without subunit interactions or ion binding affecting the comparisons (166). Numerous variants of Csp had previously been characterised including a hyperthermophilic variant from *Thermotoga maritima* (*TmCsp*) which was seen to have a clear mechanical unfolding signature (136-137). Csp are widely conserved across all kingdoms of life including prokaryotic and eukaryotic variants. For the eukaryotic variants three bivalve molluscs were considered: *Laternula*, *Mya* and *Chlamys* which live at different water temperatures (167-169). When aligned, the sequences between the bivalves were found to be very near identical suggesting that these organisms are adapted at the system level to temperature

rather than at the protein level (Figure 3.6). It was therefore decided to focus on the prokaryotic Csp's.



Figure 3.6: Alignment of cold shock domain containing proteins from bivalves. *Laternula* (-12 to 3 °C) lives at the coldest temperatures of the three, *Mya* at (-5 to 6 °C) (and *Chlamys the highest* (0 to 8 °C)(167-169). Sequences were obtained from British Antarctic Survey databases. Blue indicates conserved residues. Bars indicate conservation with yellow representing full conservation

A hyperthermophilic Csp variant originating from *Thermotoga maritima* has been previously well characterised (166)(170-171). *Thermotoga maritima* is a bacteria isolated from deep sea thermal vents off the coast of Italy. Data on cold adapted Csp variants is sparse although recently detailed analysis has been conducted on a Csp from psychrotrophic bacteria *Listeria monocytogenes*, isolated from a fridge in Korea. For this study two published sequences of Antarctic bacterial Csp's from *Psychromonas ingrahamii* and *Psychrobacter* PB6 were selected (172-175). Both have proposed structures based on bioinformatic modelling approaches. To select a third sequence to study, the closest cold-adapted homologue to the *Thermotoga* Csp sequence was found using the homology search tool BLAST (176). This was a Csp from the Antarctic bacteria *Shewanella livingstonensis* (177). Amino acid sequences were optimised for *E. coli* codon preference (sequences shown in appendix S1).

3.2.2 Multiple sequence alignments of Csp sequences

To discover to what extent the sequences of Csp's isolated in bacteria, living at different temperatures differ from each other, a selection of Csp sequences were aligned in the bioinformatics tool Jalview (140). All the sequences contained between 66 and 73 amino acids and were aligned using a ClustalW algorithm. The

alignment in Figure 3.7 shows a very high degree of conservation between the Csp's with large sections being identical (highlighted in blue). The sequences of β -sheet regions and those involved in nucleotide binding are particularly highly conserved. This fits with the theory of temperature adaptations involving only a small number of changes and residues with a functional or structural role being conserved. There is an increase of the highly flexible amino acid residue glycine in the cold-adapted variants with a cluster of 3 glycine residues seen in within the largest loop in the *Psychrobacter* PB6 sequence from residues 40-43. The length of loops was similar across all Csp's studied indicating no adaption in terms of loop lengths. The positions of hydrophobic residues were generally highly conserved with no clear temperature adaption pattern for either number or size.

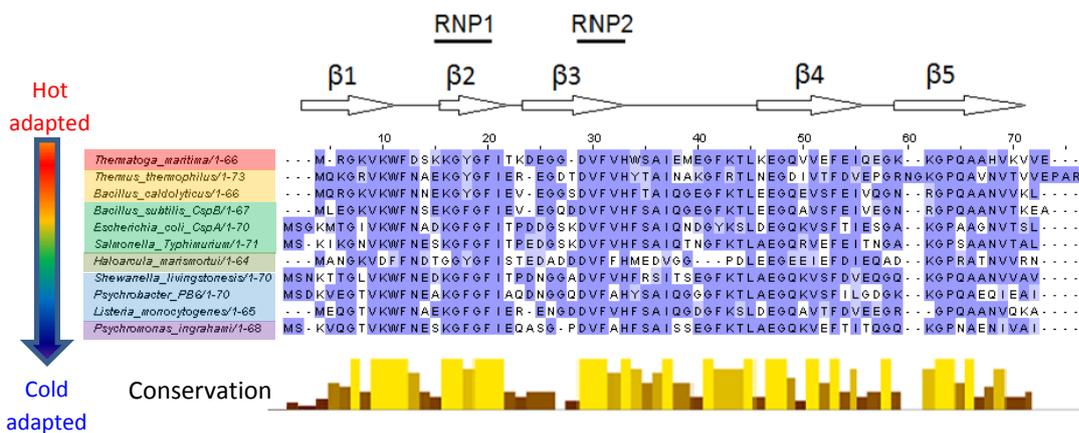


Figure 3.7: Multiple sequence alignments of different temperature adapted cold shock proteins from extreme hot adapted (top) to extreme cold adapted (bottom). The sequences are shown alongside the secondary structure and RNA binding sites above. The alignment is colour coded based on residue type and homology with deeper blue representing conservation. A measure of the degree of conservation at each site is indicated by the height and colour of the conservation bar below with yellow representing greatest conservation. The Csp sequences are colour coded by temperature adaption with Hyperthermophilic (shown in red): *Thermotoga maritima* (1G6P), Thermophilic (shown in orange): *Thermus Thermophilus* (3AOJ), *Bacillus caldolyticus* (2HAX). Mesophilic (shown in green): *Bacillus subtilis* (3PF4), *E. coli* (1MJC), *Salmonella typhimurium* (3I2Z), Halophilic (shown in olive): *Haloarcula marismortui* Psychrotrophic studied for this investigation *Shewanella livingstonensis* and *Psychrobacter PB6* (shown in blue), *Psychromonas ingrahamii* (shown in purple).

The positions of the 8 aromatic residues were fully conserved between all the Csp sequences and this is probably due to them playing a crucial role in the binding of DNA (Figure 3.7). There were some variations in the specific aromatic residue present at each position with more of the larger aromatic residues of tryptophan and tyrosine present in the hot-adapted Csps (Table 3.1). This may agree with theories of more specific binding in hot-adapted enzymes and less in cold-adapted forms (section 1.4.3). Mutation of the aromatic residues to short hydrophobic residues showed that the aromatic residues are important to both Csp thermostability and function (178).

The location and nature of charged residues was relatively highly conserved across the Csps. One of the clearest differences was a general trend of increasing numbers of charged residues in Csp sequences from organisms living at high temperatures and fewer present in Csps from bacteria living at low temperatures (Figure 3.3 and Table 3.1). Charged residues are known to mediate increased thermostability through salt bridges in hot adapted proteins relative to other homologues. Mutation studies on the thermophilic Csp, *BcCsp* derived from *Bacillus caldolyticus* and mesophilic *BsCsp* show that the differences in thermostability are controlled by a very small number of substitutions which alter the protein electrostatics (179). There has been debate as to whether a reduced number of salt bridges in cold-adapted variants is due to lack of selective pressure for their conservation or active evolutionary pressure to substitute them to increase flexibility. The balance of positive and negative charges remains similar across most Csps apart from *Haloarcula Marismortui* which has far more negative charges but this is commonly seen in organisms living in high salt conditions. *TmCsp* shows the greatest number of charged residues of the studied Csps with a particularly high number of lysine residues.

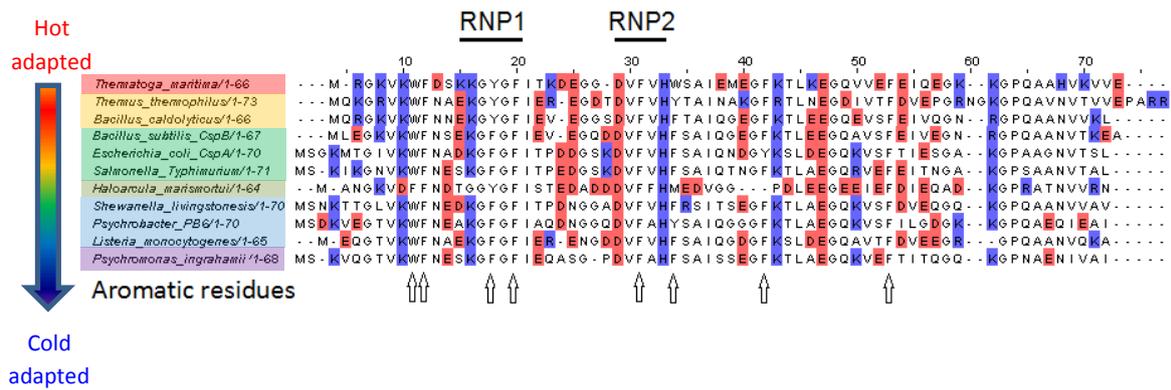


Figure 3.8: Csp alignment highlighting charged residues with acidic residues in red and basic residues in blue. The sequences are ordered in groups from hyperthermophilic shown top to extreme psychrophile bottom. The positions of the 8 conserved aromatic residues are indicated by arrows.

Table 3.1: The number of charged residues present in different Csp from hyperthermophilic to extreme psychrophilic Csp shown from top to bottom

Species of Csp origin	Residues	Positively charged	Negatively charged	Charged total	Aromatic
<i>Thermotoga maritima</i>	66	13	11	24	8 (5F 1Y 2W)
<i>Thermus thermophilus</i>	73	12	9	21	8 (5F 2Y 1W)
<i>Bacillus caldolyticus</i>	66	8	9	17	8 (6F 1Y 1W)
<i>Bacillus subtilis</i>	66	7	12	19	8 (7F 1W)
<i>Escherichia coli</i>	69	8	8	16	8 (7F 1W)
<i>Salmonella typhimurium</i>	71	10	7	17	8 (7F 1W)
<i>Haloarcula marismortui</i>	64	5	18	23	8 (7F 1Y)
<i>Shewanella livingstonensis</i>	70	8	9	17	8 (7F 1W)
<i>Psychrobacter PB6</i>	69	8	8	16	8 (6F 1Y 1W)
<i>Listeria monocytogenes</i>	65	7	12	19	8 (7F 1W)
<i>Psychromonas ingrahamii</i>	68	7	7	14	8 (7F 1W)

The aligned sequences were generated into a phylogenetic tree to ascertain more information as to the evolutionary relationship between the Csp (Figure 3.9). The phylogenetic tree shows a close evolutionary relationship between the thermophilic adapted proteins which are likely to have maintained more similarity due to the selective evolutionary pressures of high temperatures. The close relationship

between hot-adapted Csps could be because only a very limited combination of mutations is capable of producing sufficient thermostability at high temperatures, leading to highly conserved structural regions. The cold-adapted sequences here appear to show a more distant evolutionary relationship. Life in the past was adapted to live at higher temperatures and so psychrophiles may have diverged genetically from each other as they became cold-adapted. Diversity between psychrophiles may be as a result of there being a much wider array of separate approaches to achieve greater flexibility than greater thermostability. Alternatively a natural drift to lower stabilities may occur in the absence of strong selection pressure towards greater stabilities. It is very difficult to distinguish between these theories without additional Csp dynamics studies.

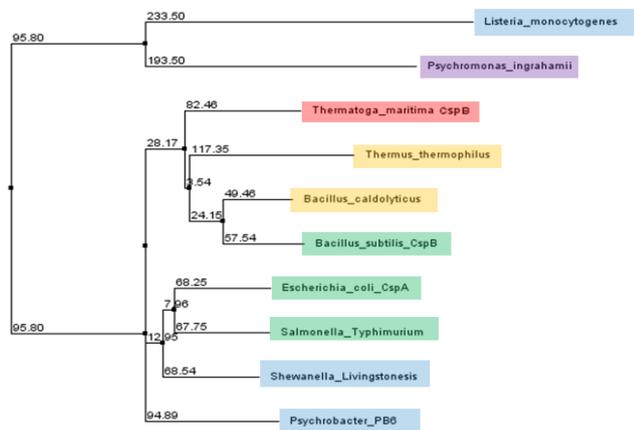


Figure 3.9: Phylogenetic tree of a selection of bacterial Csps. Hot-adapted Csps shown in red and orange, temperate in green and cold-adapted in blue and purple.

The percentage homology score across the sequences showed a greater percentage identity between the cold-adapted variants (60-70%) and lower levels between the cold-adapted Csps and the hyperthermophilic *TmCsp* variant (50-60%) showing that patterns of conservation are related to temperature (Table 3.2). This may be related to earlier life being hot-adapted and psychrophilic proteins evolving away from that (section 1.2). The cold adapted sequences showed high percentage similarity to each other despite the phylogenetic tree suggesting a more distant relationship.

Table 3.2: Percentage sequence identity between the different Csp homologues

	<i>TmCsp</i>	<i>BsCsp</i>	<i>S/Csp</i>	PB6 Csp	<i>PsiCsp</i>
<i>TmCsp</i>	-				
<i>BsCsp</i>	64.2	-			
<i>S/Csp</i>	60.6	65.2	-		
PB6 Csp	55.8	56.9	63.8	-	
<i>PsiCsp</i>	60.3	57.4	65.2	68.1	-

3.2.3 Generation of over-expression plasmids for *BsCsp* and 3 cold-adapted Csp

The sequences for the Csp were synthesised in pEX vectors by Eurofins. Csp are known to affect *E. coli* growth rates and so the Csp sequences had to be transferred in to an alternate plasmid under tighter regulatory control (pET3a). Previous studies have shown reasonable expression yields for *TmCsp* using a pET3a expression vector controlled by a T7 viral promoter and a lac operon based system (181). To transfer the Csp sequences from pEX to pET3a firstly PCR primers were designed with sections complementary to the 5' and 3' ends of each Csp sequence. These were used to amplify the amount of the Csp DNA present and add specific *XhoI* and *MluI* restriction sites respectively at the 5' and 3' ends of each Csp (see section 2.2). The amplified DNA could then be seen by clear bands at just over 220 bp on an agarose gel (Figure 3.10).

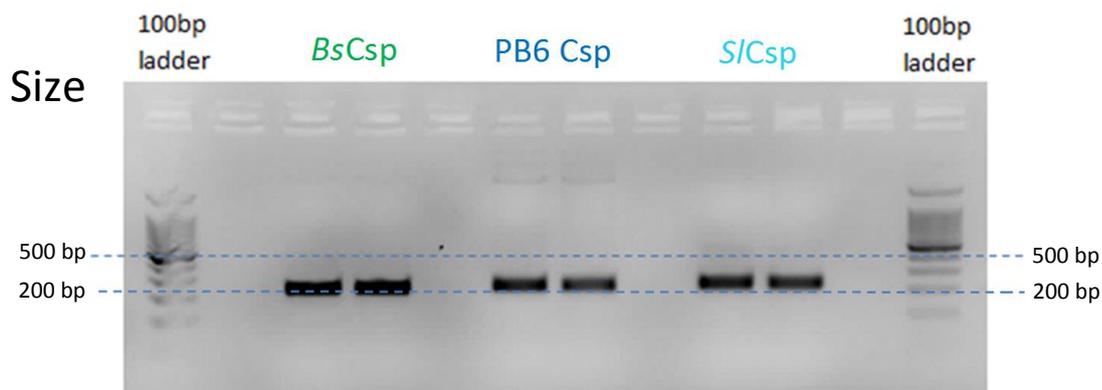


Figure 3.10: 1.5% agarose gel showing addition of *XhoI* and *MluI* restriction sites to Csp sequences by PCR. The desired product is approximately 220 bp for each Csp.

The PCR product was extracted from the agarose gel and cleaved by *XhoI* and *MluI* to generate sticky ends. The same restriction sites were used to cut a pET3a vector supplied by Dr Toni Hoffmann which contained a hexahistidine tag sequence upstream of the insertion site. Correct cleavage of the vector could be observed by the correct size of the previous insert which was around 1 kbp in size (Figure 3.11). Each Csp DNA sequence was a separately ligated into cut vector and transformed into DH5 α cells for amplification. Correct insertion of the Csp sequences was verified through sequencing of mini-prep extracted plasmid.

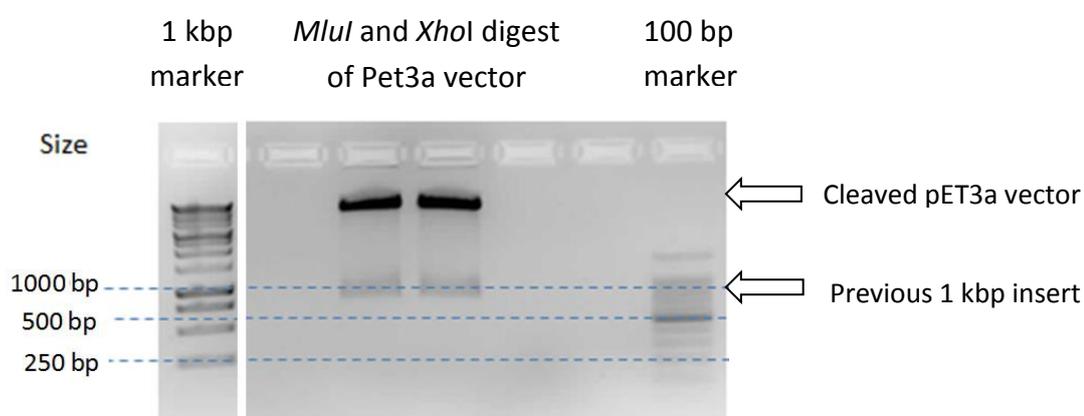


Figure 3.11 Agarose gel displaying products of restriction digestion of pET3a plasmid containing (TmCsp)₄-I27-TmCsp-I27 and *XhoI* and *MluI*. Arrows right illustrate position of cleaved vector and insert.

3.2.4 Optimising expression of the monomers

The sequenced plasmid was transformed into BL21 (DE3) pLysS cells for expression. Cells containing plasmids with Csp sequences were grown substantially slower than untransformed equivalent *E. coli* cells. The pLysS plasmids help to decrease the strain of Csp on the *E. coli* cells prior to induction as it produces T7 lysozyme which is an inhibitor of the viral T7 polymerase and so further suppresses endogenous expression of the Csp. Carbenicillin antibiotic was included in all expression media as previous studies had reported frequent losses of TmCsp containing plasmids (179). Initial protein expression trials were conducted at the commonly used temperature of 37 °C, however only very limited levels of protein expression were seen in the SDS PAGE gels at the correct mass of 8.5 kDa (Figure 3.12).

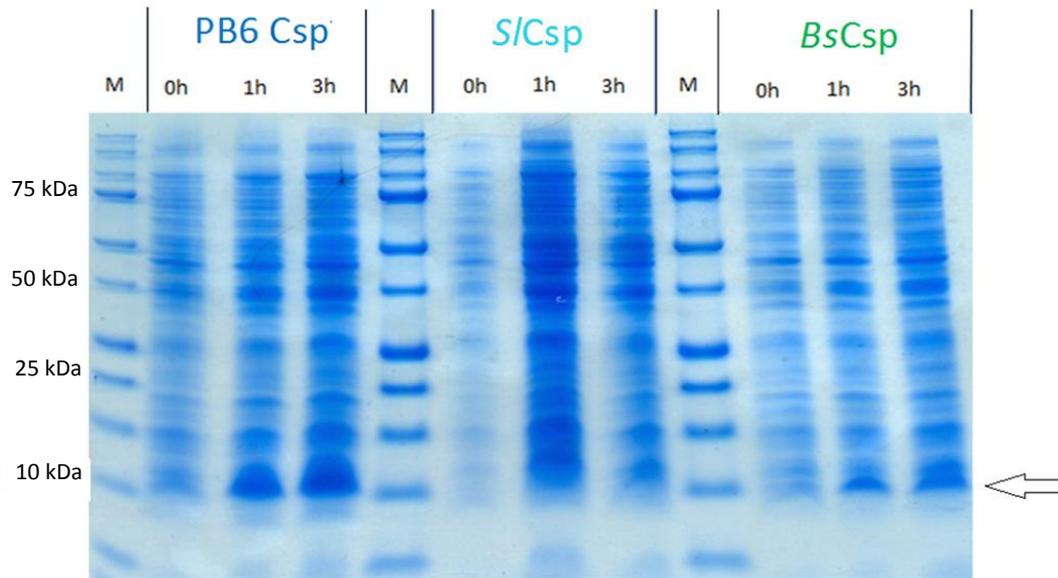


Figure 3.12: Expression profile of BL21 DE3 cells transformed with Csp sequences, induced with IPTG and incubated at 37 °C. Arrow indicates expected mass of Csp (around 8.5 kDa)

Biologically Csp are expressed in response to a drop in temperature but are not expressed at high levels at the optimal growth temperatures of their respective bacteria (115)(180). Using this information, a lower expression temperature of 26 °C was trialled in agreement with a method used by *Welker et al.* for *TmCsp* expression (179). This produced a significant improvement in expression of all the Csp which was highlighted by lysing the cells and performing Ni-NTA pull-down assays (Figure 3.13). The Csp contain an N-terminal hexahistidine affinity tag to allow binding to nickel beads while endogenous *E. coli* proteins are removed over multiple wash steps. The elution fraction bands clearly show a band at around 8 kDa. While lower temperatures may further improve expression, these also greatly reduce bacterial growth rates so a temperature of 26 °C was used for all Csp expressions. A period of overnight growth after induction gave more pronounced expression and greater cell yield (Figure 3.13).

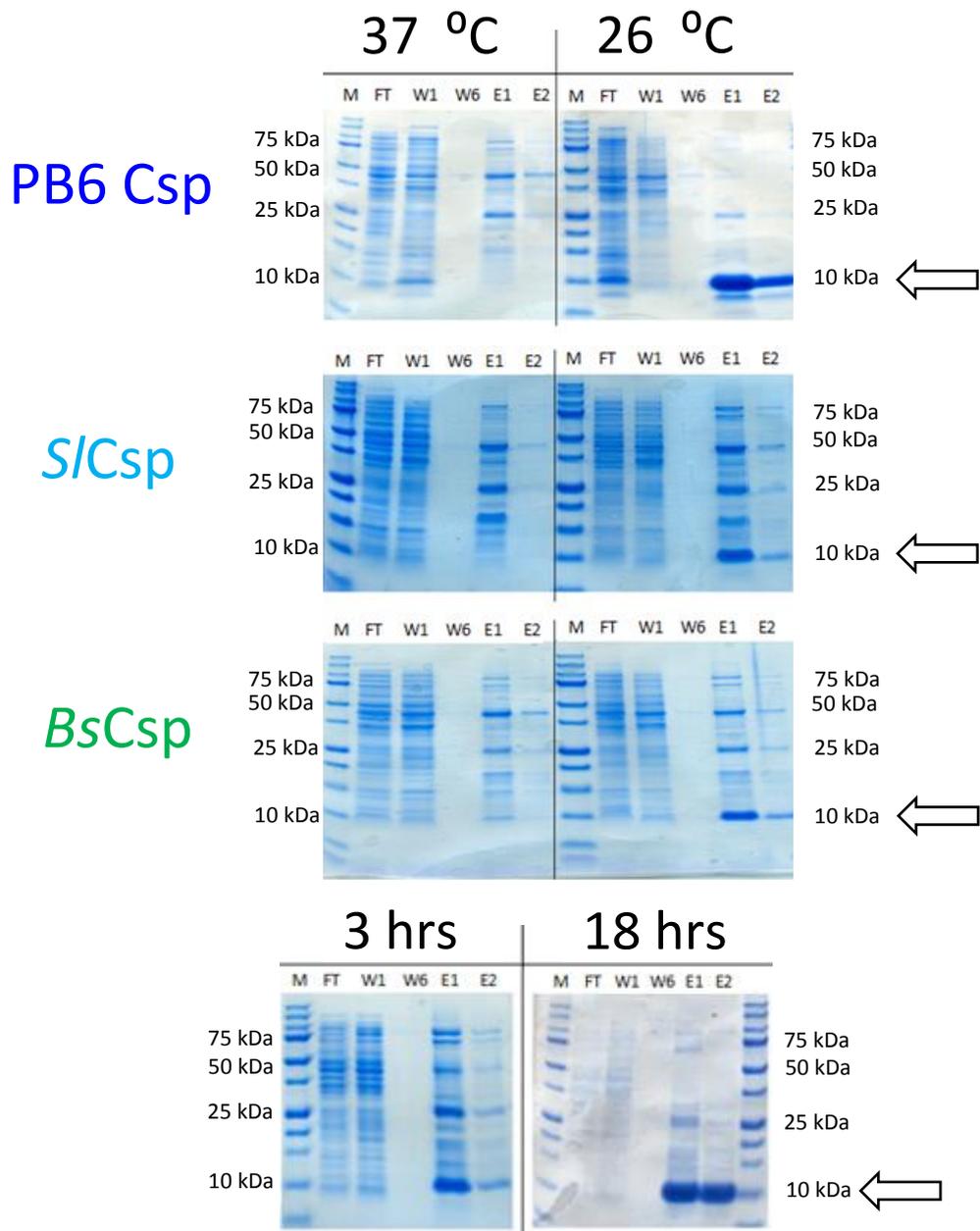


Figure 3.13: Ni-NTA pull-down assays contrasting expression of Csps at 37 °C and 26 °C. M represents protein markers, FT represents *E. coli* proteins which did not bind the Ni-NTA beads, W1 and W6 the solution from the 1st and 6th washes with buffer respectively and E1/E2 lanes show imidazole mediated elution of protein bound to the nickel beads. Arrow indicates expected mass of Csps. Top three graphs show expression of each Csp after 3 hours and bottom graph shows comparison of expression of PB6 Csp at 26 °C after 3 and 18 hours.

3.3 Purification of monomeric Csps

3.3.1 Ni-NTA affinity chromatography

Purification of the Csp monomers was performed by a three step process (as detailed in methods 2.4). An AKTA workstation was used to control flow rates within columns and SDS-PAGE (Sodium dodecyl sulphate polyacrylamide gel electrophoresis) gels used at the end of each purification step to identify the fractions containing over-expressed protein. 6 litres of *E. coli* were grown in each case with IPTG (1mM final concentration) used to induce expression (181). In the first purification step of a Ni-NTA affinity column, most proteins pass through the column without binding and then after washing away non-specifically bound proteins, imidazole is used to elute specifically bound proteins from the Ni beads. A mixture of proteins was present within the eluted fractions but the largest band observed was around 10 kDa which agrees with 8.5 kDa expected for Csps (Figure 3.14). Selected fractions were pooled for dialysis.

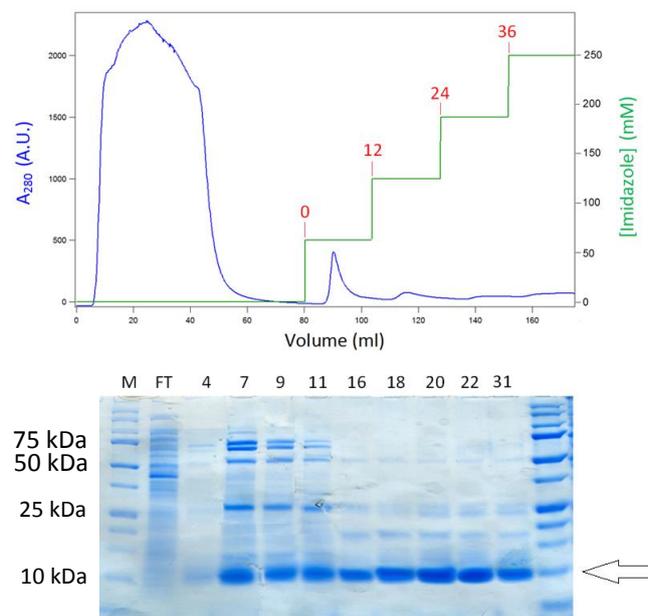


Figure 3.14: Ni-NTA affinity purification of PB6 Csp. Top: Typical elution profile for affinity purification of Csps using an AKTA prime workstation. The blue line and left axis represents the absorbance at 280 nm and the green line and right axis the concentration of imidazole applied to the column. Specific fraction numbers are indicated in red. Bottom: Analysis of selected collected fractions by SDS-PAGE. Expected Csp mass 8.5 kDa. FT represents sample which did not bind to the column diluted 1:19.

3.3.2 Ion-exchange chromatography

The pooled fractions from the Ni-NTA affinity column were dialysed into 50 mM salt buffer (see section 2.4.2). To remove DNA and RNA the sample was loaded on to a Resource Q anion-exchange column (For the need to perform this step see section 3.4.3). The Csp protein flows through the column without binding. When a gradient of increasing salt concentration is applied the negatively charged DNA remains bound to the column and is eluted with an increasing gradient of salt (all peaks from 50 ml onwards). An additional 40 ml of 1M NaCl is used to ensure all DNA is removed before further samples are added to the column. While the DNA like the protein has strong absorbance at 280 nm it does not produce a band on an SDS-PAGE gel (Figure 3.15)

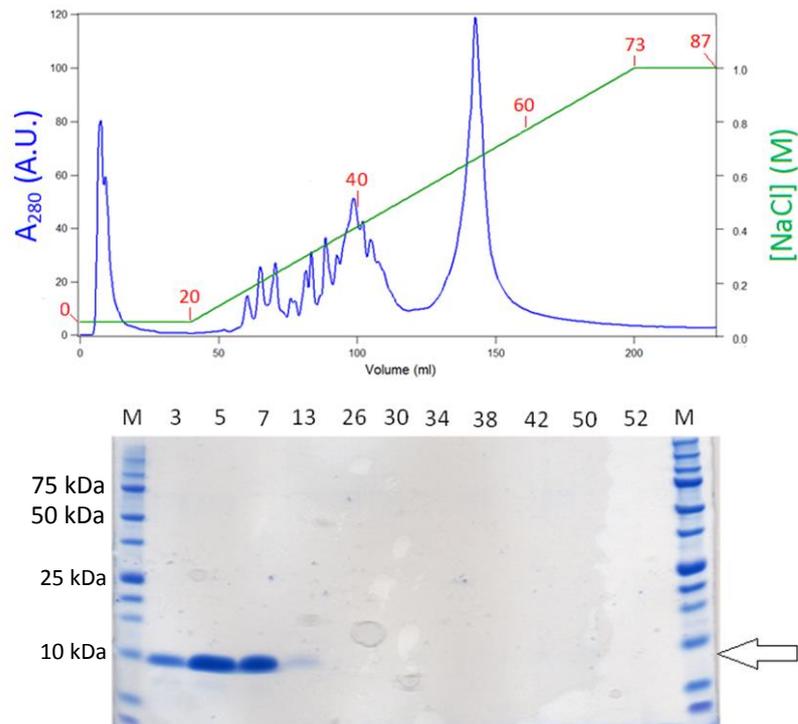


Figure 3.15: Ion exchange separation of PB6 Csp and SDS-PAGE showing fractions eluted from a resource Q column. Top: Blue line and left axis represents the absorbance at 280 nm. Green line represents the NaCl concentration applied to the column. Elution fractions were 2 ml in size for 40 ml and 3 ml subsequently with fraction numbers indicated on graph. Bottom: SDS-PAGE gel of selected fractions with arrow indicating mass of Csp

3.3.3 Size exclusion chromatography

Relevant fractions were pooled and the volume reduced to 3.5 ml. The sample was loaded on to a Sup75 size-exclusion column (see methods 2.4.3). Smaller molecules travel more slowly through the column as they can enter holes in the sephadex beads while larger molecules can't fit through these holes and so travel faster through the column. This means that molecules are separated by size with larger molecules eluted earlier and smaller molecules later. The elution volume is proportional to $-\log_{10}(M_r)$. A clear large peak could be seen and respective fractions separated and pooled (Figure 3.16)

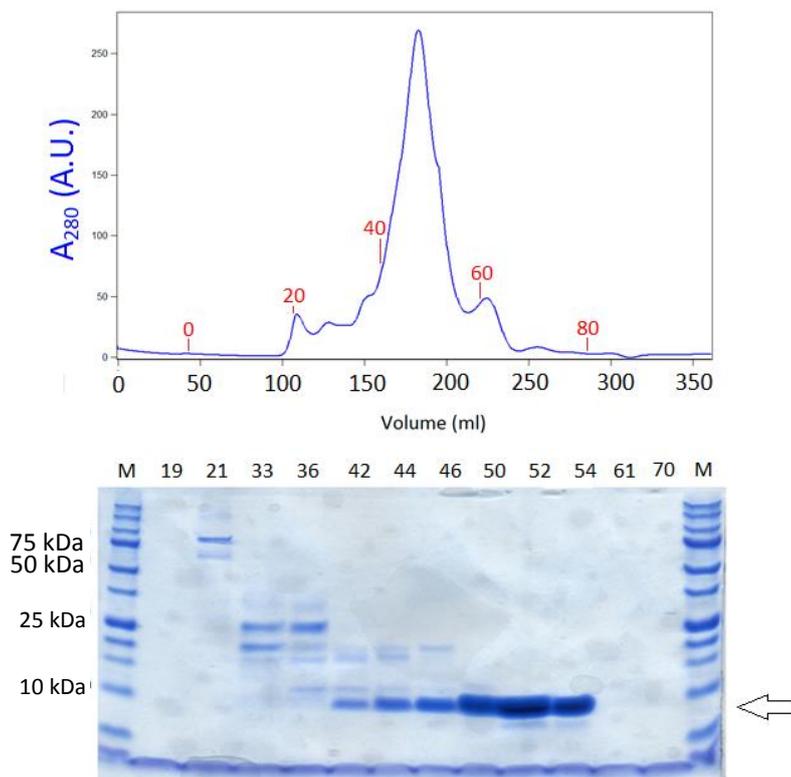


Figure 3.16: Size exclusion purification of PB6 Csp monomer using a Sup75 column. Top: Absorbance profile with blue line representing the absorbance at 280 nm. Bottom SDS-PAGE showing relevant fractions with arrow indicating Csp expected size of 8.5 kDa. Fractions were 3 ml in volume throughout.

3.3.4 Confirmation of protein identity by MS

To verify that the correct protein sequence had been produced, the atomic mass of the protein was confirmed with mass spectrometry experiments performed by Dr James Ault (Figure 3.17).

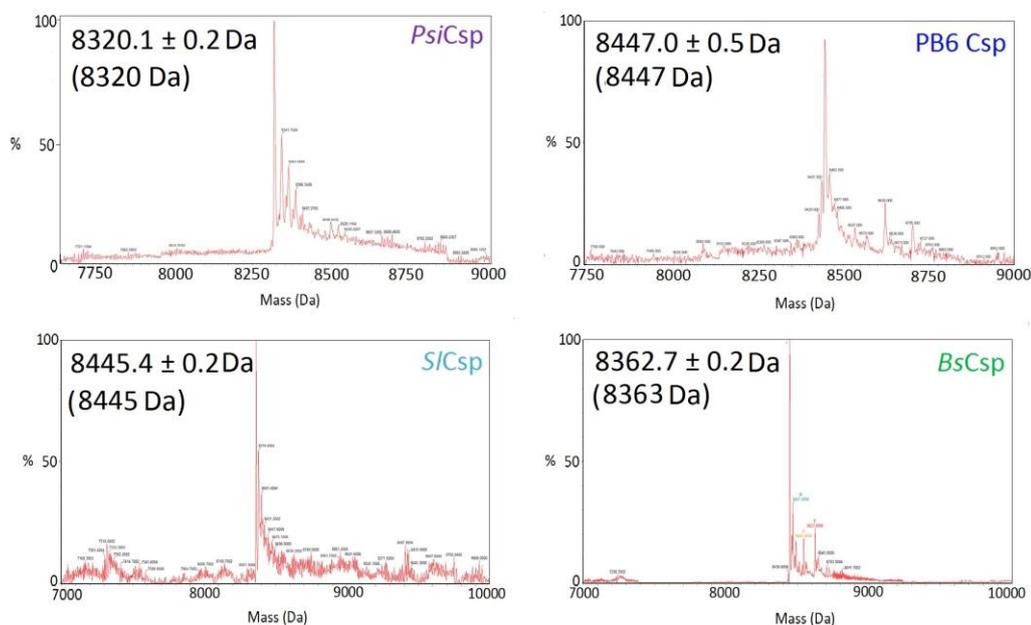


Figure 3.17: Mass spectra of the purified Csp with indicated masses of the largest peaks. Observed mass of the largest peak is indicated top left with the predicted mass based on protein sequence indicated in brackets.

3.4. Biophysical characterisation of the monomers

3.4.1 CD spectra to confirm folding

For all the Thermostability and mechanical stability measurements a consistent buffer containing 63 mM sodium phosphate was used. Phosphate buffer was selected as the pH of this buffer is known to be resistant against temperature-dependent pH changes and previously reported dimerization of *BsCsp* in some buffers does not occur in phosphate buffer (183). A concentration of 63 mM was selected for the sodium phosphate buffer as it has the same ionic strength as phosphate buffered saline allowing it to be a good mimic for cell cytoplasm.

If accurate parameters are to be obtained for the thermostability of the Csp's then the first step is to confirm that the protein is folded prior to conducting unfolding experiments. This was tested using circular dichroism spectroscopy (182). Each protein was diluted to 0.3 mg/ml in 63 mM sodium phosphate buffer and spectra recorded between 190 and 280 nm at 4 °C. All of the Csp spectra were seen to have peaks at 200 nm and 225 nm which is characteristic of folded Csp's (179). An interesting pattern was that the cold adapted Csp's all contained two CD peaks of similar intensity while the mesophilic *BsCsp* had a larger peak at 200 nm and a similar sized peak to the cold adapted forms at 220 nm. This may indicate some small differences in secondary structure between the Csp's. To test ability of the Csp's to unfold and refold, samples were heated to 85 °C to denature the protein and then the protein was seen to fully refold when the temperature was cooled down to 4 °C (Figure 3.18). This implies that correct folding is intrinsic to the protein sequence without the need for cellular chaperones and that aggregation of the thermally unfolded protein was minimal.

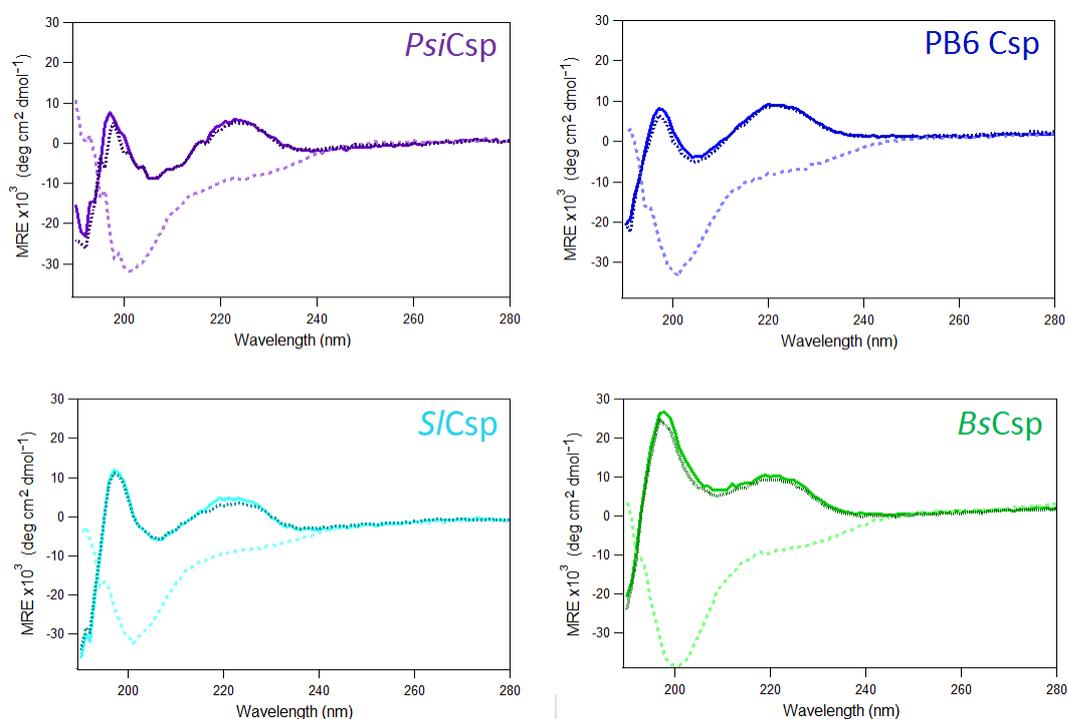


Figure 3.18: CD spectra between 190-280 nm validating folded state of the Csp's and re-folding capability. To account for the different masses of the proteins and to allow comparison between data mean residue ellipticity was plotted. Solid line indicates initial state at 4 °C, dashed line represents the Csp after 15 min at 85 °C and dotted line the Csp after 15 min returned to 4 °C.

3.4.2 Thermal denaturation followed by change in circular dichroism

It is relatively simple to compare the stabilities of protein homologues at the same individual temperature, however the objective of the study is to observe differences that occur over a range of temperatures. This requires the establishment of values for three different thermal parameters: the enthalpy of protein unfolding ΔH , the thermal mid-point of unfolding T_m and the change in specific heat capacity between folded and unfolded states ΔC_p .

Protein unfolding transitions can be followed by increasing the temperature of a Csp sample and monitoring loss of a folded protein signal. As a folded or unfolded protein is heated the relative rotation of left and right circularly polarised light (ellipticity signal) or fluorescence signal, changes in a linear fashion. This produces clear folded and unfolded protein baselines. Using circular dichroism, unfolding transitions of proteins can be followed by a large change in ellipticity at 222.6 nm as secondary structure elements are lost upon unfolding. Heating Csp at a constant rate of temperature increase gives a highly reproducible unfolding profile as shown in Figure 3.19.

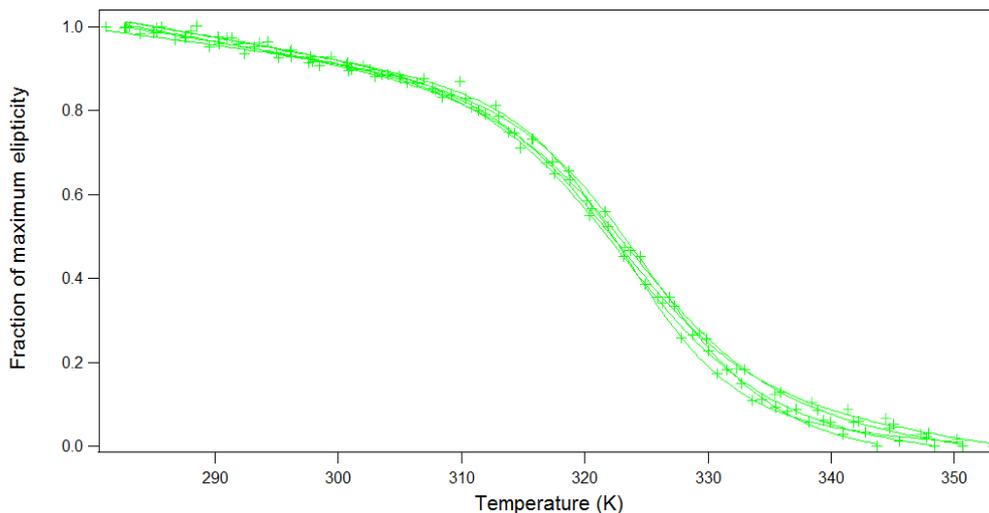


Figure 3.19: An overlay of 5 repeats of thermal unfolding of *BsCsp* followed by circular dichroism at 222.6 nm in 63 mM sodium phosphate buffer.

During heating the protein sample can be broken down in to 2 constituent parts: the fraction of the protein that remains in the folded form (F) and the fraction of the protein which has already unfolded (U) at a particular temperature. These can then be expressed in terms of the total protein and in terms of the equilibrium constant (Equations 3.5, 3.6 and 3.7).

$$\text{Eqn 3.5: } K_{eq} = \frac{U}{F}$$

$$\text{Eqn 3.6: } fF = \frac{F}{Total} = \frac{F}{F+U} = \frac{\frac{F}{F}}{\frac{F}{F} + \frac{U}{F}} = \frac{1}{1+K_{eq}}$$

$$\text{Eqn 3.7: } fU = \frac{U}{Total} = \frac{U}{F+U} = \frac{\frac{U}{F}}{\frac{F}{F} + \frac{U}{F}} = \frac{K_{eq}}{1+K_{eq}}$$

The observed fluorescence signal can be broken down in to the fraction of folded and unfolded protein present and their relative fluorescence intensities. S_F and S_U are the fluorescence or ellipticity intensities of the fully folded and unfolded states respectively. T represents temperature (in K).

$$\text{Eqn 3.8: } Signal = (f_F \times S_F) + (f_U \times S_U)$$

$$\text{Eqn 3.9: } Signal = \left(\frac{1}{1+K_{eq}} \times S_F\right) + \left(\frac{K_{eq}}{1+K_{eq}} \times S_U\right)$$

$$\text{Eqn 3.10: } Signal = \left(\frac{S_F}{1+K_{eq}}\right) + \left(\frac{S_U \times K_{eq}}{1+K_{eq}}\right)$$

$$\text{Eqn 3.11: } Signal = \frac{S_F + (S_U \times K_{eq})}{1+K_{eq}}$$

$$\text{Eqn 3.12: } K_{eq} = e^{\frac{-\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T}\right)}$$

$$\text{Eqn 3.13 } Signal = \frac{S_F + (S_U \times e^{\frac{-\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T}\right)})}{1 + e^{\frac{-\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T}\right)}}$$

As the pre-transition and post-transition baselines are not flat, a correction factor is added to approximate the folded and unfolded signal where a_F is the folded signal,

b_F the pre-transition gradient, a_U the unfolded signal and b_U the post-transition baseline gradient (Equation 3.14).

$$\text{Eqn 3.14: } f(T) = \frac{(a_F + b_F T) e^{-\frac{\Delta H}{R}(\frac{1}{T_m} - \frac{1}{T})} + (a_U + b_U T)}{1 + e^{-\frac{\Delta H}{R}(\frac{1}{T_m} - \frac{1}{T})}}$$

3.4.3 DNA contamination issues and using the barycentric median

CD thermal denaturation experiments monitor the loss of protein secondary structure upon unfolding while tryptophan fluorescence studies monitor loss of tertiary structure during the same process. If the unfolding transition is two-state as previously indicated by CD and DSC studies in the literature (170-180) then these two processes should occur concurrently giving closely matched values of T_m from both techniques. There have been limited studies predicting a folding intermediate for Csp (114). The initial results obtained for the Csp by tryptophan fluorescence were around 3 °C lower than the values obtained by CD. This difference was observed even though the rates of temperature increase in both techniques were optimised to be equivalent with the aid of a temperature probe. One notable aspect was some curvature present within the pre-transitional baseline of the fluorescence profiles. Csp binds to single stranded DNA or RNA as part of their biological function. DNA and RNA also have a significant fluorescence signal it could be deviation from linearity in the pre-transition baseline was as a result of DNA dissociation. The samples were therefore tested for polynucleotide contamination.

The samples were found to have an $A_{260/280}$ ratio of 1.1 - 1.2 which suggested that a significant level of DNA or RNA contamination was present. DNA and RNA give a strong fluorescence signal at 280 nm and so if present within the Csp sample could distort the observed T_m value. The concentrations of protein sample for each experiment were also calculated based on spectroscopic values so any U.V. active contaminant could cause a discrepancy between the perceived and actual Csp concentrations. Attempts were made to remove the nucleotide contamination by

binding the protein to a Ni-NTA affinity column and using a pulse of 2 M NaCl. A second strategy involved unfolding the Csps in 8 M urea and then refolding the protein by dialysis into fresh buffer to allow refolding of the Csp samples. Neither of these methods proved effective so an additional ion-exchange step was incorporated into the purification strategy and this was found to decrease the levels of DNA/RNA contaminant to below 2% w/w.

The monitoring of protein unfolding using fluorimetry is typically followed by exciting tryptophan residues with light at a specific wavelength and then measuring how the fluorescence signal at a single emission wavelength changes with temperature. This gives linear changes for the folded and unfolded variants with a large change in signal around the protein unfolding transition. Fluorescence values at a single wavelength are however fairly susceptible to noise. To avoid this, an alternative fluorescence measure can be used called the barycentric median (λ_{BCM}). λ_{BCM} utilises the intensity values from a whole range of wavelengths (in this case 320 nm to 380 nm) rather than one individual wavelength, averaging out instrument noise. When the protein unfolds the fluorescence peak shifts in wavelength and this can be followed by a change in λ_{BCM} . The value of λ_{BCM} for each fluorescence curve is calculated by multiplying each wavelength by the respective signal intensity at that wavelength and then dividing the sum of these numbers by the sum of intensities. The result (λ_{BCM}) is a wavelength figure which represents the 'centre of mass' of each fluorescence curve. Following a shift in this wavelength value against increasing temperature or denaturant concentration allows the transition from folded to unfolded protein to be followed as solvent exposure of the tryptophan changes (Figure 3.20). As the method cancels out signal intensity it is less susceptible to small differences in protein concentration. The extent of the change in λ_{BCM} during protein unfolding varied between the Csps due to the different number of tryptophan residues in each protein. A larger number of tryptophan residues give a more defined transition for *TmCsp* ($\Delta\lambda_{\text{BCM}} = 1.3$ nm) relative to PB6 Csp ($\Delta\lambda_{\text{BCM}} = 0.6$ nm).

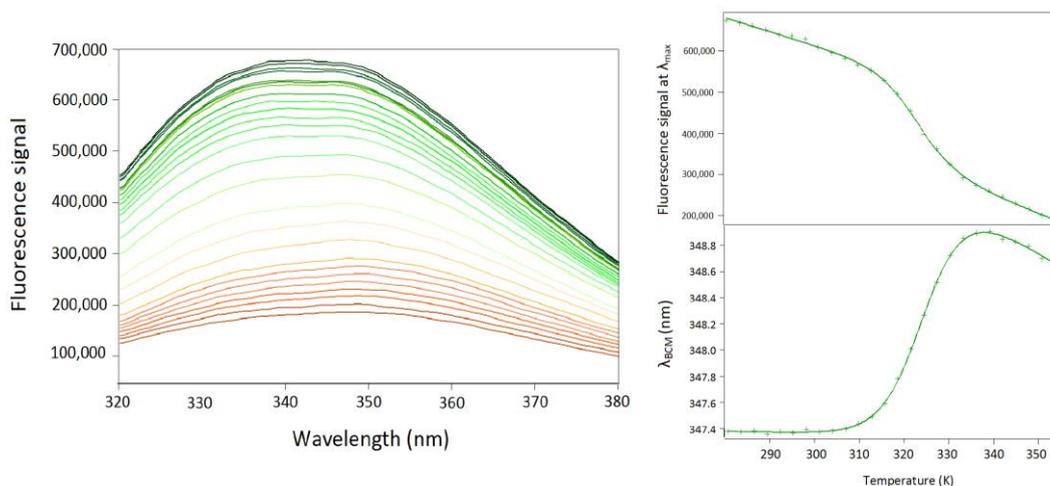


Figure 3.20: Fluorescence spectra of *BsCsp* from 7 °C to 82 °C in 63 mM sodium phosphate buffer. At 7 °C shown dark green and 82 °C shown dark red. A clear shift to higher wavelengths can be seen as protein unfolding occurs. The insert (top right) shows the changes followed at a single wavelength and (bottom right) follows the λ_{BCM} 320 nm - 380 nm against temperature.

The equations for the fitting of denaturation curves rely on both folded and unfolded states contributing equally to the fluorescence signal. This is however rarely the case and such differences can be seen in the relative signal heights of the folded (green) and unfolded (red) curves for *BsCsp* in Figure 3.20. If the fluorescence intensity of the unfolded state is substantially different to that of the folded state this distorts the perceived value of the mid-point T_m . To compensate for this, a correction factor called Q is applied which scales the contribution of the fluorescence intensity of the folded state to be equivalent to that of the unfolded state (see methods 2.7.3).

3.4.4 Thermal denaturation values obtained by circular dichroism and fluorescence

Thermal denaturation studies were then conducted on Csp's which had been through the extra DNA purification step (Figure 3.21) and the transitions fitted to equation 2.7. These showed that the values of T_m obtained by thermal unfolding monitored by change in λ_{BCM} for fluorescence data agreed with the T_m values

extracted from CD ellipticity data (see Table 3.3 for T_m values). The values conformed within 1 °C for all CspS apart from $TmCsp$ where the difference was 1.7 °C. With $TmCsp$ the larger differences appear to be caused by difficulties in establishing a reliable post-translational baseline as the T_m is relatively close to the maximum working temperature of the instruments. The agreement between the secondary structure unfolding reported by CD values and tertiary structure unfolding reported by fluorescence values now agrees with the literature theory of Csp unfolding being a simple two-state process.

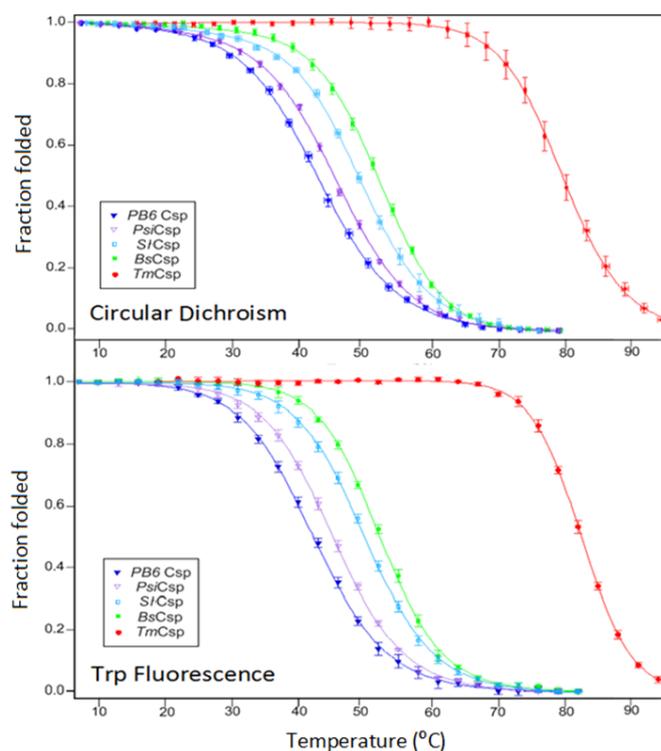


Figure 3.21: Thermal unfolding of CspS using CD and tryptophan fluorescence followed by λ_{BCM} in 63 mM sodium phosphate buffer. Top image displays CD data and bottom tryptophan fluorescence data. Error bars represent standard deviation between repeats. CD temperature error was obtained from standard deviation of probe measured temperatures.

Table 3.3: Comparison of stabilities between different CspS determined by CD and tryptophan fluorescence in 63 mM sodium phosphate buffer. Errors derive from the standard deviation of the fit obtained values between repeats

	Circular dichroism		Trp fluorescence (corrected $\lambda_{\text{BCM320-380}}$)		Reported T_m values
	$\Delta H^{\text{F-U}}$ (kJ/mol)	T_m (°C)	$\Delta H^{\text{F-U}}$ (kJ/mol)	T_m (°C)	T_m (°C) references
<i>PsiCsp</i>	-159 ± 9	46.1 ± 0.5	-155 ± 4	45.6 ± 0.6	-
PB6 Csp	-146 ± 6	43.2 ± 0.6	-150 ± 4	42.8 ± 0.4	-
<i>SlCsp</i>	-166 ± 6	49.7 ± 0.4	-171 ± 5	50.1 ± 0.7	-
<i>BsCsp</i>	-175 ± 5	52.5 ± 0.6	-181 ± 6	52.3 ± 0.3	50.3 (154) 52.9 (166) 53.4 (171) 53.6 (179) 53.4 (183) 53.8 (184)
<i>TmCsp</i>	-223 ± 18	80.2 ± 0.8	-271 ± 4	81.9 ± 0.9	85 (166) 81.1 (170) 82.0 (171)

As well as the thermal mid-point T_m Equation 2.7 also outputs a value of ΔH for the unfolding transition which describes how sharp the transition is. The Csp derived from *Thermotoga maritima* (*TmCsp*) showed a significantly greater enthalpy of unfolding of -223 kJ/mol than the mesophilic *BsCsp* at -174.8 kJ/mol and a greater T_m value of 80.2 °C compared to 52.5 °C (see Table 3.3). This conformed closely to data from previous studies and small differences would be expected as a result of differences in the buffers used and the presence or absence of a hexa-histidine tag. *TmCsp* is stable to higher temperatures than the other CspS and unfolds with a sharper transition (more negative ΔH). This is striking considering how similar the sequences and overall structures are to each other. A study by *Motono et al.* suggested additional stabilisation of the C-terminus which is the section of the CspS where unfolding initiates in the other CspS (185). Between the CspS the relative unfolding enthalpies and T_m values appear very closely coupled.

All of the cold-adapted CspS showed a reduced thermostability relative to the mesophilic Csp and less sharp unfolding transitions (less negative ΔH values). The

reductions in thermostability were small however compared to the differences between *BsCsp* and *TmCsp*. PB6 Csp showed the largest reduction in T_m value to around 9 °C below that of *BsCsp*. One surprising aspect was that *PsiCsp* derived from *Psychromonas ingrahamii* which is the Csp sequence from the bacterium adapted to the lowest optimum growth temperature of those studied, actually has a higher T_m value than that of the psychrotroph PB6 Csp.

3.4.5 Determination of the change in specific heat capacity upon unfolding

The third parameter required to determine the temperature dependence of protein thermostability is the change in specific heat capacity ΔC_p between folded and unfolded states. The amount of energy required to heat an unfolded protein by 1 °C is much higher than that of the folded state due increased exposure of amino acid residues to solvent which are buried in the folded protein. As protein homologues have very similar sizes and compactness of folding structure they are likely to have similar ΔC_p values (162)(186). Change in specific heat capacity ΔC_p correlates very strongly to the change in solvent exposed surface area upon unfolding. The change in surface area upon unfolding is also related to protein size. This may mean that smaller proteins are likely to have smaller ΔC_p values giving broader thermostability curves. This can allow small proteins which generally have lower stabilities to more easily maintain thermostability at high and low temperatures (186).

A comparison of 6 different thermophilic proteins with their mesophilic homologue presented by *H. Zhou* showed in all 6 cases a reduced value of ΔC_p (187). A reduced ΔC_p in hyperthermophilic proteins increases T_m by widening thermostability curves (154)(see section 3.1.2). The decrease in ΔC_p is believed to be due to maintenance of residual structure in the unfolded state which reduces the entropic benefits of unfolding (159). The degree of extension of unfolded proteins may depend on electrostatics, where an increase in matching charges and polar groups in thermophiles leads to increased repulsion giving a more extended conformation (186). The pattern of ΔC_p values for cold-adapted proteins is more variable with

some showing larger values than their mesophilic homologues and others lower values (see section 3.1.4).

Values of ΔC_p can be obtained directly through the differences in the amount of energy required to heat folded and unfolded protein through differential scanning calorimetry (DSC). These involve the heating of protein samples at a constant rate and observation of the amount of energy needed to perform this heating. The difference in the amount of energy required to heat the folded state and unfolded state by 1 °C represents ΔC_p . There can be problems in establishing correct baselines in DSC runs leading to large fluctuations in observed ΔC_p values (188). This is usually minimised through repeats and by conducting experiments at a range of pH levels.

An alternative method is to calculate ΔC_p by monitoring the dependence of ΔH and T_m on each other when the proteins are destabilised using denaturants or pH. This is because change in specific heat capacity is also the dependence of unfolding enthalpy ΔH on the thermal mid-point T_m when a protein is thermally unfolded. It is possible to determine this through destabilising the protein by adding chemical denaturants such as guanidine hydrochloride or urea which decrease both the values of ΔH and T_m obtained by protein thermal denaturation. The extent of destabilisation becomes more severe as the concentration of denaturant is increased. Performing thermal denaturation on protein samples in a series of different denaturant concentrations gives sets of pairs of ΔH and T_m from which a linear plot of ΔH against T_m can be produced. The gradient of this plot gives ΔC_p . The range of denaturant concentrations used must be selected carefully for each Csp as with concentrations too near or above the denaturant mid-point $[D]_{1/2}$ of the Csp it is not possible to establish an accurate pre-transitional baseline for accurate fitting of the thermal curves. Within this restriction the largest spread of denaturant concentrations were used to obtain the most reliable gradient. The thermal unfolding data with different denaturant concentrations is shown in Figure 3.22 with the fitting to equation 2.7 used to extract ΔH and T_m for each curve. The plot using these values to obtain ΔC_p is shown in Figure 3.23.

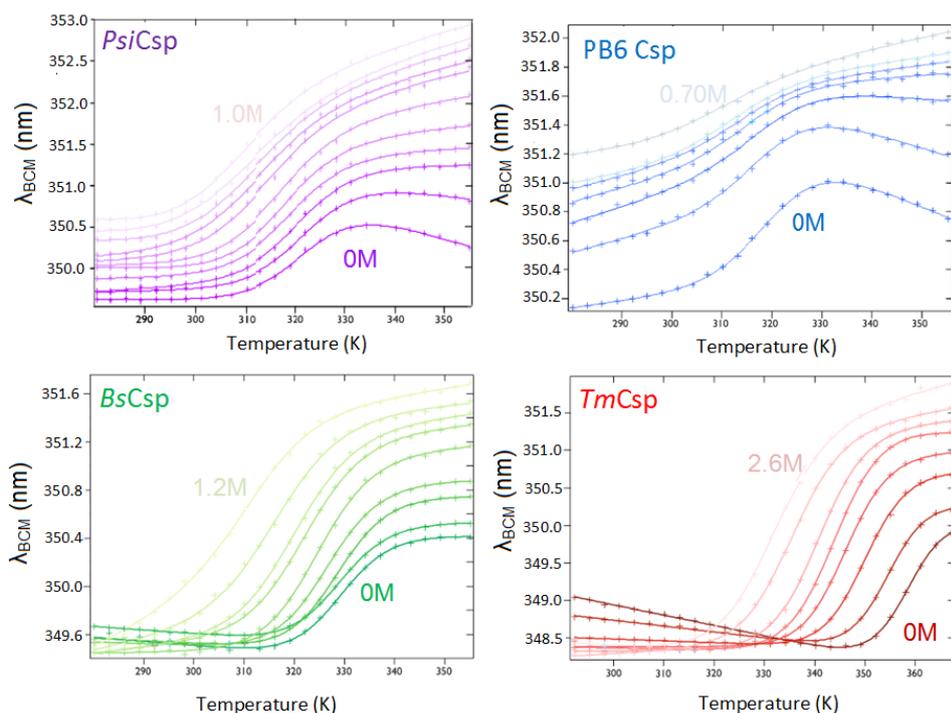


Figure 3.22: Thermal unfolding of Csp's with varying concentrations of GdnHCl measured by λ_{BCM} 320-380 nm. The bottom right curve (darkest shade) on each plot is without denaturant and each subsequent lighter curve is at increasing GdnHCl concentration. For *PsiCsp* and *PB6 Csp* the concentration of GdnHCl increases by 0.1 M with each curve, for *BsCsp* the spacing is 0.15 M and 0.4 M for *TmCsp* with a 0.2 M gap between the first curves

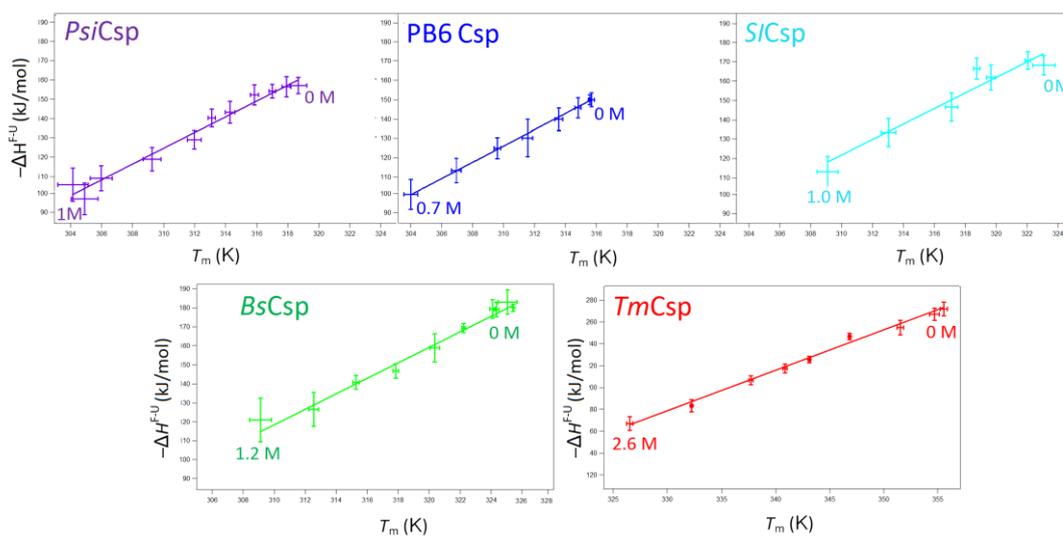


Figure 3.23: Plots of ΔH against T_m value pairs obtained with increasing concentration of GdnHCl from right to left. Errors for each data point derive from uncertainties in the fitting of the denaturant curve. A weighted linear fit was applied to the data with the significance of each data point inversely proportional to the size of the uncertainty of data points in the series.

3.4.6 Comparing thermostability parameters between the Csp

The obtained values for ΔC_p are greatest for PB6 Csp and lowest for the hyperthermophilic *TmCsp*. The differences between the ΔC_p values are relatively small (Table 3.4 and Figure 3.24, right) which may be due to the highly similar 3D structures of Csp β -barrels. The value of ΔC_p of 3.9 ± 0.1 kJ/mol/K is slightly below that of 4.5 ± 0.5 kJ/mol/K reported experimentally by *Wassenberg et al.* by DSC although the data they used for fitting ΔC_p contained substantial variation (170). It does however agree with changes in accessible surface area projections based on data from other proteins (186). This study showed a very small decrease in ΔC_p for *TmCsp* relative to *BsCsp* while a previous comparison had suggested an increase. In the comparison however *BsCsp* and *TmCsp* ΔC_p values were obtained by different techniques. The value for *TmCsp* was obtained from a fit to DSC data though there was a large error value of that fit.

Table 3.4: Thermostability parameters for the Csp. Errors in T_m and ΔH are the standard deviation between repeats and errors in ΔC_p derive from uncertainties in the weighted linear ΔH against T_m fits

Name	Category	T_m ($^{\circ}$ C)	ΔH^{F-U} (kJ/mol)	ΔC_p (kJ/mol/K)
<i>PsiCsp</i>	Psychrophile	45.3 ± 0.6	-155 ± 4	4.15 ± 0.25
<i>PB6 Csp</i>	Psychrotroph	42.8 ± 0.4	-150 ± 4	4.31 ± 0.42
<i>SlCsp</i>	Psychrotroph	50.1 ± 0.7	-170 ± 5	4.02 ± 0.31
<i>BsCsp</i>	Mesophile	52.3 ± 0.3	-181 ± 6	4.08 ± 0.15
<i>TmCsp</i>	Thermophile	81.9 ± 0.9	-271 ± 4	3.86 ± 0.16

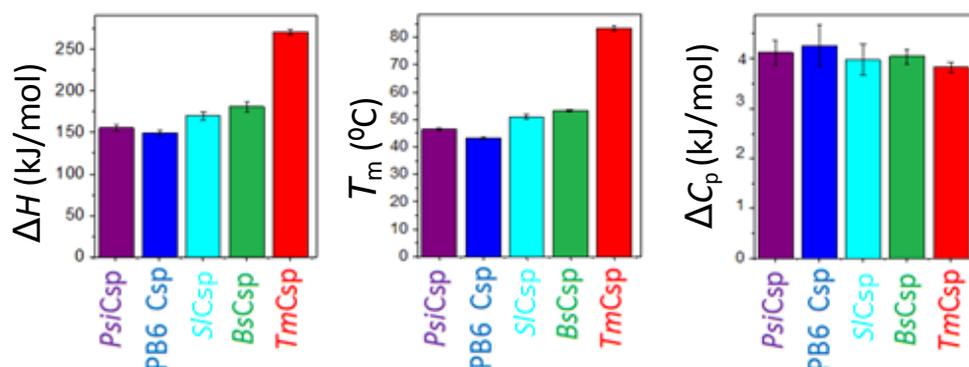


Figure 3.24: Comparison of thermostability parameters between the Csp. Values taken from fluorescence data

3.4.7 Thermostability temperature dependence curves

The thermostability parameters of ΔH , T_m and ΔC_p and their associated errors for each Csp are inserted into equation 2.8 and this is used to generate temperature dependence curves of $-\Delta G^{F-U}$ for each Csp (Figure 3.25). The T_m parameter indicates the temperature at which the curve intercepts the x-axis on the high temperature side. The width and height of the curves are determined by both ΔH and ΔC_p .

The thermostability of proteins follows a bell-shaped pattern shape mainly due the temperature dependence of the hydrophobic effect. Typically for proteins this centres around an optimum of around 10 °C which is also true for the Csp (53). The points at which the line crosses zero represent the limits of protein thermostability. At higher temperatures this is the point of thermal denaturation where molecules have more kinetic energy so there is an increased preference for unfolding (153). At low temperatures the limit of thermostability is the point of cold-denaturation. As temperature decreases, water adopts a more ordered structure which increases the magnitude of the thermodynamically favourable gain in entropy occurring upon protein unfolding. In this case the highly ordered arrangement of water molecules is disrupted by unfolding of the protein. Cold denaturation is however generally very difficult to observe as it usually only occurs at temperatures below the freezing point of water (189).

The error bars on the thermostability curves are very narrow at the high temperature side as this is constrained by the value of T_m which can be well characterised. The point of interception on the cold side however shows a larger uncertainty as it derived on a projection based on the output of the Gibbs-Helmholtz equation. It is not possible to perform experiments under 0 °C without the addition of salts or pressure which would further complicate analysis and place the protein under highly non-native conditions.

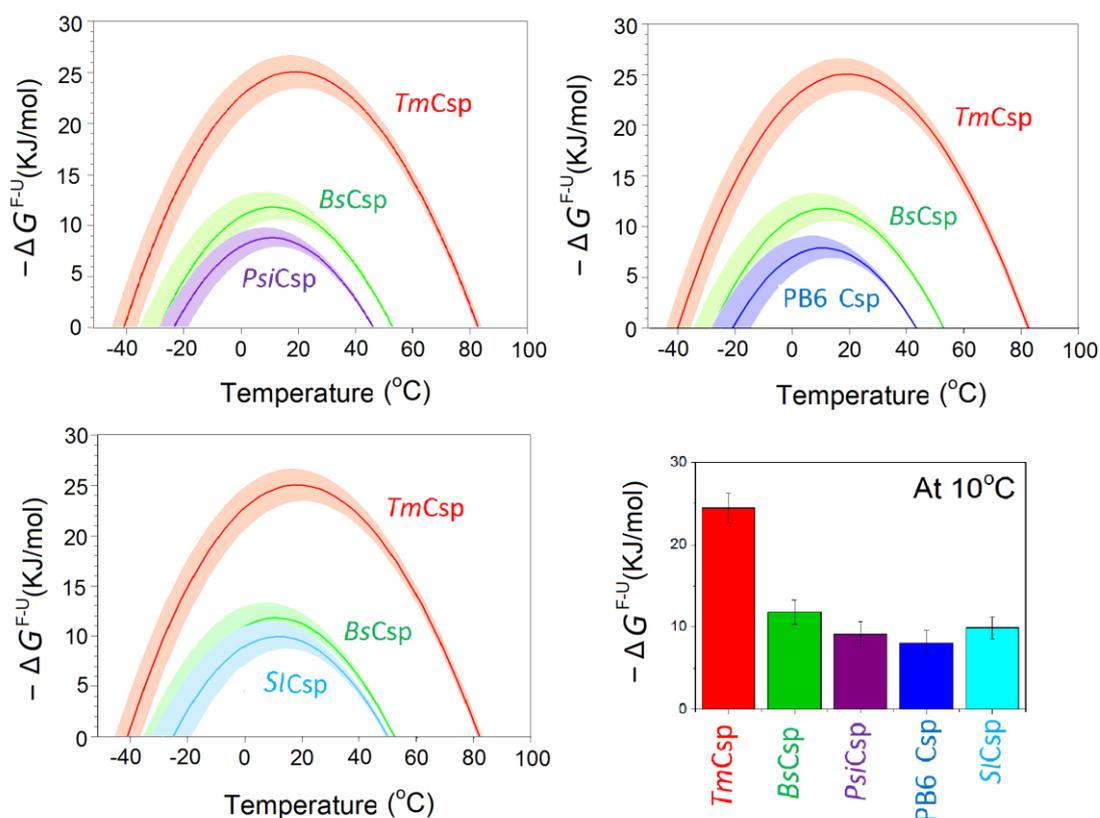


Figure 3.25: Thermostability profile for the Csp's in 63 mM sodium phosphate buffer derived from the Gibbs-Helmholtz equation. All profiles show TmCsp in red and BsCsp in green and include in addition PsiCsp (purple) top left, PB6 Csp (blue) top right and SlCsp (cyan) bottom left. Shaded region shows the uncertainty in the Gibbs-Helmholtz equation projections.

The Csp temperature dependence (Figure 3.25) shows the hyperthermophilic Csp variant TmCsp is more stable than the other Csp's across the range of temperatures including sub-zero temperatures. While hot-adapted proteins are more stable than

temperate or cold-adapted at high temperatures a common misconception regarding extremophilic proteins is that equivalently cold-adapted proteins would be more thermostable at lower temperatures. In fact what is seen here is that *TmCsp* is still more thermostable at low temperatures and remains folded to lower temperatures than the cold adapted Csp variants (16)(41). Similar findings have been reported for other protein families such the α -amylases (figure 1.5). The greater thermostability of *TmCsp* optimises it to remain folded at the high temperatures at which *Thermotoga maritima* lives but would potentially make it too rigid to function at more moderate temperatures. In a biotechnology context it may also show why thermophilic proteins are easier to use and store while the conditions psychrophilic proteins are stored at must be carefully considered.

The relative T_m value of *TmCsp* is around 30 °C above that of *BsCsp*, however the temperatures at which *TmCsp* is most stable at (termed T_s) is only 7.5 °C greater than *BsCsp* (Table 3.5). This mirrors findings for other homologue comparisons (41)(153)(164). The thermostability curve of the cold-adapted Csps did not exhibit a significant cold-temperature shift as the T_s values of between 9.9 °C and 10.7 °C for the cold-adapted Csps were very close to the T_s of the mesophilic *BsCsp* at 10.9 °C.

Table 3.5: Maximum thermostability parameters for the Csps in 63 mM sodium phosphate buffer

Csp Name	Temperature adaption	ΔG_{\max} (kJ/mol)	T_s (°C)
<i>PsiCsp</i>	Psychrophile	-8.8	10.3
<i>PB6 Csp</i>	Psychrotroph	-7.9	9.9
<i>SlCsp</i>	Psychrotroph	-9.9	10.7
<i>BsCsp</i>	Mesophile	-11.8	10.9
<i>TmCsp</i>	Thermophile	-25.1	18.4

The cold adapted Csp's show only modest decreases in maximum thermostability ΔG_{\max} relative to the mesophilic *BsCsp* with a 33% reduction for PB6 Csp, 26% for *PsiCsp* and only 16% for *SlCsp* (Table 3.5). This may be because of the need to maintain thermostability as further destabilisation could lead to transient unfolding of the Csp which would be selected against through evolution. Another factor is that destabilisation of cold-adapted proteins also reduces the range of temperatures over which the proteins remain stable suggesting a reason that many psychrophilic organisms can only survive in a narrow range of growth temperatures. This may make psychrophiles highly susceptible to changes in environmental temperature which potentially presents a real risk to psychrophilic organisms in the future from global warming.

Of the Csp's studied *PsiCsp* derives from the bacteria with the lowest optimal growth temperature and also lowest minimal growth temperature of $-12\text{ }^{\circ}\text{C}$ so *PsiCsp* may be expected to show the lowest thermostability. In terms of both maximum $\Delta G^{\text{F-U}}$ value and T_m PB6 Csp was found to be least thermostable. In sequence terms the number of charged residues is known to be significant in temperature adaptation and indeed in this case *PsiCsp* has the fewest charged residues (Table 3.1). Another factor in cold adaption is the number and position of glycine residues as glycine only contains a hydrogen atom as a variable group so is highly flexible. The sequence of PB6 Csp though contains a cluster of 3 adjacent glycine residues so this region may serve a flexible hinge region playing a significant role in the lower thermostability of PB6 Csp (87).

3.4.8 Which parameters are important in the thermostability differences between the Csp's?

The thermostability curves of the Csp's show very different values of T_m and ΔG_{\max} despite the overall topology of the Csp being conserved. Section 3.1.2 showed how thermostability curve profiles of extremophilic proteins can adapt through changes

to T_s , ΔC_p and ΔH (see figure 3.2). The T_s values of the Csps are within 9 °C for all Csps despite the very different T_m values. A lower ΔC_p value can expand the thermostability of thermophilic protein to higher temperatures however in this case the values of ΔC_p are also very similar for all the Csps and the small differences are not sufficient to significantly affect the thermostability profile (Table 3.4). This means that the crucial factor in the adaptation of Csps to different temperatures is changes in ΔH . This pattern is similar to that seen in other protein families (41). *TmCsp* is able to achieve a greater T_m than the other Csps through increased number of electrostatic interactions and increased hydrophobic packing whilst maintaining the overall protein topology.

3.4.9 Comparing the thermostabilities of proteins at the growth temperatures of their respective organisms

While clear comparisons can be made between the thermostabilities of the Csps at individual temperatures this gives little consideration of the conditions that these Csp proteins would be found at in nature. If we consider the thermostability of the proteins at the temperature at which their host bacteria grow at then the order of stabilities is reversed with *TmCsp* at the very right of its thermostability curve being only very marginally stable and the cold adapted forms being near their optimal thermostability.

To allow direct comparisons of the stabilities of Csps at the optimal growth temperatures (T_{opt}) of each organism the thermostability curves for each Csp was moved so that the optimal growth temperature was 0 °C for each. As Csps are upregulated below the optimum temperature of the organism (under conditions of cold shock) the thermostabilities above and below T_{opt} were investigated. On Figure 3.27 the dotted line represents the respective optimum growth temperature of the bacteria (*PsiCsp* $T_{opt} = 5$ °C, *PB6 Csp* $T_{opt} = 22$ °C, *BsCsp* $T_{opt} = 37$ °C, *TmCsp* $T_{opt} = 80$ °C). The right hand side shows what happens to the thermostability above T_{opt} and

the left side below T_{opt} . One point which seems extremely interesting is at about 8 °C below the T_{opt} of each species where the thermostabilities of the Csp's are nearly identical (Figure 3.26). This suggests that the vast differences in the protein thermostability curves could actually be case of natural selection towards optimising the proteins towards similar thermostabilities at their operating temperature. The idea of studying protein thermostability at room temperature may actually conceal the true concept of protein temperature adaption.

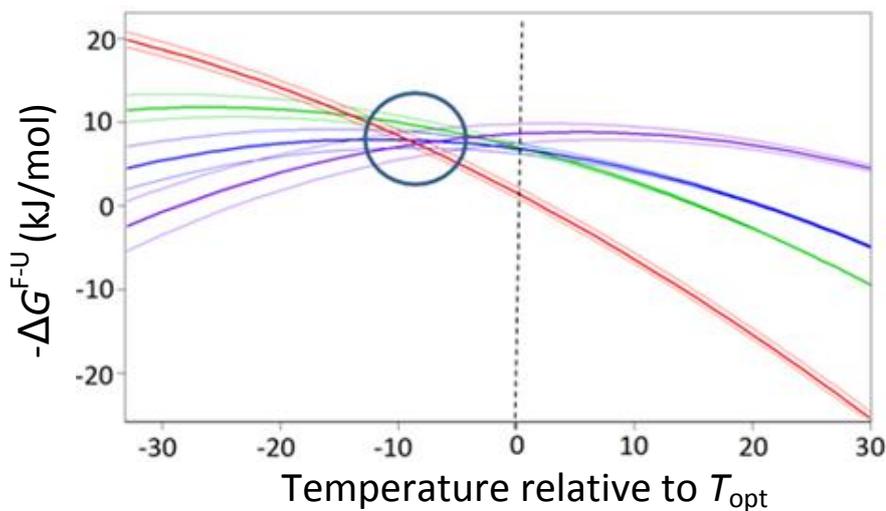


Figure 3.26: Comparisons of Csp thermostabilities relative to the optimum growth temperatures of their host bacteria. *PsiCsp* $T_{opt} = 5$ °C (purple), *PB6 Csp* $T_{opt} = 22$ °C (blue), *BsCsp* $T_{opt} = 37$ °C (green) and *TmCsp* $T_{opt} = 80$ °C (red).

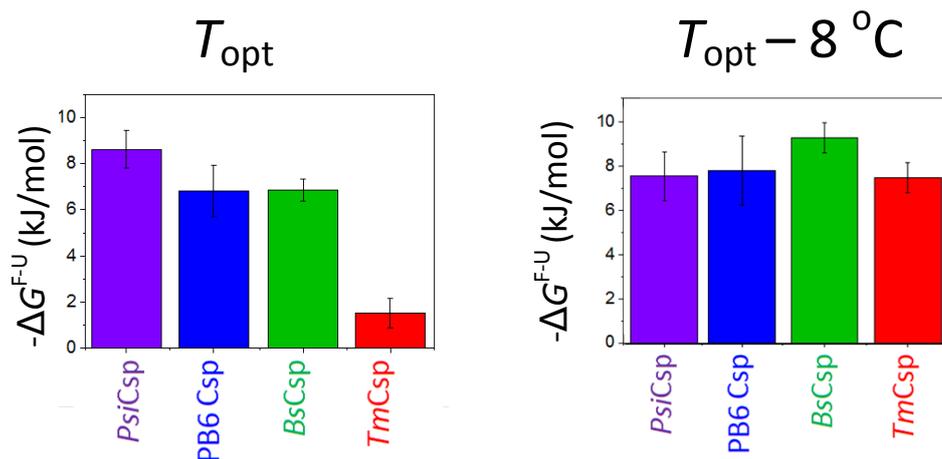


Figure 3.27: Histogram showing the relative thermostabilities of the Csp's at the respective optimum growth temperatures of their host bacteria and at 8 °C below that temperature.

The point of intersection of the curves (Figure 3.26) is not exactly at the T_{opt} for each organism falling at about 8 °C below this temperature. The T_{opt} however, may not reflect the actual growth temperature of the organism, as values of T_{opt} are generally obtained in nutrient rich media rather than in environmental conditions. This is especially applicable for psychrophiles which are likely to be outcompeted by more mesophilic bacteria at moderate temperatures. It is also worth noting that due to their function, Csp's are not expressed at T_{opt} but only at a reduced temperature. Only very limited studies have been performed into the temperatures each bacteria expresses its respective Csp's at, so drawing connections between the thermostability findings and Csp expression temperatures is complex.

Performing a similar correlation with T_{opt} using α -amylase data shows that there is no exact location of equal thermostability relative to T_{opt} however around +10 °C there is a region of similar thermostability (41). This point of similar thermostability is higher than T_{opt} in contrast to Csp's where it was lower than T_{opt} . Extremophilic adaptation however, is generally a compromise where adaption strategies represent trends rather than strict rules. Few studies include a cold-adapted variant and often thermostability values were not determined by the same method.

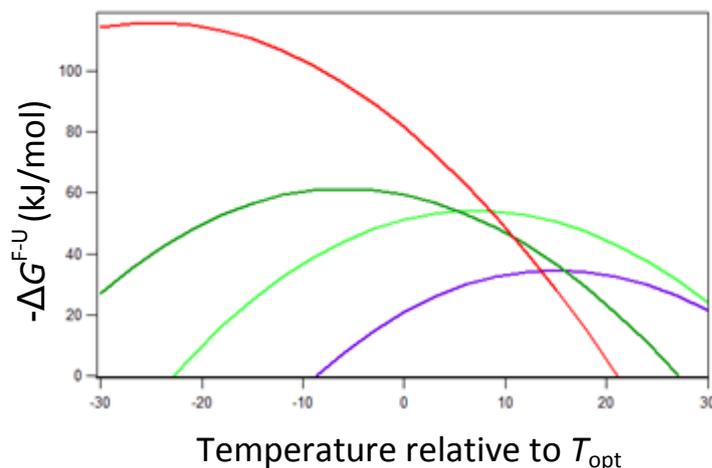


Figure 3.28: Thermostability of α -amylases compared relative to the optimum growth temperatures of their host organisms. Comparison of the thermostabilities of α -amylase enzymes *Thermobifida fusca* (TFA) in red, pig pancreas (PPA) in green, *Drosophila melanogaster* in light green (DMA) and *Pseudoalteromonas haloplanktis* (AHA) in purple. Data derived from (41).

3.4.10 Chemical denaturation using GdnHCl and urea

As discussed above, the generation of thermostability curves using ΔH , ΔC_p and T_m is capable of predicting ΔG^{F-U} at a range of temperatures. To test these predictions the ΔG^{F-U} values were measured using chemical denaturation experiments at 10 °C and 23 °C. To do this the concentrations of a denaturant in the protein sample can be set allowing the folded and unfolded protein states to reach equilibrium. The amount of folded protein present in different concentrations of denaturant can be monitored by fluorescence due to the different fluorescence levels of aromatic residues in the folded and unfolded protein states. This can be fitted with equation 2.9. The mid-point of chemical denaturation is temperature dependent so it is important to closely regulate temperature. The relative ability of the different Csp to resist denaturation by GdnHCl (shown in Figure 3.29 and Table 3.5) appears to very closely mimic the relative thermostabilities to temperature. *TmCsp* remains folded to far higher concentrations of guanidine hydrochloride than the other Csp. The cold-adapted Csp show only slightly reduced resistances to denaturants than *BsCsp*. As expected a slight reduction in all the Csp was observed at 23 °C relative to 10 °C. The values obtained at 23 °C agree with those from (166) which gave values of ΔG^{F-U} of -11.3 kJ/mol with a mid-point of 1.5 M for *BsCsp* and ΔG^{F-U} of -26.2 kJ/mol with a mid-point of 3.3 M for *TmCsp*. These also agree for *BsCsp* with -11.3 kJ/mol from (179) and -11.4 kJ/mol (180).

The *m*-value represents how ΔG^{F-U} relates to the mid-point $[D]_{1/2}$ of chemical denaturation. There is a small increase in the *m*-values between the Csp going from the psychrophilic to hyperthermophilic Csp. Models of the accessible surface area of 45 folded proteins derived from crystal structures and their equivalent unfolded protein chains showed that the *m*-value strongly correlates to the change in solvent accessible surface area occurring upon protein unfolding (186). *m*-values have been shown to have close correlation between GdnHCl and urea equilibrium denaturation experiments of 0.9 showing that the effects are independent of the denaturant used (186).

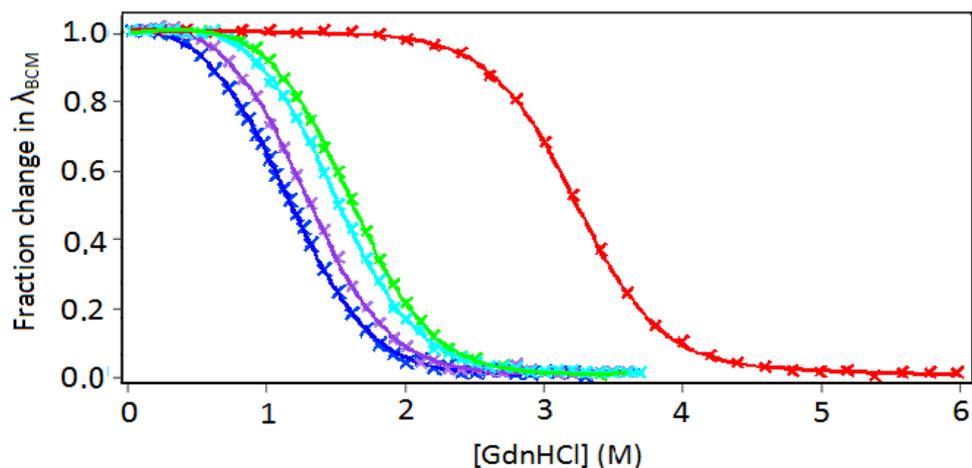


Figure 3.29: Equilibrium curve showing the unfolding of Csp in 63 mM sodium phosphate buffer in response to guanidine hydrochloride denaturant at 10 °C.

Table 3.6: Parameters derived from fitting of equilibrium denaturation curves performed with guanidine hydrochloride at 10 °C and 23 °C in 63 mM sodium phosphate buffer. Errors indicated derived from fitting.

Temperature	protein	ΔG^{F-U} (kJ/mol)	m (kJ/M/mol)	$[\text{GdnHCl}]_{1/2}$ (M)
10°C	<i>PsiCsp</i>	-9.2 ± 0.8	7.5 ± 0.4	1.24
	<i>PB6 Csp</i>	-7.9 ± 0.6	7.4 ± 0.4	1.08
	<i>SlCsp</i>	-11.0 ± 0.6	7.6 ± 0.3	1.45
	<i>BsCsp</i>	-11.9 ± 0.4	7.6 ± 0.2	1.56
	<i>TmCsp</i>	-24.8 ± 1.7	7.8 ± 0.4	3.21
23°C	<i>PsiCsp</i>	-8.3 ± 0.6	7.5 ± 0.3	1.12
	<i>PB6 Csp</i>	-7.0 ± 0.7	7.3 ± 0.4	0.94
	<i>SlCsp</i>	-10.6 ± 0.7	7.5 ± 0.4	1.41
	<i>BsCsp</i>	-11.2 ± 0.7	7.6 ± 0.3	1.49
	<i>TmCsp</i>	-25.9 ± 0.6	7.8 ± 0.1	3.28

To confirm if any specific effects were due to guanidine hydrochloride, equivalent experiments were also performed with urea which is a less powerful denaturant. The same hierarchy of chemical stabilities was observed with the Csp's although *TmCsp* was not included as it cannot be unfolded by 9 M urea and greater concentrations can exceed the solubility limit of urea.

Table 3.7: Parameters derived from fitting of equilibrium denaturation curves performed with urea at 10 °C in 63 mM sodium phosphate buffer. Errors indicated derived from fitting.

protein	ΔG^{F-U} (kJ/mol)	m (kJ/mol/M)	$[\text{urea}]_{1/2}$ (M)
<i>PsiCsp</i>	-8.6 ± 0.5	2.9 ± 0.1	2.96
PB6 Csp	-7.7 ± 1.0	2.8 ± 0.2	2.73
<i>BsCsp</i>	-11.3 ± 0.8	3.0 ± 0.2	3.88

The values of ΔG^{F-U} obtained from fitting of the equilibrium curves agree strongly with those observed from the Gibbs-Helmholtz equation projections (Figure 3.30) meaning that the results from both methods reinforce each other. The values from stopped-flow experiments were also in agreement (section 4.2.1).

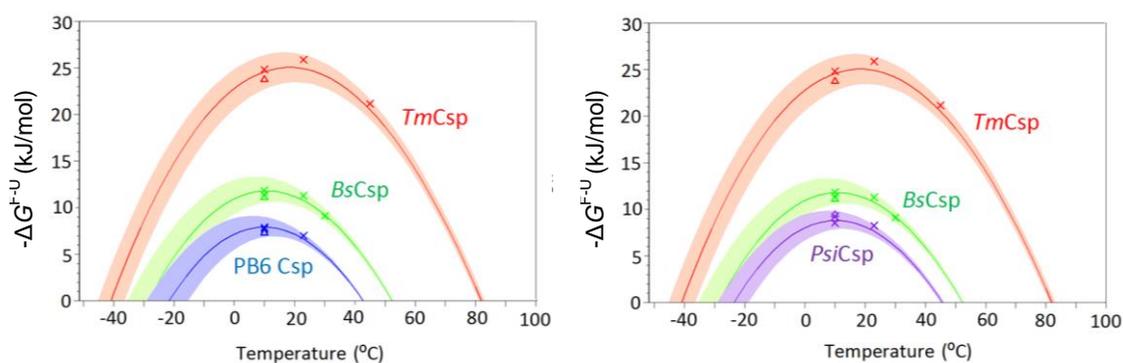


Figure 3.30: Comparison of Gibbs-Helmholtz projection with chemical denaturation approaches: Plots showing equilibrium curve values (cross symbol) and stopped flow values (triangles) overlaid on the Gibbs-Helmholtz equation projections. The left graph includes PB6 Csp and the right graph *PsiCsp*.

3.4.11 Insights into the properties of extreme hot and extreme cold-adapted Csps

The optimum growth temperatures of each organism (T_{opt}) correlates with the maximum thermostabilities of the Csps (Figure 3.31). At lower temperatures the pattern is more mixed and the data is sparse. Here the thermostability of *PsiCsp* from an extreme psychrophile is greater than that of PB6 Csp derived from a psychrotroph. This raises the question of what an extreme psychrophile needs to achieve to access further into the sub-zero range. As temperature decreases the arrangement of water molecules becomes more ordered effect eventually leading to failure of the hydrophobic effect termed cold denaturation. High salt conditions often found in cold environments may offer some protection against this but cold denaturation still limits the temperatures at which proteins can remain folded. To remain folded at extreme cold temperatures a protein may require increased thermostability. This may mean that extreme psychrophilic proteins actually need to be more thermostable than mildly cold-adapted proteins and so could surprisingly be more similar to mesophilic proteins. This perspective though does not factor in the kinetics and protein folding issues at low temperatures. It may be the balance of coping with all these factors that is important to protein function at cold extremes.

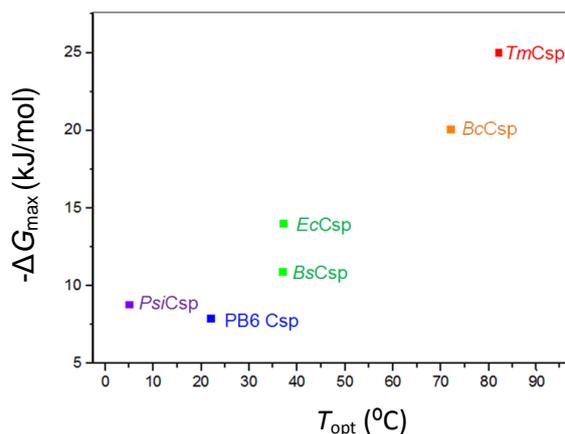


Figure 3.31: Plot of thermostability of the Csps against the optimum growth temperature of their host organism. *E. coli* CspA and *BcCsp* data added from literature (171).

3.5 Cold shock protein variants

To attempt to elucidate more about what features make *TmCsp* highly thermostable, three protein mutants were produced by Dr Toni Hoffmann. The first was a variant of *BsCsp* with the 3 mutations E3R, S48E and T64K. The other 2 were variants of *TmCsp* with mutations V62A and L40A. All three mutants were purified by the same method (detailed in section 2.4). The properties of these mutants could be analysed by the same method used for the temperature adapted Csps (sections 2.7 and 3.4) to identify differences between wild type and mutant thermostabilities. For the advantages and disadvantages of this approach see section 1.3.1.

3.5.1 Studies of a charged triple mutant of *BsCsp*

One of the key patterns shown in the comparative analysis of the Csp sequences was an increase in the numbers of charged residues present in the thermophilic and hyperthermophilic Csp sequences. A set of 3 of these additional charged residues present in the hyperthermophilic *TmCsp* but not in the mesophilic *BsCsp* are 3R, 48E and 64K which are thought to constitute a highly thermostabilising ionic cluster region. These charged residues are proposed to produce a clamp through the formation of salt bridges between β -strands 1 and 4 and a second new salt bridge between β -strands 4 and 5 (Figure 3.32). This provides a linkage between β -sheet 1 and 5 which is suggested to increase the thermostability of *TmCsp* (190). Dr Toni Hoffmann synthesised and purified a mutant of the *BsCsp* which contained these 3 substitutions E3R, S48E and T64K from the hyperthermophilic *TmCsp*. This mutant will be referred to as the charged triple mutant (CTM). Here thermal and chemical denaturation studies of *BsCsp* CTM are performed to show to what extent changing these 3 residues converts the properties of *BsCsp* to one which closer resembles the hyperthermophilic *TmCsp*.

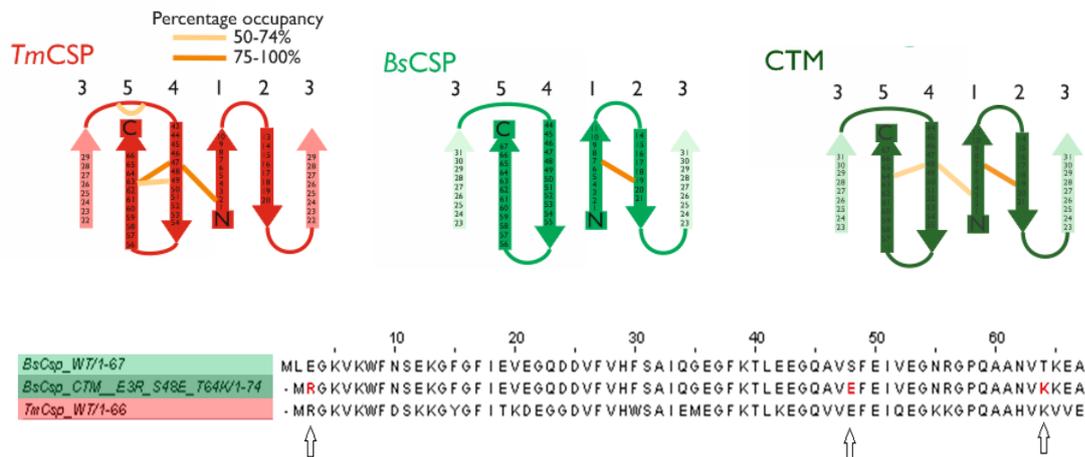


Figure 3.32: Secondary structure and sequence comparisons of *BsCsp* WT, *BsCsp* CTM and *TmCsp*.

Top: Secondary structure topology with locations of proposed salt bridges indicated in orange.

Bottom: Sequence alignment displaying mutations of CTM from *BsCsp* indicated by arrows and red highlighting

The *BsCsp* CTM showed an increased resistance to both temperature and chemical denaturant relative to *BsCsp*. A substantial increase in thermal mid-point of 10 °C was observed from 52 °C to 62 °C which is around 1/3 of the thermostability gap between *BsCsp* and *TmCsp* (Figure 3.33 top and Table 3.7). In terms of chemical stability a similar pattern is observed to the thermostabilities with the CTM more chemically stable than *BsCsp* but with only an increase of -2 kJ/mol in ΔG^{F-U} above WT *BsCsp* relative to an increase of -13 kJ/mol for *TmCsp* (Figure 3.33 bottom and Table 3.8).

For the natural Csp's an increase in T_m was consistently met with a corresponding increase in ΔH . With the CTM however, despite a significant increase in T_m , no significant increase in ΔH was observed. It could be that the thermostabilising effects of the salt bridges on unfolding enthalpy rely on the surrounding residues present in *TmCsp* and not just the salt bridges alone. The salt bridges may have an entropic role in *TmCsp* by tethering the β -sheets together. The increased enthalpy of *TmCsp* is also governed by increased hydrophobic packing effects as well as electrostatic interactions.

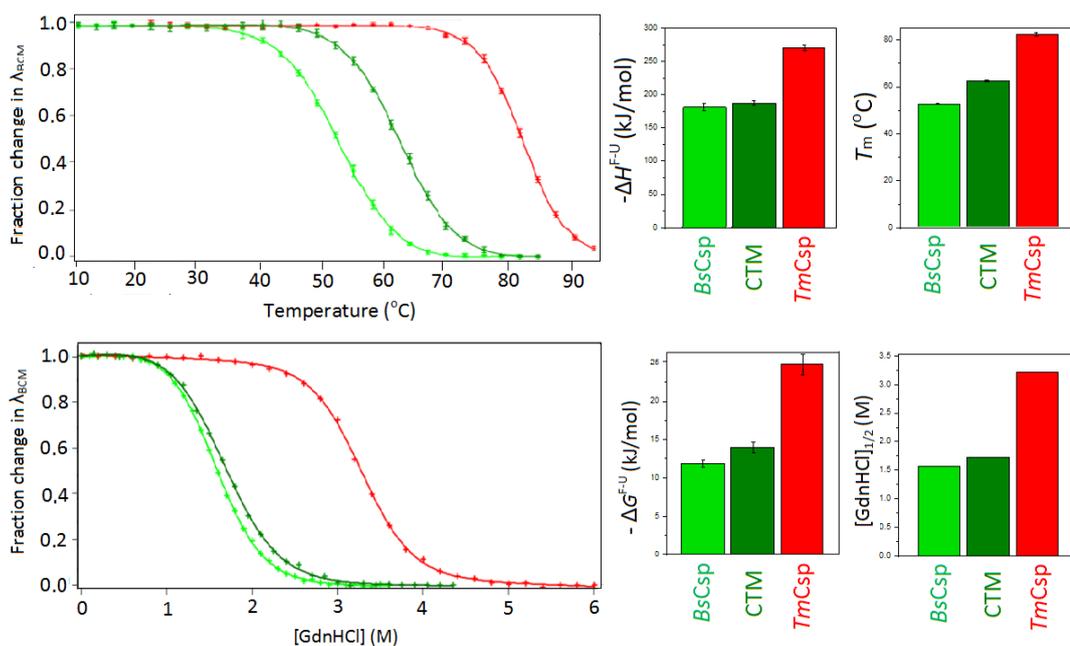


Figure 3.33: Thermal unfolding of *BsCsp* CTM compared with *BsCsp* and *TmCsp*: Thermal denaturation of *TmCsp* (red), *BsCsp* (green) and the charged triple mutant of *BsCsp* (dark green) followed by change in λ_{BCM} by fluorescence. Bar charts illustrating change in ΔH and T_m are shown on the right. Bottom: Chemical denaturation at 10 °C. Change in $-\Delta G^{\text{F-U}}$ and denaturant mid-point are illustrate in bar charts right

Table 3.8: Parameters obtained from fitting of thermal denaturation of *TmCsp*, *BsCsp* and *BsCsp* CTM

	CD		PTI (corrected λ_{BCM})	
	$\Delta H^{\text{F-U}}$ (kJ/mol)	T_m (K)	$\Delta H^{\text{F-U}}$ (kJ/mol)	T_m (K)
<i>BsCsp</i>	-175 ± 5	52.5 ± 0.6	-181 ± 6	52.3 ± 0.3
<i>BsCsp</i> CTM	-172 ± 7	62.4 ± 0.6	-187 ± 3	62.2 ± 0.4
<i>TmCsp</i>	-223 ± 18	80.2 ± 0.8	-271 ± 4	81.9 ± 0.9

Table 3.9: Parameters extracted from chemical denaturation of the Csp's at 10 °C by GdnHCl

Protein	ΔG^{F-U} (kJ/mol)	m (kJ/mol/M)	$[\text{GdnHCl}]_{1/2}$ (M)	$\Delta\Delta G^{F-U}$ from <i>BsCsp</i> (kJ/mol)
<i>BsCsp</i>	-11.9 ± 0.4	7.6 ± 0.2	1.56	
CTM	-14.0 ± 0.7	8.0 ± 0.3	1.73	-2.1
<i>TmCsp</i>	-24.8 ± 1.4	7.8 ± 0.4	3.21	-12.9

The ΔC_p value of CTM was similar to that of *TmCsp*. The unfolding enthalpy against T_m plot appeared to shift to have similar values to *TmCsp* with high concentration of GdnHCl. A thermostability profile was constructed as before for the CTM protein and this contrasted with those of *BsCsp* and *TmCsp* (see appendix Figure S3 for intermediate ΔC_p determination plot). The maximum $-\Delta G^{F-U}$ of CTM was increased by a small amount over that of *BsCsp* but this was minor in comparison to the far higher maximum ΔG^{F-U} value for *TmCsp* (Figure 3.34 and Table 3.10). A slight shift to higher temperatures from 11 °C to 15 °C was observed in the temperature of maximum thermostability. It was expected that the thermostability of *BsCsp* would be increased to a greater extent in the CTM. The reason that there was not a greater improvement in thermostability may be that residues around the salt-bridges were not modified to those in *TmCsp* and these may play an important role in the stabilizing effect of this cluster (187). The additional salt bridge from residues 63 to 49 not formed in CTM could be important to the electrostatic network of *TmCsp*. These results may highlight an obstacle of rational design strategies in that simply adding salt bridges into a protein may not produce the expected increase in thermostability. In some cases this may interfere with solvent interactions actually decreasing thermostability (187).

Table 3.10: Thermostability parameters for *TmCsp*, *BsCsp* and CTM in 63 mM sodium phosphate buffer. ΔC_p values were obtained from gradients on Figure S3.

	T_m (°C)	ΔH^{F-U} (kJ/mol)	ΔC_p (kJ/mol/K)	max ΔG^{F-U} (kJ/mol)	Temp of max ΔG^{F-U} (°C)
<i>BsCsp</i>	52.3 ± 0.3	-181 ± 6	4.08 ± 0.15	-11.8	10.9
CTM	62.2 ± 0.4	-189 ± 6	3.87 ± 0.22	-13.4	15.3
<i>TmCsp</i>	81.9 ± 0.8	-271 ± 4	3.86 ± 0.10	-25.1	18.4

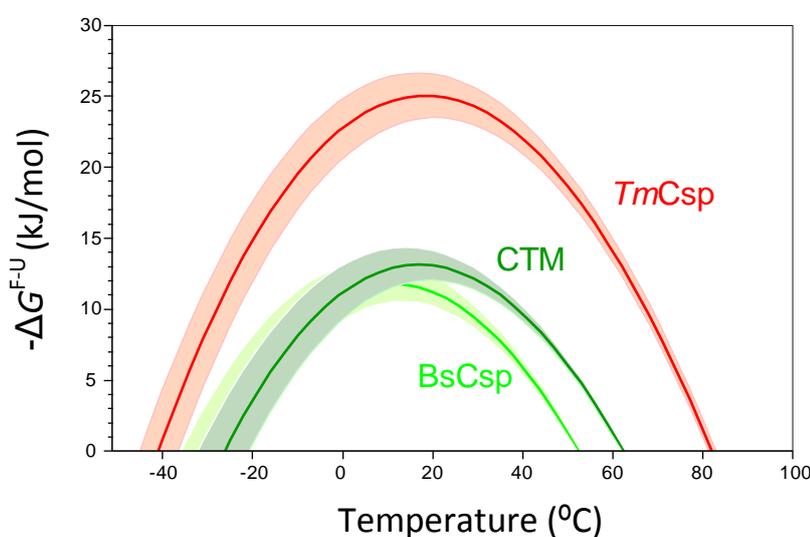


Figure 3.34: Thermostability plot for *BsCsp* CTM with *TmCsp* in red, *BsCsp* in light green and CTM in dark green.

3.5.2. *Thermotoga maritima* hydrophobic deletion mutants

One of the other major factors proposed to affect thermostability is increased levels of hydrophobic core packing in thermophiles. To observe the significance of this to stabilising *TmCsp* two further mutants were developed by Dr Toni Hoffmann. These were *TmCsp* V62A and L40A which both substitute larger hydrophobic residues for smaller hydrophobic residues to decrease levels of hydrophobic core interactions. V62A is located in β -strand 5 and L40A in the loop connecting β 3 and β 4 (Figure 3.37).

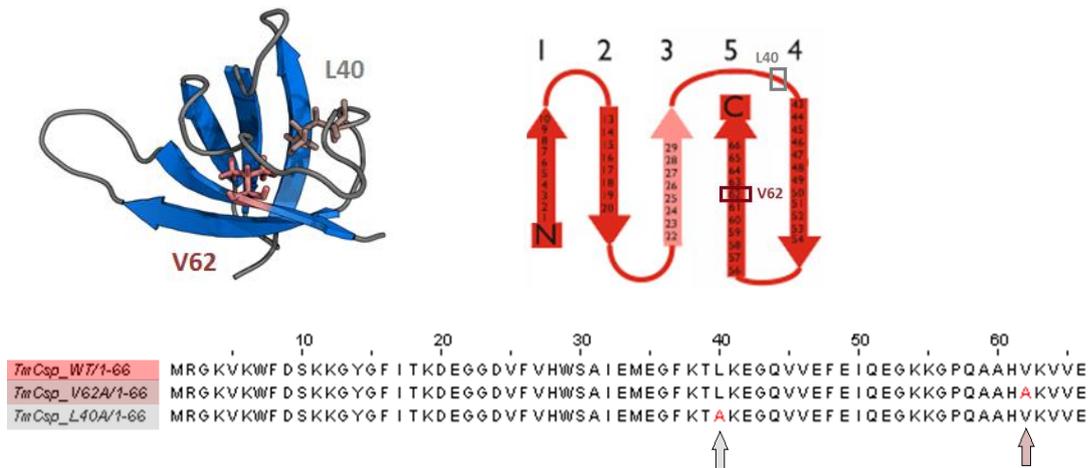


Figure 3.35: Location of the modifications in the hydrophobic deletion variants. Top left: Position of amino acid substitution in *TmCsp* tertiary structure with β -sheet shown in blue. Top right: Positions in mutation in secondary structure. Bottom: Alignment of wild type Csp with the two hydrophobic deletion mutants. The mutations are indicated by arrows below the alignment.

As expected the hydrophobic deletion mutants showed a large decrease in the resistance of the protein to thermal and chemical denaturation (Figure 3.36, Table 3.11 and Table 3.12)(191). The T_m values of the mutants at 61 °C and 54 °C were substantially lower than the 81 °C - 82 °C for the wild type *TmCsp*. Another notable factor is that while the T_m value of L40A remains higher than that of *BsCsp* the unfolding enthalpy of the thermal unfolding transition is lower (Table 3.10).

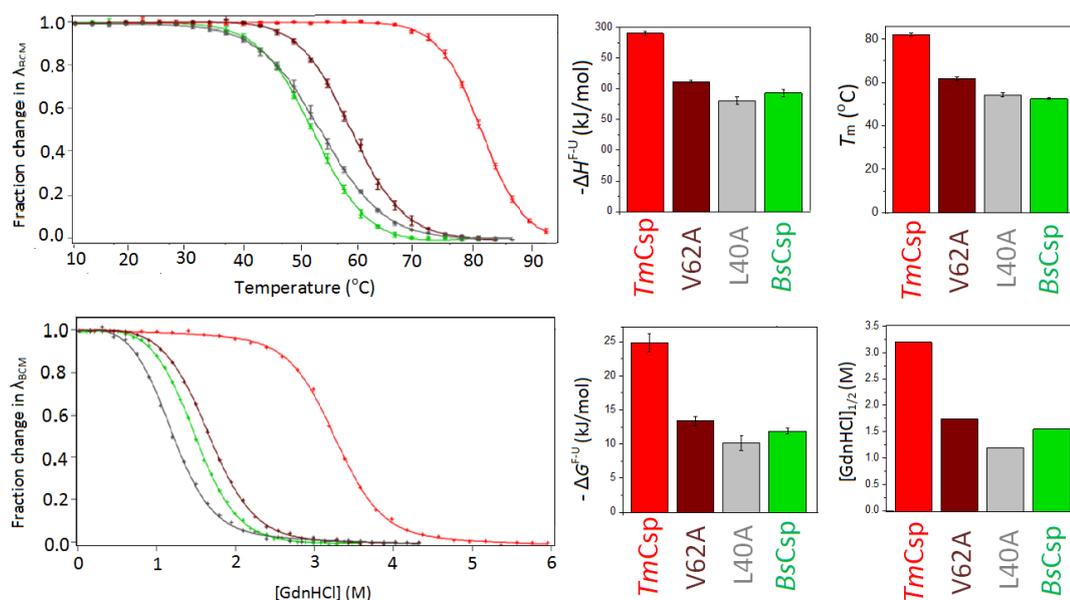


Figure 3.36 Comparison of the thermal and chemical stabilities of the hydrophobic deletion mutants $BsCsp$ and $TmCsp$ measured by tryptophan fluorescence. Top: Thermal unfolding of the $TmCsp$ mutants compared with $TmCsp$ and $BsCsp$ wild type. $TmCsp$ is shown in red, $BsCsp$ in green, $V62A$ in brown and $L40A$ in grey. Bottom: Chemical denaturation at $10^{\circ}C$.

Table 3.11: Thermal unfolding parameters including $V62A$ and $L40A$ in 63 mM sodium phosphate buffer

	CD		PTI (corrected λ_{BCM})	
	ΔH^{F-U} (kJ/mol)	T_m ($^{\circ}C$)	ΔH^{F-U} (kJ/mol)	T_m ($^{\circ}C$)
$BsCsp$	-175 ± 5	52.5 ± 0.6	-181 ± 6	52.3 ± 0.3
$TmCsp$	-223 ± 17	80.2 ± 0.8	-271 ± 4	81.9 ± 0.9
$TmCsp$ $V62A$	-188 ± 6	61.3 ± 0.4	-198 ± 2	61.1 ± 0.9
$TmCsp$ $L40A$	-160 ± 5	54.1 ± 0.8	-168 ± 6	53.9 ± 0.4

Table 3.12: Values obtained from fits to equilibrium denaturation curves with GdnHCl at 10 °C in 63 mM sodium phosphate buffer

Protein	ΔG^{F-U} (kJ/mol)	m (kJ/mol/M)	$[\text{GdnHCl}]_{1/2}$ (M)	$\Delta\Delta G^{F-U}$ from $TmCsp$ (kJ/mol)
<i>BsCsp</i>	-11.9 ± 0.4	7.6 ± 0.2	1.56	12.9
<i>TmCsp</i>	-24.8 ± 1.4	7.8 ± 0.4	3.21	-
<i>TmCsp</i> V62A	-13.4 ± 0.7	7.6 ± 0.3	1.76	11.4
<i>TmCsp</i> L40A	-10.2 ± 1.1	8.4 ± 0.6	1.21	14.6

Thermostability curves were constructed for V62A and L40 (see appendix S3 for intermediate plot). The maximal thermostability of the hydrophobic mutants is significantly reduced and this appears to be across all temperatures (Figure 3.37 and Table 3.13). A small decrease in the temperature at which the proteins are most stable at (T_s) is observed and the temperature predicted for cold denaturation is not as low as predicted for the wild type *TmCsp*. This result highlights a global collapse of the thermostability curves rather than a shifting and does appear to make the *TmCsp* thermostability profile resemble that of *BsCsp*.

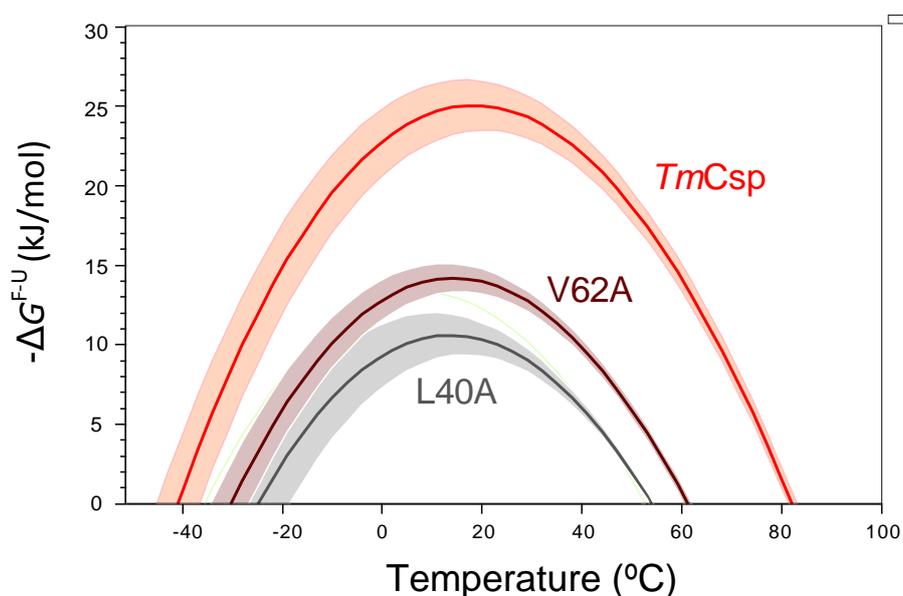


Figure 3.37: Thermostability projection of *TmCsp* V62A and L40A in 63 mM sodium phosphate buffer compared to *TmCsp* WT

Table 3.13: Thermostability parameters of V62A and L40A mutants in 63 mM sodium phosphate buffer

	T_m (°C)	ΔH^{F-U} (kJ/mol)	ΔC_p (kJ/mol/K)	max ΔG^{F-U} (kJ/mol)	Temp of max ΔG^{F-U} (°C)
<i>BsCsp</i>	52.2 ± 0.3	-181 ± 6	4.08 ± 0.15	-11.8	10.9
<i>TmCsp</i>	81.9 ± 0.8	-271 ± 4	3.86 ± 0.10	-25.1	18.4
V62A	61.0 ± 0.4	-197 ± 2	3.89 ± 0.17	-14.2	14.4
L40A	53.8 ± 0.4	-168 ± 6	3.91 ± 0.26	-10.6	11.6

The thermostability of the V62A mutant at 10 °C showed a decrease of 46% from the value of -24.8 kJ/mol for the wild type *TmCsp* to -13.4 kJ/mol in V62A which is a similar thermostability to the charged triple mutant. The decrease in ΔG^{F-U} was even more dramatic for the L40A mutant with a reduction of 56% from -25.1 kJ/mol *TmCsp* to -10.9 kJ/mol. This thermostability is actually below that shown by the mesophilic *BsCsp*. This highlights the significance of the stabilising role of hydrophobic packing in the core of hyperthermophiles. It illustrates quite how optimised their hydrophobic core is that a single change in the size of one side chain has such a profound effect on Csp thermostability. Building up a combination of mutations to generate thermostability is very complex and is extremely vulnerable to mutations. Mutations that affect hydrophobic packing such as those presented are able to completely abolish the effects of hyperthermophilic adaption.

3.6 Discussion

Using recombinant DNA technology and expression in *E. coli* followed by column purification it proved possible to obtain protein samples of Csps adapted to vastly different temperatures. Despite showing very high levels of sequence similarity, the Csps had vastly different thermal stabilities with a trend of increasing T_m and greater numbers of charged residues being apparent from cold-adapted to hot-adapted variants. The unfolding transition was sharper for the more hot-adapted Csps and shallower in cold-adapted variants. Cold-adapted Csps showed only a

moderate reduction in the mid-points of both chemical ($[D]_{1/2}$ values) and thermal transitions (T_m values). This highlights the need to preserve protein thermostability meaning that there are limits to the extent that these proteins can destabilise either through genetic drift due to a lack of selection pressure or through selective pressure to increase flexibility. It appears that a global expansion or collapse of the thermostability profiles caused by different ΔH values is the key mediator of the thermostability differences. Only slight changes in the temperature of maximal thermostability were observed despite vast changes in T_m . This shows the significance of the hydrophobic effect towards protein thermostability. It appears as if the vastly different stabilities of the Csps are based around optimising the proteins to hit a 'sweet spot' in terms of a similar thermostability at the temperature each Csp operates. This gives a very similar value of ΔG^{F-U} at around 8 °C below the respective optimum growth temperature of each bacterium. From thermostability data alone it is difficult to see if this correlates to equivalent levels of protein dynamics. This pattern does appear to apply to other protein families beyond Csps, though this may be more a trend rather than a specific defined rule.

Using a range of mutants generated by Dr Toni Hoffmann that were designed to serve as an intermediate between the mesophilic *BsCsp* and hyperthermophilic *TmCsp* the traits of Csp hot-adaption were examined. By altering the ionic bonding network of *BsCsp* it was possible to increase the T_m of a mesophilic Csp however ΔG_{max} was not substantially increased. The effects of decreasing the size of individual hydrophobic residues in *TmCsp* were much larger showing that effective hydrophobic packing is crucial to thermostable proteins. The large decreases in thermostability for one single unfavourable mutation show why rational design of thermostable proteins has proved very difficult. Extremophilic proteins have already evolved many of the traits desired in proteins such as high thermostability or high flexibility. Designing cold-adapted proteins may however prove complex as they are stable over narrower temperature ranges so care must be taken when choosing a strategy.

4 Characterising cold shock protein flexibility by NMR and kinetics

4.1 Introduction

In the previous sections it has been shown how thermophilic adaption produces increased protein thermostability and that psychrophilic proteins exhibit reduced thermostability with the hypothesis that this correlates to an increase in flexibility (192). In this section we begin to explore how true this is for the cold shock proteins and also consider the effect that DNA binding has on flexibility. A greater understanding of the dynamics of psychrophilic proteins could provide insights into the strategies which could be used to improve the flexibility of mesophilic and thermophilic enzymes and potentially increase their catalytic rates. The ability to predict the dynamics of proteins and find initial scaffolds on which to begin directed evolution studies is important and research into extremophiles may offer solutions.

Differences in the amino acid composition of extremophilic proteins can have a significant effect on flexibility. *Best et al.* showed how an increased level of the hydrophobic core packing that was observed between two proteins with similar folds correlated with a reduction in dynamics which infers that thermophilic proteins which have greater levels of hydrophobic packing will be more rigid (193). Thermophilic and hyperthermophilic proteins also contain greater numbers of charged residues capable of forming salt bridges which may further constrain the structure of these proteins. There is also a shift in the composition of positively charged residues in hyperthermophilic Csp's towards a greater proportion of lysine residues rather than arginine residues. The side chain of lysine can rotate more freely than that of arginine, potentially forming different patterns of interactions during rotations (194). This increase in lysine residues has been observed in malate dehydrogenase from the hyperthermophile *Methanococcus jannaschii* and is also observed in *TmCsp* (see section 3.2.2) (195). Psychrophilic Csp's exhibit reduced numbers of the structurally constraining proline and arginine residues and elevated numbers of the highly flexible glycine residues, often in clusters to provide regions of local flexibility (87)(196).

4.1.1 Flexibility comparisons between temperature adapted homologues

The concept of an inverse relationship between thermostability and flexibility was generated due to thermophilic proteins exhibiting slower rates of hydrogen-deuterium exchange, indicating less exposure of parts of the protein to solvent which implies less dynamics (196). An example is that deuterium-hydrogen exchange was found to be slower in *Thermotoga maritima* glyceraldehyde-3-phosphate dehydrogenase compared with mesophilic variants and a similar finding was made for isopropylmalate dehydrogenase variants (197). Lower proteolysis rates seen in thermophiles such as L-asparaginase and β -galactosidase compared to their mesophilic homologues have also been cited as implying lower flexibility in thermophiles (198). The reduced rates of catalytic activity in some hyperthermophiles have also been used to justify the concept of reduced dynamics in hot-adapted proteins (199). *Lam et al.* showed that the additional electrostatic interactions in a thermophilic acylphosphatase variant from *Pyrococcus horikoshii* were responsible for it exhibiting lower reaction rates implying the lower activity was directly linked to the thermophilic adaptations (200).

Other studies showing similar results include fluorescence quenching experiments of a phosphoglycerate kinase from the thermophile *Thermus thermophilus* which showed a slower rate quenching relative to a yeast homologue (201). Molecular dynamics studies contrasting a hyperthermophilic rubredoxin from *Pyrococcus furiosus* and a mesophilic variant from *Desulfovibrio vulgaris* showed that the hyperthermophilic rubredoxin was slightly less flexible (202). Molecular dynamics folding and unfolding models which suggested a greater rigidity in thermophilic proteins over mesophilic proteins (203). A similar result was seen for comparisons of the crystal structures of 3-isopropylmalate dehydrogenases from *Thermus thermophilus* and *E. coli* (204). These suggest reduced internal flexibility in the thermophilic or hyperthermophilic variant. When the comparisons of phosphoglycerate kinases and rubredoxins were instead made at the optimal growth temperatures of the organisms the proteins derive from the levels of protein dynamics were similar.

The differences in flexibility between psychrophilic and mesophilic protein homologues have mainly been highlighted using fluorescence quenching experiments which in general have suggested that psychrophilic proteins exhibit increased levels of micro-unfolding events. α -amylase from a psychrophilic variant exhibits faster rates of acrylamide quenching than a mesophilic alpha amylase (43). The proportion of time spent in each conformation and rate of interchange between conformations could significantly affect the rates observed in exchange and quenching experiments. A comparison of crystal structure B-factors between 20 psychrophilic and mesophilic protein homologue pairs suggested that psychrophilic versions were more flexible as they contain larger cavities to interact with water (205). The presence and size of cavities within proteins can significantly affect the way a protein interacts with the solvent around it. These cavities are often large and hydrophilic in psychrophilic proteins while they are often blocked in thermophilic proteins (206). *Ohmura et al.* showed that large cavities introduced into lysozyme can be destabilizing (158).

4.1.2 Controversy over the theory that thermophiles and hyperthermophiles are less dynamic

While a large range of studies using an array of different techniques have shown psychrophiles to exhibit increased levels of dynamics and thermophiles to exhibit reduced levels of dynamics a significant number of studies disagree with this. Molecular dynamics simulations on flavoenzymes showed no significant differences in flexibility between thermophilic and mesophilic variants, while in a separate study increased flexibility of a thermophilic subtilisin-E, over a mesophilic homologue was suggested (207-208). MD simulations of α -amylases also suggest more internal fluctuations in a thermophilic α amylase than in a mesophilic variant (209). A study of RNaseH enzymes showed that *Thermus thermophilus* RNaseH exhibits both regions of increased and decreased rigidity compared to a mesophilic variant (210).

One of the most thermally stable proteins determined to date is rubredoxin from the hyperthermophile *Pyrococcus furiosus* which is stable to over 100 °C (211). This protein would be expected to be extremely rigid at lower temperatures and so exhibit low rates of hydrogen-deuterium exchange. In contrast to this expectation, however, high rates of exchange were seen even at 28 °C for all backbone amides. One of the organisms identified as growing at the highest temperatures is the obligate hyperthermophile *Geogemma barossii* which does not grow below 80 °C (212). Neutron scattering studies of dihydrofolate reductase homologues from this organism showed a slightly greater flexibility than the mesophilic *E. coli* variant (213).

These conflicting reports present a confused picture but on balance the reported studies suggest a trend of increased rigidity in thermophiles and increased dynamics in psychrophiles. Individual comparison however seems to be highly specific to the protein homologues with various different adaption strategies being applied in different protein families. From this variation it becomes clear that flexibility is not simply the inverse of thermostability and is far more complex. For example activity and proteolysis rates depend on other factors of protein topology so may not be an accurate measure of dynamics. It is important to consider that differences between homologues are generally not uniformly distributed across a protein and that the various different techniques to assess flexibility vary in the extent to which they reflect localised thermostability or global thermostability. Additional factors such as ligand binding may also affect dynamics (214).

4.1.3 Global and local protein flexibility adaptations

Comparisons of psychrophilic, mesophilic and thermophilic α -amylase homologues suggest global thermostability differences between the homologues (41). This is not true of all temperature-adapted homologues as many may display much more

localised changes in flexibility (201). Molecular dynamics studies on mesophilic and psychrophilic trypsin-like serine protease homologues suggested no differences in global dynamics but significant differences in the flexibility of small localised regions within the protein, mostly close to the active site (215). As is seen across most homologue comparisons the active site catalytic residues were however highly conserved to maintain function of the protein. Similar findings were obtained for lysozyme where initial reports using hydrogen deuterium exchange suggested no flexibility differences between thermophilic and mesophilic homologues however later studies found significant differences in the relative active site flexibilities (196). As a consequence of increased flexibility the active site of many psychrophilic enzymes has been found to be more heat labile than the rest of the protein which could leave this region susceptible to denaturation (196). MD simulations on uracil DNA glycosylase that showed changes in the flexibility of a loop which recognises DNA, is crucial to psychrophilic adaption of this enzyme (216). Similarly hydrogen-deuterium exchange studies suggested that a small region of the active site of dihydrofolate reductase generated large differences in dynamics between a mesophilic and thermophilic homologue (217). Consequently, care must be taken with the interpretation of flexibility findings. Differences in activity between mesophilic adenylate kinase enzymes and a hyperthermophilic variant from *Aquifex aeolicus* were initially used to imply global flexibility differences. It was later found that these differences related to a small region of the protein responsible for the opening and closing of a lid domain while there was little change in the dynamics of the rest of the protein (218).

It may not just be residues near the active site which are significant, as various loop regions can show significant adaption in extremophiles such as the stabilization of loops in a thermophilic adenylosuccinate synthetase (219). A region in phosphoglycerate kinases shows significant variations in thermophiles and mesophiles as it regulates opening of the active site cleft while being distant from the active site itself (196).

4.1.4 Combining flexibility and thermostability properties of extremophilic enzymes

Efforts have been made to try to combine improved flexibility of the active site of psychrophilic enzymes with the thermostability of mesophilic and thermophilic proteins. A mutational study on a psychrophilic subtilisin showed that, by adding stabilizing mutations to the protein outside of the active site it is possible to improve thermostability while maintaining active site flexibility (220). Another study showed that by incorporating residues from a psychrophilic homologue it was possible to increase activity of a thermophilic homologue and in a number of cases not diminish thermostability (45). Psychrophilic enzymes have more open active sites with less substrate specificity so it is potentially easier to modify their specificity (196).

4.1.5 Dynamics of cold-shock proteins

Previous sections (Section 3.4) highlighted how the thermostabilities of different cold shock proteins (Csps) vary and are adapted to the different temperatures their host organisms inhabit. This section aims to uncover preliminary details of the extent to which these differences in thermostability reflect in changes to protein dynamics and the way these proteins fold and unfold. It may be postulated that the psychrotrophic PB6 Csp which shows the lowest thermostability would be more flexible and display a greater sensitivity to temperature changes (see Figure 3.25). The hyperthermophilic *TmCsp* would be expected to show the least flexibility as it exhibits the greatest thermostability and contains increased number of charged residues to form additional salt-bridge connections that may restrict protein movement (Section 3.3.2). In a previous study *Motono et al.* discovered differences in the flexibility of the C-terminal region of the Csps with more rigidity in thermophilic variants relative to mesophilic variants (185).

4.1.6 Studies of protein folding and unfolding kinetics

One of the major challenges for psychrophilic organisms is the folding of proteins to a native form as solutions become more viscous at low temperatures. This means that exchange between conformations becomes slower and also at low temperatures there are more issues with protein mis-folding (74). A weaker hydrophobic core may slow the initial collapse of an unfolded protein and the presence of increased numbers of highly flexible glycine residues could increase the degrees of freedom in parts of psychrophilic proteins (87). *In vivo*, psychrophiles have an increased number of protein folding chaperones to mitigate the impact of some of these effects. Psychrophilic proteins generally have fewer salt bridges and disulphide bonds so therefore may exhibit simplified folding landscapes.

There are also factors which complicate the folding of thermophilic proteins. Thermophilic proteins contain increased numbers of hydrophobic residues which could accelerate the rate of initial protein collapse in terms of protein folding but can also make thermophilic proteins significantly more prone to aggregation effects. (221). Hyperthermophiles have significant numbers of additional complex interactions such as salt bridge networks which must be formed during protein folding. Where folding intermediates are seen for folding of a thermophilic homologue but not a mesophilic form it can be difficult to determine if the intermediates are on-pathway, an indication of multiple folding pathways or protein mis-folding (222). The unfolding rate of proteins is correlated to protein thermostability with more stable proteins taking longer to unfold (223). The increased thermostability of thermophilic proteins produces generally slower unfolding rates. Sometimes the initial phase of unfolding may occur at a different location in the protein for thermophilic proteins as some regions have different relative stabilities compared to their mesophilic homologues.

4.1.7 Previous folding and unfolding studies on Csps

Previous studies of Csp folding have suggested a common folding mechanism between the mesophilic *BsCsp*, thermophilic *BcCsp* and hyperthermophilic *TmCsp* with similar rapid folding rates. They did not find evidence of any different folding intermediates between the Csps suggesting that the additional interactions of *TmCsp* over *BsCsp* do not occur via a clear folding intermediate evidenced by the lack of rollovers in the chevron plots for *BsCsp*, *BcCsp* and *TmCsp* (224-226). Förster resonance energy transfer studies by *Magg et al.* showed that folding of the thermophilic *BcCsp* proceeds by rapid collapse with no significant energy barrier between extended and collapsed forms (224). They found that a crucial hairpin of $\beta 1$ and $\beta 2$ was present in the transition state but did not form in the initial hydrophobic collapse of the protein during folding.

Protein unfolding initiates at the part of the protein structure which is thermodynamically weakest which for Csps is the C-terminal region (224). Molecular dynamics simulations by *Motono et al.* reveals that unfolding of the C-terminal region of *TmCsp* occurs at the same time as other regions of the protein rather than the first stage of unfolding (185). This difference is caused by additional salt bridges in *TmCsp* which increases the thermostability of the C-terminus. Stabilization of the weakest part of mesophilic Csps provides *TmCsp* with increased thermostability relative to its homologues and slower unfolding rates (171).

4.2 Results

4.2.1. Folding and unfolding of the Csp homologues

The folding status of a protein can be monitored by a change in the fluorescence signal at 350 nm as the fluorescence emission of tryptophan residues is dependent on the extent to which they are exposed to solvent or shielded within a folded form. In Csps a higher fluorescence signal is associated with the folded state. The change

in fluorescence can be followed by stopped flow kinetics which involve the rapid mixing of two solutions and recording of the change in fluorescence signal. To monitor Csp refolding, an unfolded protein in a high concentration of denaturant was combined 1:9 with buffer with a lower denaturant concentration (see Methods 2.8). The lower the denaturant concentration, the faster the initial folding rates (Figure 4.1 left). The curves can be fit to an exponential function which generates values for the rate constant of the reaction. Similarly protein unfolding can be followed by combining a solution of folded protein in buffer with buffer containing a greater concentration of denaturant. The higher the denaturant concentration the greater the initial unfolding rate and fitting of each curve to an exponential yields the observed rate constants. This approach was taken for 30 μ M samples of PB6 Csp, *PsiCsp*, *BsCsp* and *TmCsp*. The fitted data for PB6 Csp is shown in Figure 4.1.

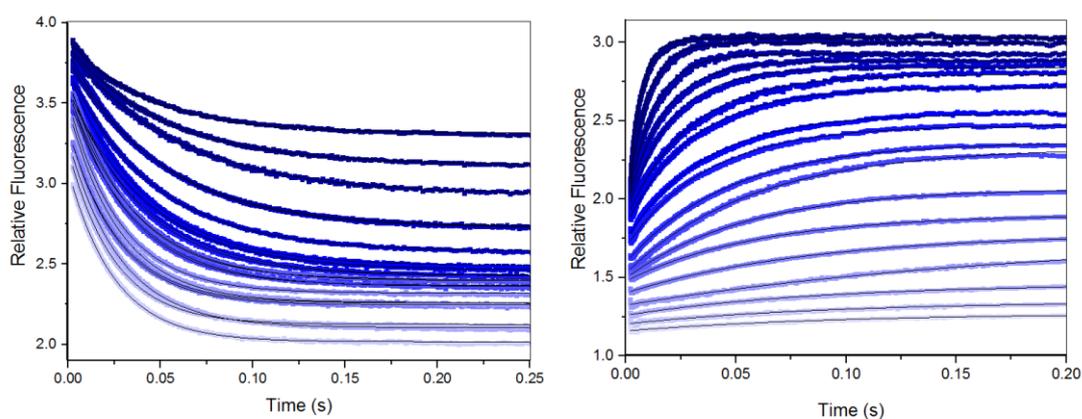


Figure 4.1 PB6 Csp folding and unfolding conducted using stopped flow mixing and followed by fluorescence changes. Representative folding data shown left and unfolding data, right for 30 μ M PB6 Csp at 10 $^{\circ}$ C in 63 mM sodium phosphate buffer. Left: Folding from 4.5 M [GdnHCl] into 0.125 M [GdnHCl] (darkest blue) to 2.25M (lightest blue). Right: unfolding from 0 M [GdnHCl] into greater [GdnHCl] of 1.5 M [GdnHCl] darkest blue to 4 M lightest blue. Each run was performed 8 times and the kinetic traces averaged.

The folding and unfolding rate constants were plotted against guanadine hydrochloride concentration to generate a chevron plot with the left arm in each trace representing protein refolding and the right arm representing unfolding (Figure 4.2) (236). The refolding of all of the Csp's was rapid and exhibited very

similar rate constants ($k_f = 194 \text{ s}^{-1} - 344 \text{ s}^{-1}$). The unfolding rate constants (k_u) were dramatically slower and significantly more varied. The unfolding rate constants appear to directly reflect the relative thermostabilities of the Csp. The refolding rate constants can be seen as a representation of the energy barrier between the unfolded state and transition state, while the unfolding rate constants represent the energy barrier between the folded state and the transition state. As the folding rate constants extrapolated to 0 M denaturant are so similar between all the Csp (Table 4.2) this together with the beta-Tanford (β_T) values (Table 4.1) suggests that the energy barrier between unfolded and transition state is similar between the Csp. The differences in protein thermostabilities derive from the differences in unfolding rate constants which implies that the differences in thermostabilisation between the Csp are thermodynamic rather than kinetic. This also suggests that the mechanism of protein folding is fully conserved between the different β -barrel Csp. This makes evolutionary sense as the Csp β -barrel topology is seen across so many kingdoms of life and in organisms growing at a wide range of temperatures. The presence of intermediates is often observed by curvature at very low or high denaturant concentrations though none were seen here which agrees with the theory of Csp folding being a two-state process.

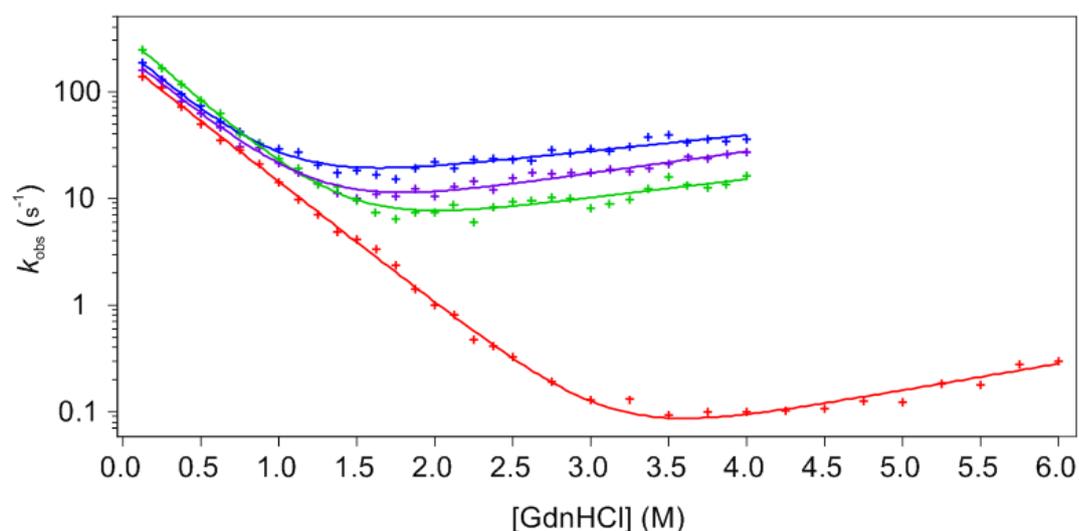


Figure 4.2: Chevron plot of Csp folding and unfolding at 10 °C. PB6 Csp in blue, *PsiCsp* in purple, *BsCsp* in green and *TmCsp* in red

A linear extrapolation method is used to calculate the rate constants of folding and unfolding under conditions of no denaturant which are located at the y-intercept. These intercept points can then be used to obtain ΔG^{F-U} for the protein (see Methods 2.8). m -values (m_f and m_u) are obtained from the gradients of the linear regions of the folding and unfolding arms of the multiplied by RT . These are added to give the M-value of unfolding. These can then be compared with values obtained from equilibrium curves performed at the same temperature (Table 4.1 and Table 4.2). The values were found to be in close agreement suggesting that Csp folding and unfolding occur in a two-state manner. The beta-Tanford (β_T) value is a measure of the degree of solvent exposure in the transition state relative to the folded state. It ranges from 0 to 1 and can be calculated from equation 2.20. The large values obtained for the Csps (0.84-0.89) suggests that the transition state is close in compactness to the native state (147).

Table 4.1: Stopped flow kinetics derived parameters at 10 °C in 63 mM sodium phosphate buffer. K_u and k_f were determined by extrapolation of the linear regions of the refolding arms and unfolding arms of the chevron plot back to the y-axis (0 M [GdnHCl]). M-values derive from the gradient of the refolding and unfolding arms of the chevron plot respectively.

	k_f (s ⁻¹)	k_u (s ⁻¹)	m_f (kJ/M/mol)	m_u (kJ/M/mol)	β_T
<i>PsiCsp</i>	225 ± 3	4.1 ± 0.6	6.5 ± 0.2	1.0 ± 0.2	0.87
PB6 Csp	249 ± 2	9.7 ± 0.5	6.8 ± 0.2	0.8 ± 0.1	0.89
<i>BsCsp</i>	344 ± 3	3.0 ± 0.5	6.9 ± 0.2	0.9 ± 0.2	0.88
<i>TmCsp</i>	194 ± 2	0.009 ± 0.004	6.2 ± 0.1	1.4 ± 0.5	0.84

Table 4.2: Stopped flow kinetics derived parameters derived at 10 °C in 63 mM sodium phosphate buffer.

	ΔG^{F-U} (kJ/mol)	m (kJ/mol/M)	$[\text{GdnHCl}]_{1/2}$ (M)
<i>PsiCsp</i>	-9.4 ± 0.5	7.5 ± 0.3	1.25
PB6 Csp	-7.6 ± 0.3	7.7 ± 0.4	0.99
<i>BsCsp</i>	-11.1 ± 0.6	7.8 ± 0.2	1.42
<i>TmCsp</i>	-23.8 ± 2.7	7.6 ± 0.7	3.13

Table 4.3: Comparison of ΔG^{F-U} values obtained by stopped flow kinetics at 10 °C with equilibrium curve values obtained at 10 °C. Both sets were performed in 63 mM sodium phosphate buffer. No value is included for *TmCsp* in urea as the mid-point $[\text{urea}]^{1/2}$ exceeds the solubility limit of urea.

ΔG^{F-U} (kJ/mol)	Stopped flow	Equilibrium curve with Urea	Equilibrium curve with GdnHCl
<i>PsiCsp</i>	-9.4 ± 0.5	-8.6 ± 0.5	-9.0 ± 0.8
PB6 Csp	-7.6 ± 0.3	-7.7 ± 1.0	-7.9 ± 0.6
<i>BsCsp</i>	-11.1 ± 0.6	-11.3 ± 0.8	-11.9 ± 0.4
<i>TmCsp</i>	-23.8 ± 2.7	-	-24.8 ± 1.4

Previous folding studies have suggested that the folding of proteins may be more dependent on the topology of the protein secondary structure and tertiary structure rather than the specific amino acid sequence (231)(235). The unfolding mechanism of Csp is similar to that of SH3 domain proteins, dihydrofolate reductase and rubredoxins which are also composed of β -sheets (224)(228-230). Studies of β -sheet proteins containing an SH3 domain showed very high levels of conservation of regions of the proteins pertaining to folding (226)(232). This suggests that an efficient mechanism of protein folding for β -sheet proteins evolved very early and then has been maintained with only very limited refinements of this method since. This may explain why Csp folding rates and topology is conserved across all different kingdoms of life.

4.3 Expression of ^{15}N labelled protein for NMR studies.

NMR HSQC (Heteronuclear single quantum coherence spectroscopy) spectra have previously been obtained for *TmCsp*, *BsCsp*, *TaCsp* *TcCsp* and *LmCsp* (229)(233-234). As cold-adapted proteins are hypothesised to display increased conformational heterogeneity this would be expected to lead to reduced or absent cross-peaks or temperature dependent line broadening. The aim is to discover if a clear HSQC can be obtained for the psychrotroph PB6 Csp and to compare that with the signals of the other Csps.

One of the main complexities in performing NMR studies is, that in most cases, they require the protein under study to be enriched in an NMR-active isotope such as the ^{15}N labelled protein used in this study. To produce labelled protein, bacteria are grown in minimal medium which contains only a single source of nitrogen that has been labelled (Methods 2.5). This represents a challenge as growth and expression of Csps is challenging even in nutrient rich media but more difficult still in minimal media. This makes it very difficult to induce the protein when the bacteria are at a high enough growth rate. Through the use of M9 minimal media, it proved possible to obtain sufficient quantities of labelled *BsCsp* and *TmCsp* for study but insufficient levels of PB6 Csp even with optimisation of glucose levels. To attempt to resolve this problem, the PB6 Csp was fused-in-frame to a highly expressing protein called maltose binding protein (MBP) (Figure 4.3). The aim was that the MBP would dramatically improve expression of the Csp and that the MBP and PB6 Csp could then be separated during an additional purification step to yield free Csp.

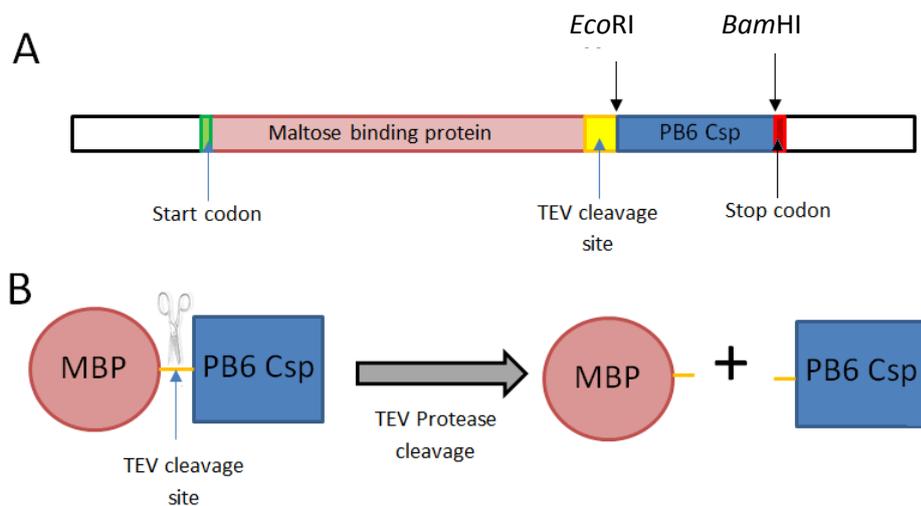


Figure 4.3: Schematic showing the use of the MBP fusion protein to express PB6 Csp in minimal medium. (A) Represents the DNA construct and (B) represents the process used to cleave expressed MBP-PB6 Csp fusion protein into MBP and PB6 Csp.

4.3.1 Over-expression and purification of ^{15}N MBP-PB6 Csp

A sample of pMal C5x plasmid was obtained from Dr Daniele Walsh containing sequences for a His tagged MBP-ferredoxin construct. The ferredoxin DNA sequence was flanked by *EcoRI* and *BamHI* restriction enzyme sites. The first step was to remove the ferredoxin sequence with restriction digestion. Concurrently PCR was performed to add the appropriate restriction enzyme sites to a PB6 sequence and to add a stop codon. The PB6 Csp sequence was ligated into the digested plasmid and transformed into XL1-Blue cells. To observe which plasmids had taken up the PB6 sequence and which still contained the ferredoxin sequence, samples of 5 colonies were grown overnight in LB and plasmid extracted as described 2.2.8. A sample of the plasmid was double digested with the same restriction enzymes. As ferredoxin is a longer sequence than PB6 Csp the plasmids which contained Csp (Figure 4.4 lanes 1-2) could be differentiated from those containing ferredoxin (Figure 4.4 lanes 3-4).

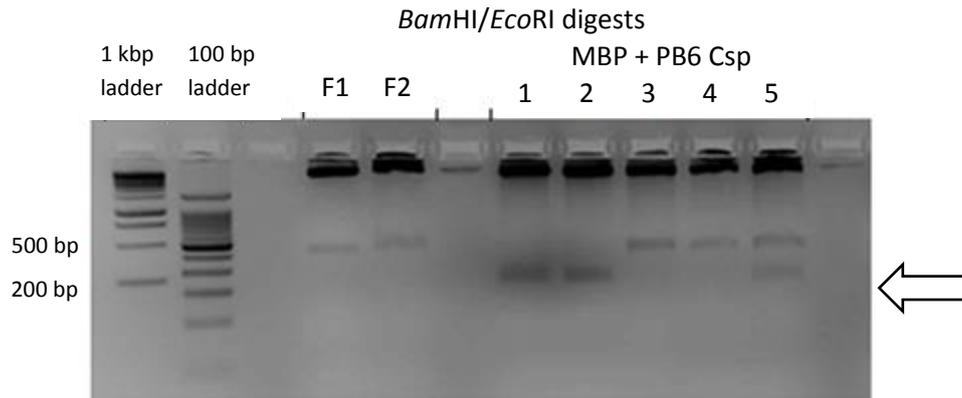


Figure 4.4: 1.5% Agarose gel showing double digest of pMal C5X plasmid containing MBP fusion proteins. F1 and F2 contain the MBP vector with ferredoxin insert and are a negative control. The remaining lanes (1-5) were used to investigate plasmids ligated with PB6 Csp. Lanes 1 and 2 contained the PB6 Csp sequence, lanes 3 and 4 the ferredoxin sequence and lane 5 a mixed colony.

Moderate expression of MBP-Csp was observed with minimal media expression using IPTG induction (Figure 4.5). It was possible to purify the MBP-PB6 Csp fusion protein using the same 3 step process as in LB expression (Figures 4.6-4.7).

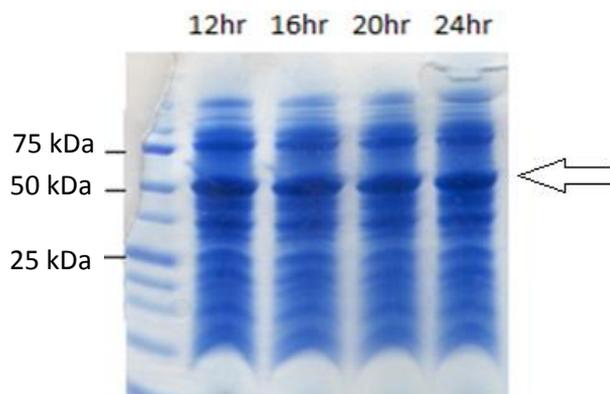


Figure 4.5: Expression of MBP-PB6 Csp in minimal media. The expected mass of the fusion protein of 53 kDa is indicated by an arrow.

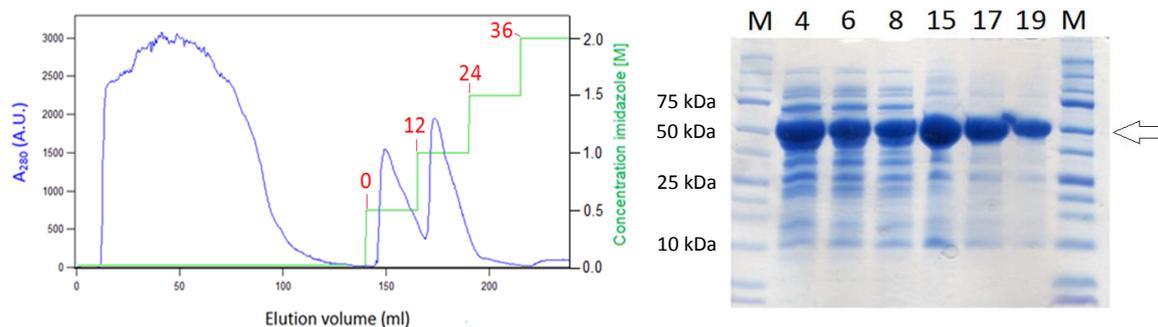


Figure 4.6: Elution profile displaying fraction eluted from Ni-NTA purification of MBP-PB6 fusion protein. Left: The blue line and left axis represents the absorbance at 280 nm. The green line represents the imidazole concentration applied to the column. Red numbers indicate fraction number. Right: SDS-PAGE gel of relevant fractions. M indicates marker lanes. The expected mass of the fusion protein is 53 kDa.

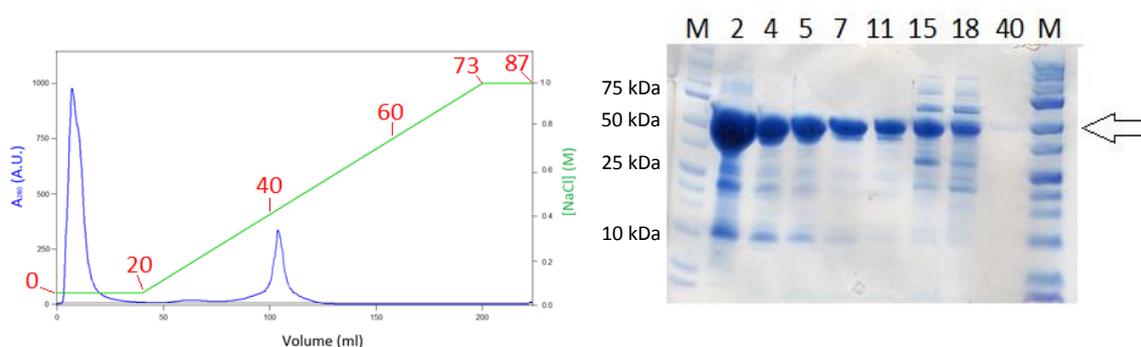


Figure 4.7: Ion exchange separation of MBP-PB6 Csp fusion protein and SDS-PAGE showing fractions eluted from a resource Q column. Left: Blue line and left axis represents the absorbance at 280 nm. Green line represents the NaCl concentration applied to the column. Red numbers indicate fraction number. Right: SDS-PAGE displaying selected relevant fractions. M represents lanes containing molecular size markers. The expected mass of the fusion protein is 53 kDa.

4.3.2 Cleavage of PB6 Csp from MBP-PB6 Csp and purification of PB6 Csp

MBP and PB6 Csp were separated through the action of TEV protease. While there was no band at the size of PB6 Csp before cleavage (8.4 kDa), a clear band can be seen after proteolysis by TEV protease (Figure 4.8). While it is hard to observe a shift in the MBP-band as the loss of PB6 Csp does not significantly alter movement of MBP through the gel, it does appear as if there is a small band corresponding to

uncut MBP and a lower stronger band relating to cut MBP in the after cleavage gel. This also serves as a positive sign that the cleavage step had proceeded near to completion.

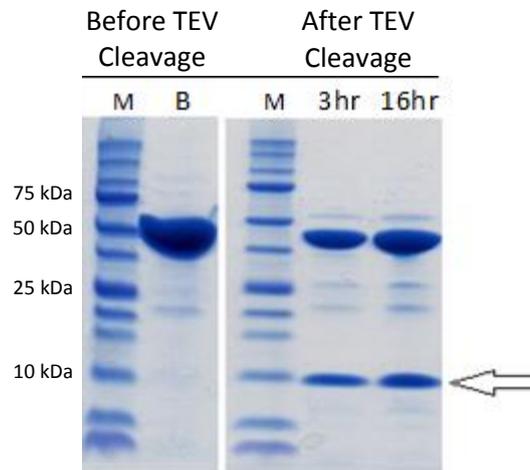


Figure 4.8: TEV cleavage of MBP-PB6 Csp to separate PB6 Csp from MBP-PB6 Csp. M represents molecular weight markers. B shows the purified fusion protein prior to cleavage. The 3 hour lane and 16 hour lane show incubation time with the first 20 minutes at room temperature and remaining time at 4 °C. The expected mass of PB6 Csp of 8.4 kDa is indicated with an arrow.

The products of the TEV cleavage reaction can then be separated by a final size exclusion step. As TEV and MBP are significantly larger than PB6 it was possible to easily separate the fractions containing PB6 Csp and MBP (Figure 4.9).

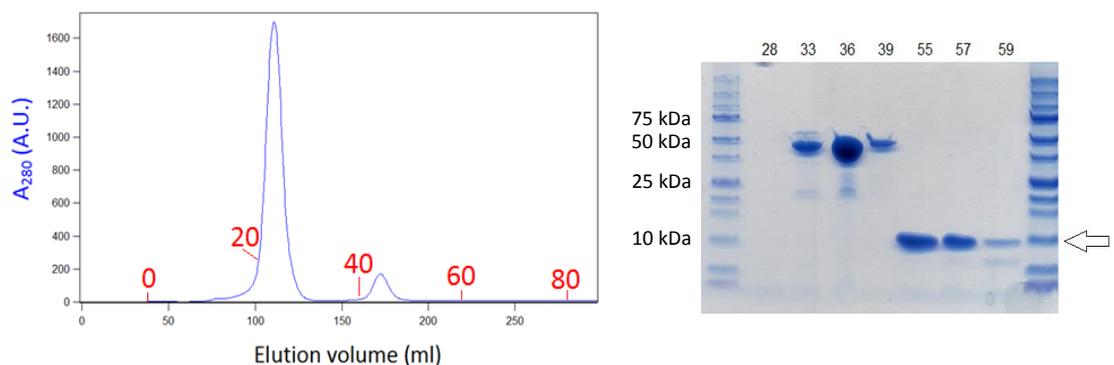


Figure 4.9: Purification of PB6 Csp from a mix of PB6 Csp, MBP, MBP-PB6 Csp and TEV by size-exclusion chromatography. Left: Blue line and left axis represents the absorbance at 280 nm. Fraction numbers are indicated in red. Right: SDS-PAGE gels of the size exclusion fractions. Edge lanes represent molecular mass markers. The expected mass of the PB6 Csp protein is 8.5 kDa

4.4 2D NMR spectra of the Csp monomers

In addition to structure determination, NMR is a powerful tool with which to elucidate dynamic processes in biomolecules (233). The section describes a preliminary investigation into the differences in dynamics between PB6 Csp, *BsCsp* and *TmCsp* and how these are affected by ligand binding and temperature.

The spins of nuclei align in a paramagnetic field and can be excited with electromagnetic radiation. This radiation can then be emitted when they relax to the ground state. The resonance of nuclei is hugely dependent on the surrounding chemical environment. In a 2D ^{15}N - ^1H HSQC (Heteronuclear Single Quantum Coherence) spectra, the resonant signal of each ^{15}N -H coupling generates a peak. These are plotted by value of chemical shift for ^{15}N and protons. In proteins cultured with ^{15}N as the only nitrogen source, each amino acid contains a ^{15}N -H bond except for proline where the side chain loops back on to the main chain nitrogen atom. Each ^{15}N -H has a specific value of chemical shift for ^{15}N and ^1H so these represent a defined co-ordinate for each residue. NMR spectra give a time averaged signal and regions which are highly dynamic will show increased spread of the signal, sometimes to such an extent that the signal falls below the noise threshold. Previous NMR assignments have been obtained for the hyperthermophilic *Thermotoga maritima* (*TmCsp*), thermophilic *Thermus aquaticus* (*TaCsp*), mesophilic *Bacillus subtilis* *BsCsp* and *E. coli* (*EcCsp*) as well as psychrotrophic *Listeria monocytogenes* (*LmCsp*) (237-240).

4.4.1 HSQC spectra of the Csp monomers

NMR spectra were recorded with the aid of Dr Lars Kuhn. It was possible to obtain HSQC spectra for *TmCsp*, *BsCsp* and PB6 Csp at 10 °C with clear well defined peaks. The number of peaks observed when excluding the side chain region (68 for PB6 Csp, 63 for *BsCsp* and 65 for *TmCsp*) was close to the number of amino acids present in the Csps (figure 3.7). The signal to noise ratio was slightly lower for

TmCsp as some of the protein precipitated and due to technical issues, spectra for *TmCsp* were performed on a lower field spectrometer (Figure 4.12). The spectra for *BsCsp* agreed with that of previously published data (237-238) while *TmCsp* showed significant deviation (239).

The spectrum for PB6 Csp as seen in Figure 4.10 shows a number of very small peaks with low dispersion around 8.5 ppm on the ^1H axis. As the protein is known to be stable at this temperature (see section 3.4 and figure 4.11), this may suggest the presence of micro-unfolding events. The *BsCsp* signal had a large number of less well defined peaks but the spectra still had a strong signal to noise ratio.

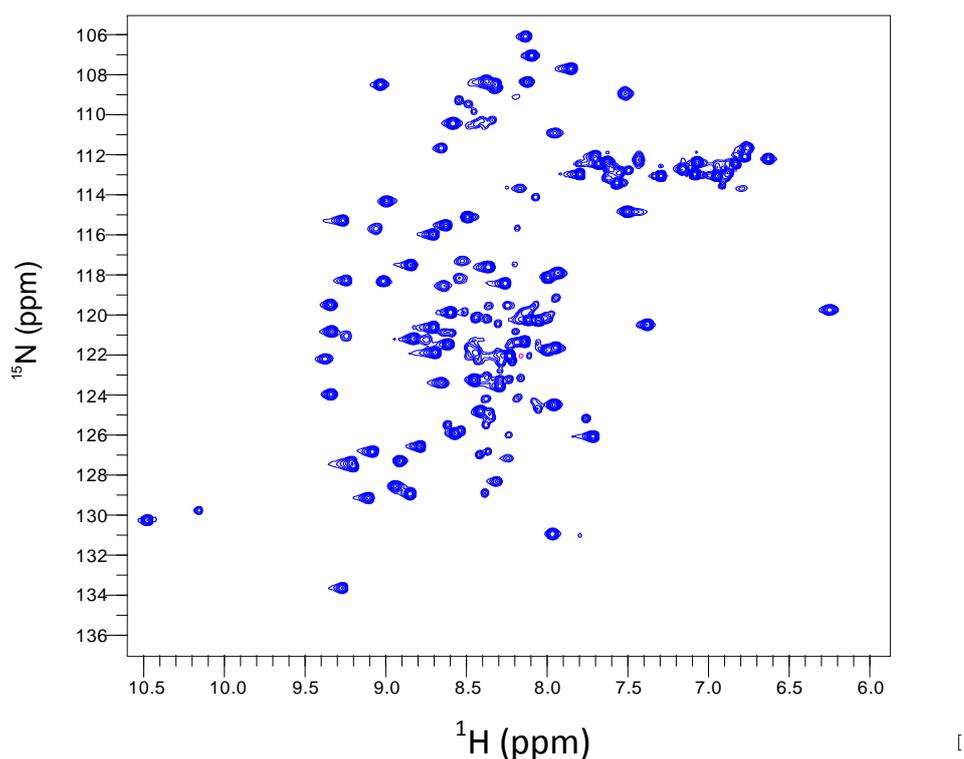


Figure 4.10: HSQC of PB6 Csp in 63 mM sodium phosphate buffer at 10 °C taken on a 750 MHz spectrometer.

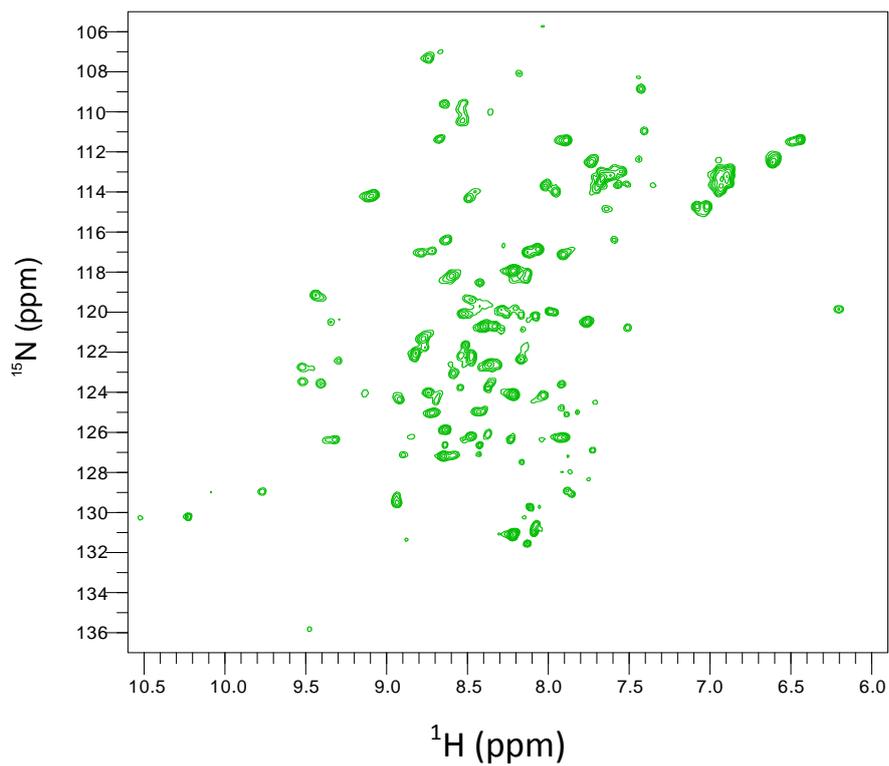


Figure 4.11: HSQC of *BsCsp* in 63 mM sodium phosphate buffer at 10 °C taken on a 750 MHz spectrometer.

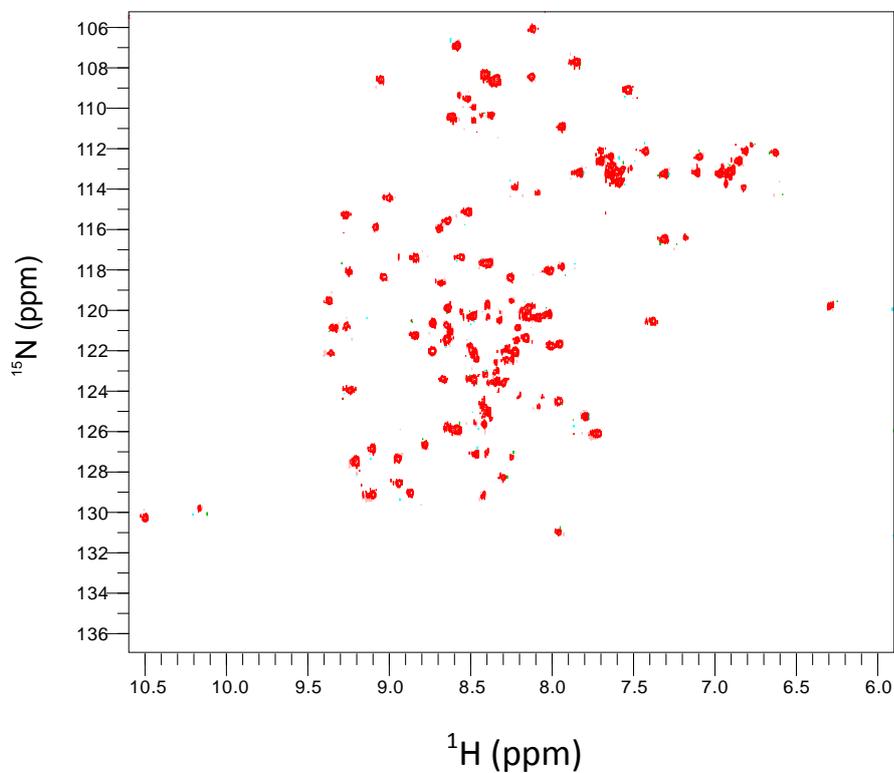


Figure 4.12: HSQC of *TmCsp* in 63 mM sodium phosphate buffer at 10 °C taken on a 600 MHz spectrometer

4.4.2 Exploring variations in dynamics with temperature for the Csp's using HSQC spectra

To generate a real understanding of the Csp's, as with previous sections it is important to study the effect of temperature on the NMR spectra. A clear change was seen in the value of chemical shift as the Csp's were heated. For PB6 Csp and *TmCsp* most residues showed a regular chemical shift change with temperature change from 10 °C to 23 °C and from 23 °C to 37 °C (Figure 4.13, 4.14 and 4.15). Some residues however exhibited no change with temperature and the overall number of peaks slightly decreased with increasing temperature. A general sharpening of peaks was observed but this is expected due to an increased tumbling rate of molecules at higher temperatures. For *TmCsp* and PB6 Csp the changes in dynamics were minor suggesting the same conformation is maintained at the different temperatures (Figure 4.13 and figure 4.15).

It may have been expected that PB6 Csp as the cold-adapted and least stable Csp would show more temperature sensitive dynamics at its ΔG^{F-U} value changes between 99% and 27% of its thermostability ΔG_{max} value while the other Csp's retain a greater percentage of ΔG_{max} . While HSQC spectra are only a snapshot of dynamics it still seems surprising that the large proportion of ΔG_{max} which is lost at 37 °C does not appear to give increased dynamics as far as peak widths. The one change that was observed in the PB6 Csp spectra was that the smaller peaks in the 1H 8 -9 ppm region were seen to disappear with increasing temperature suggesting that the state which they derive from is unfavoured at higher temperatures. This may suggest a shift occurs in a chemical equilibrium between two states with a change in temperature.

A far more substantial change was observed in *BsCsp* (Figure 4.14) with some peaks disappearing around a 1H value of 7.9 ppm as temperature was increased. This suggests this region of the protein becomes highly dynamic at higher temperatures so the signal broadens and the intensity falls below the observable threshold on the

HSQC spectra. The intensity of residues is controlled by two opposing factors. It increases as tumbling rate increases at higher temperatures but decreases if a protein becomes more dynamic. The greatest intensity was seen at 37 °C for PB6 Csp but at 23 °C for *BsCsp* and *TmCsp*

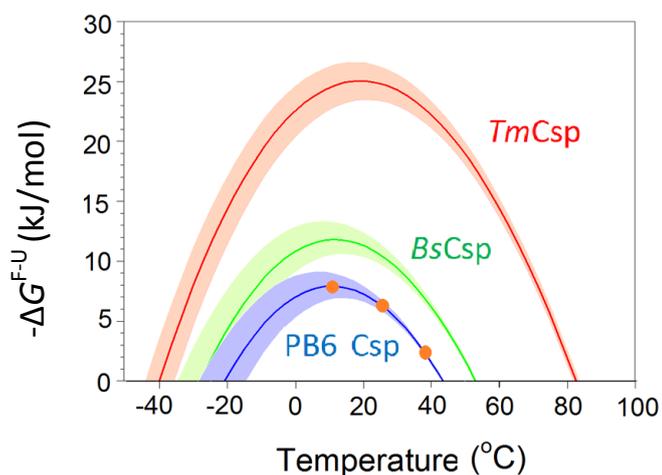
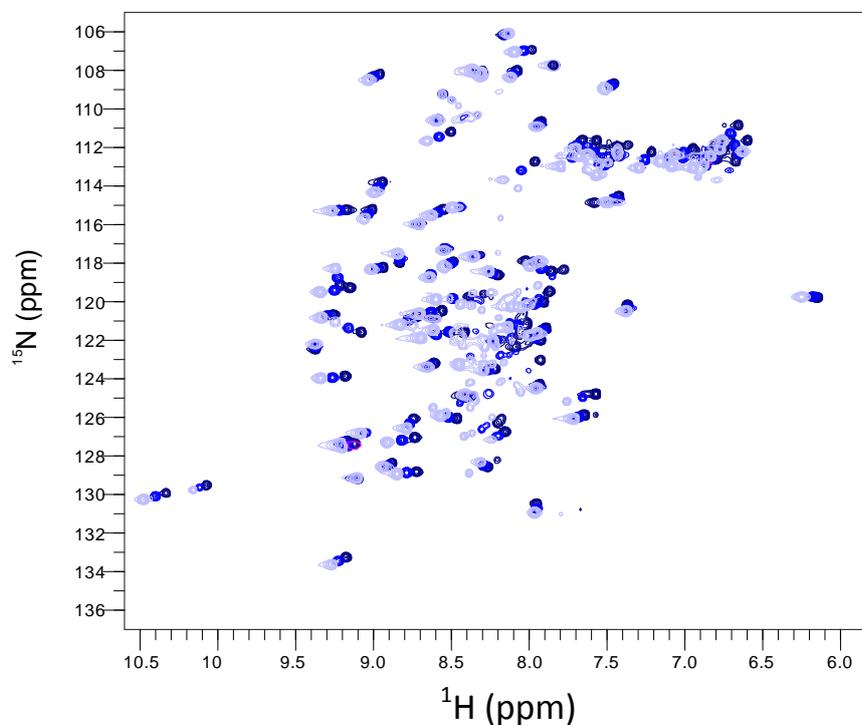


Figure 4.13: Temperature dependence of the HSQC spectra of PB6 Csp. Experiments performed in 63 mM sodium phosphate buffer on a 750 MHz spectrometer. Lightest shade 10 °C, middle shade 23 °C and darkest shade 37 °C. All spectra scaled relative to position of D₂O signal. A reproduction of figure 3.25 is shown to illustrate temperatures at which readings were taken (orange circles) relative to the thermostability profile of PB6 Csp. This is 99% of ΔG_{max} at 10 °C , 84% at 23 °C and 27% at 37 °C.

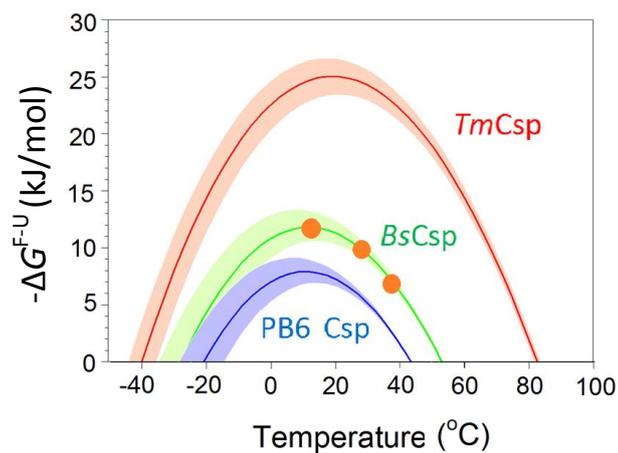
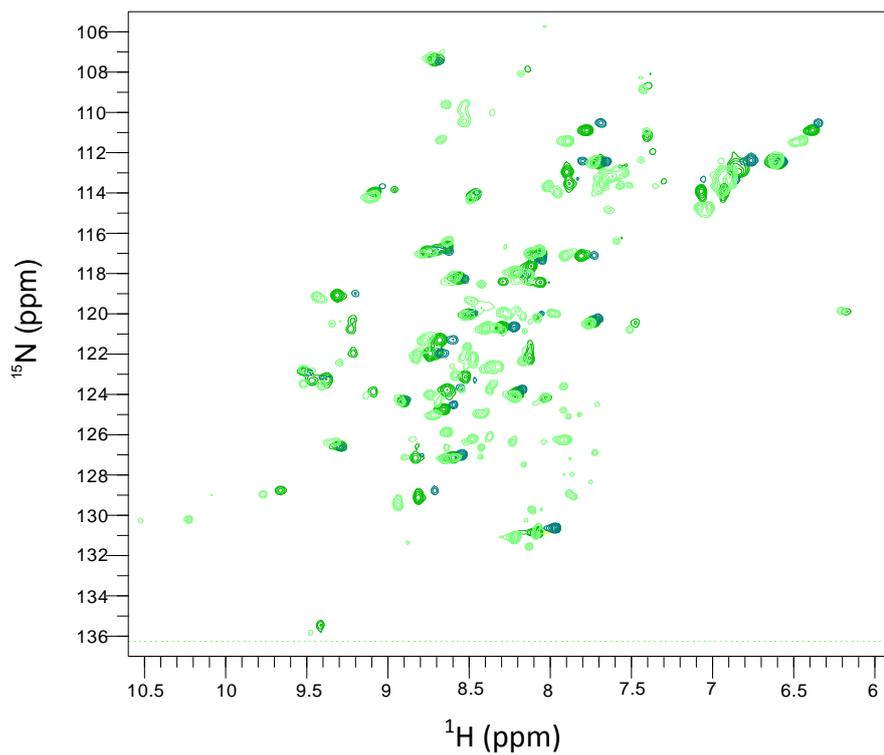


Figure 4.14: Temperature dependence of the HSQC spectra of *BsCsp*. Experiments performed in 63 mM sodium phosphate buffer on a 750 MHz spectrometer. Lightest shade 10 °C, middle shade 23 °C and darkest shade 37 °C. All spectra scaled relative to position of D₂O signal. A reproduction of figure 3.25 is shown to illustrate temperatures at which readings were taken (orange circles) relative to the thermostability profile of *BsCsp*. This is 99% of ΔG_{\max} at 10 °C, 93% at 23 °C and 59% at 37 °C.

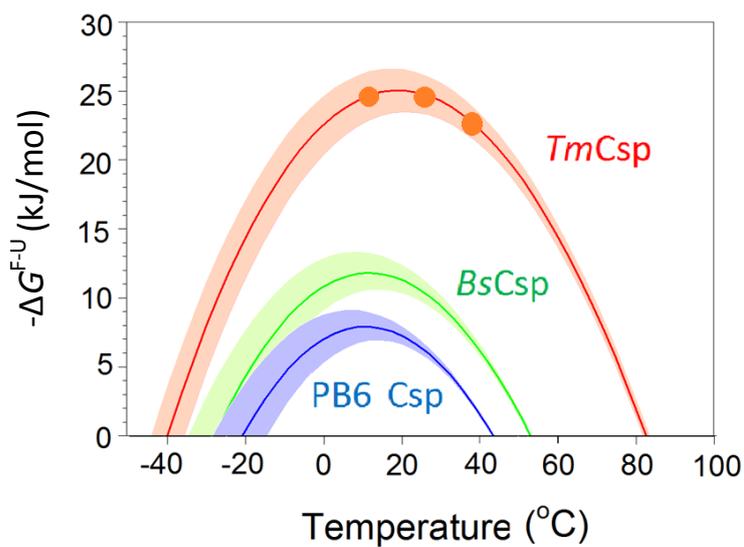
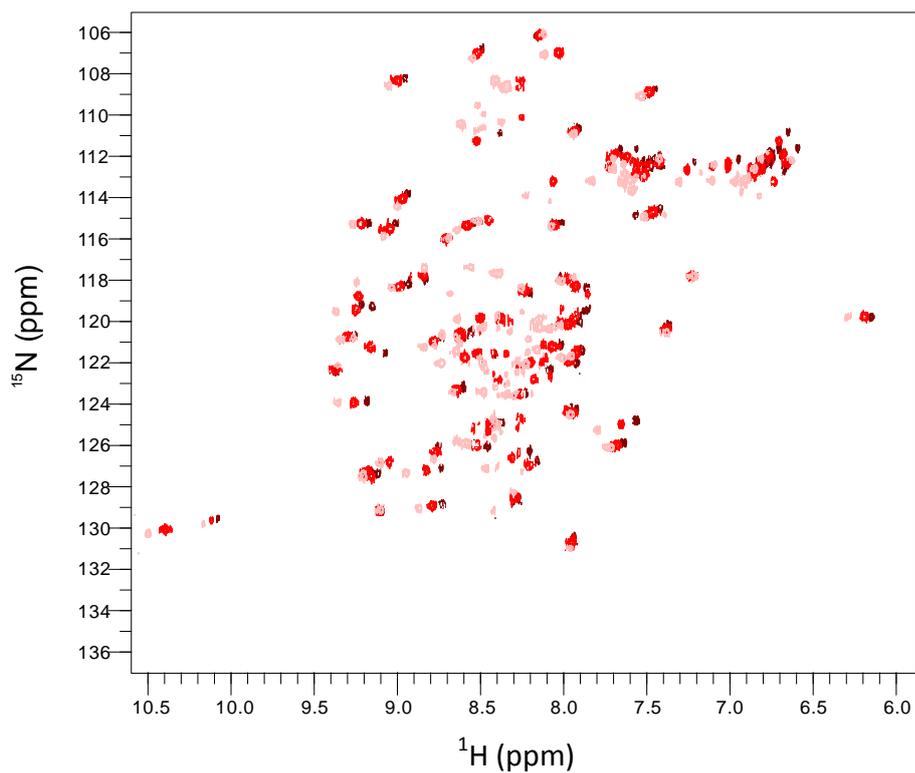
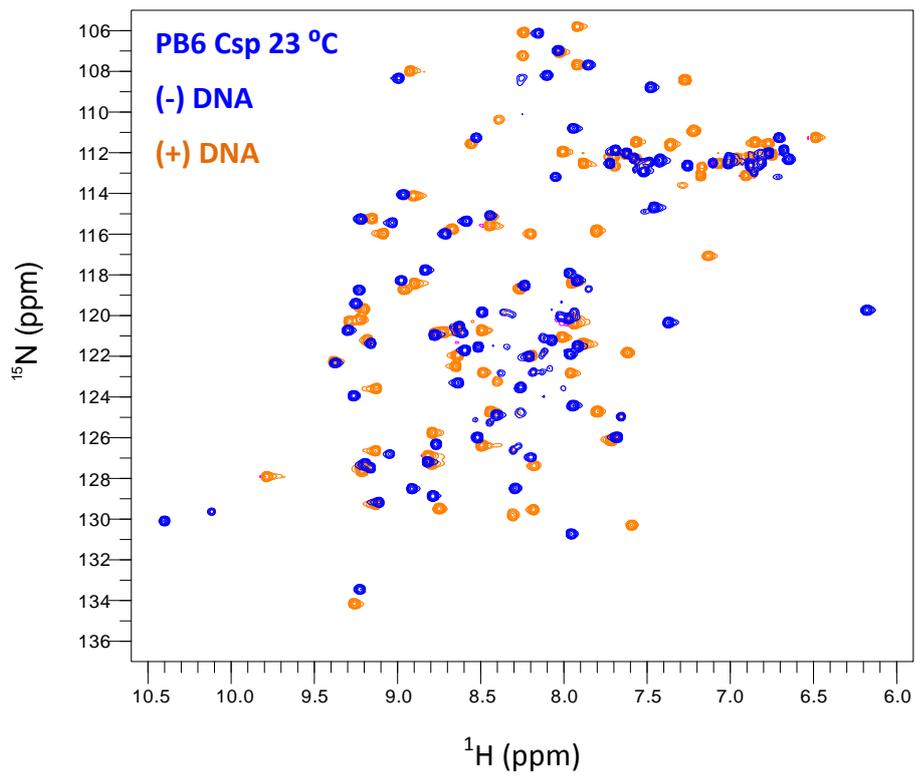
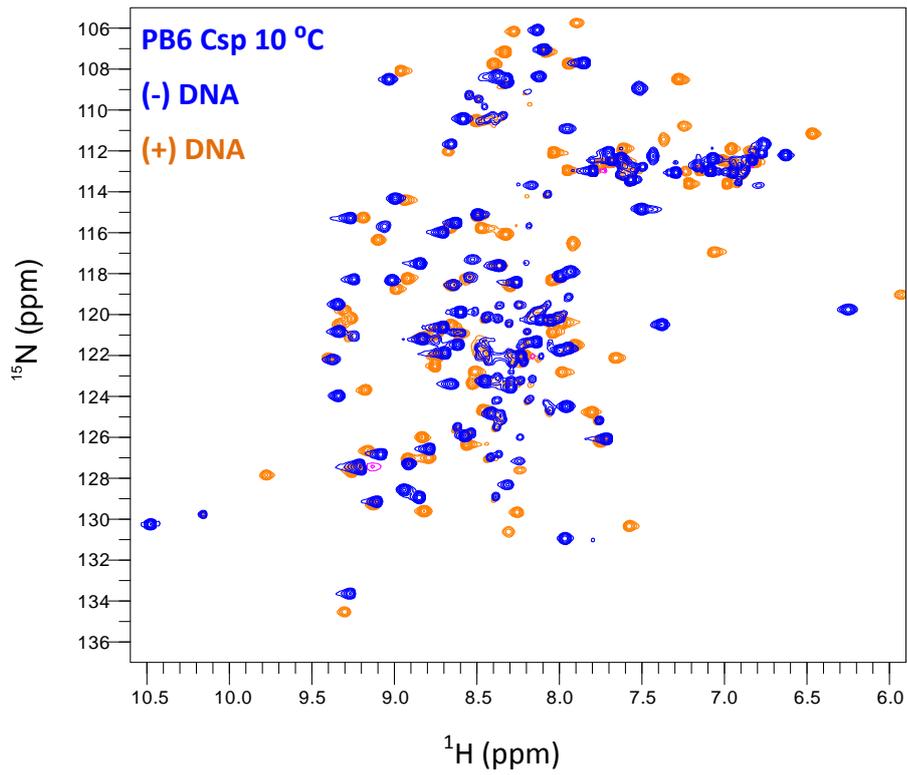


Figure 4.15: Temperature dependence of the HSQC spectra of *TmCsp*. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer. Lightest shade 10 °C, middle shade 23 °C and darkest shade 37 °C. All spectra scaled relative to position of D₂O signal. A reproduction of figure 3.25 is shown to illustrate temperatures at which readings were taken (orange circles) relative to the thermostability profile of *TmCsp*. This is 99% of ΔG_{\max} at 10 °C , 99% at 23 °C and 91% at 37 °C.

4.4.3 Changes in the HSQC spectra of the Csps upon binding of DNA

Part of the function of Csps is their ability to bind to single stranded DNA and RNA. It has been proposed that ligand binding stabilizes protein structures for example in the ligand binding to a protein called small ubiquitin-like modifier (214). To assess the effect of ligand binding in Csps an oligonucleotide of dT7 was used as it is the sequence which binds to *BsCsp* and *TmCsp* with greatest affinity (237). Observing changes to the HSQC spectra can provide information as to how much the binding of DNA causes adjustments in the structure of the protein and the width of the peaks can give insights into any stabilization effects. It was therefore decided to investigate how the NMR spectra of the Csps varied in the presence of a 1:1 molar ratio of DNA.

The spectra in Figure 4.16, 4.17 and 4.18 shows the difference in HSQC pattern between DNA-bound and DNA-free Csps at 10 °C. For PB6 Csp and *TmCsp* (Figure 4.16 and 4.18) very large shifts in a significant number of residues especially in the hydrophobic core region (7.5 to 9.5 ppm on the ^1H axis). This implies a conformation change due to the binding of DNA. It could be that the DNA directly perturbs the structure through a steric clash or makes interactions which cause a change in the relative thermostabilities of different conformations. The signals for the DNA bound forms appear to be slightly more punctate with smaller line width than the DNA-free peaks indicating a more rigid structure. The binding also cause the small weak intensity peaks at 10 °C to disappear (around 8.5 ppm on ^1H axis). A clear change is seen in all spectra at 10.3 - 10.4 ppm of the ^1H axis where the tryptophan peak is present in the unbound form but has moved in the DNA bound form. This fits with the involvement of aromatic residues in DNA binding.



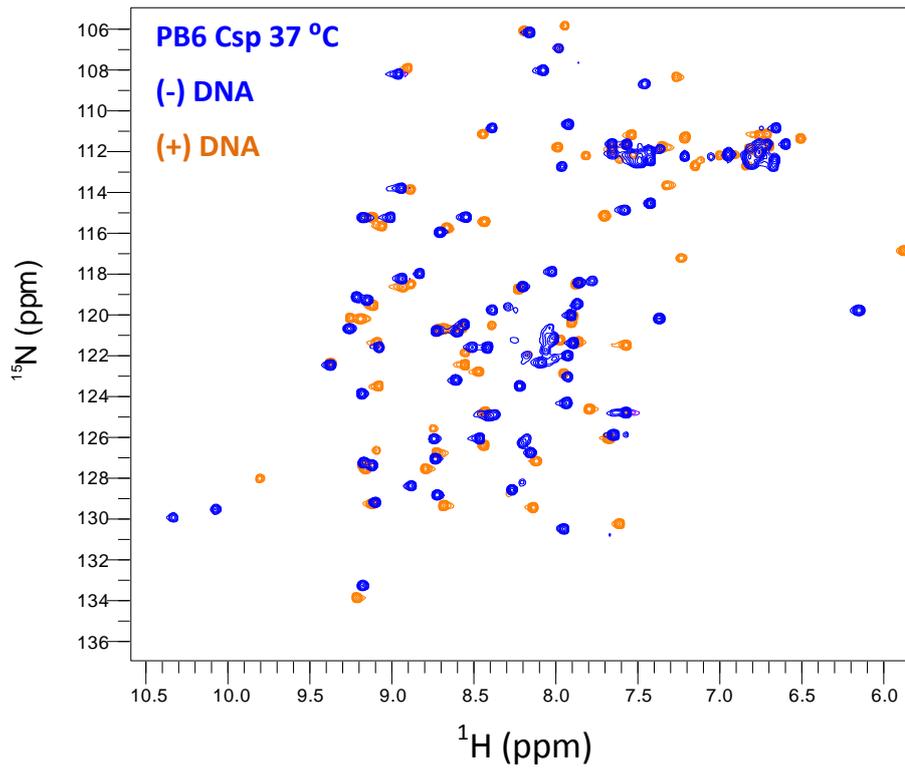
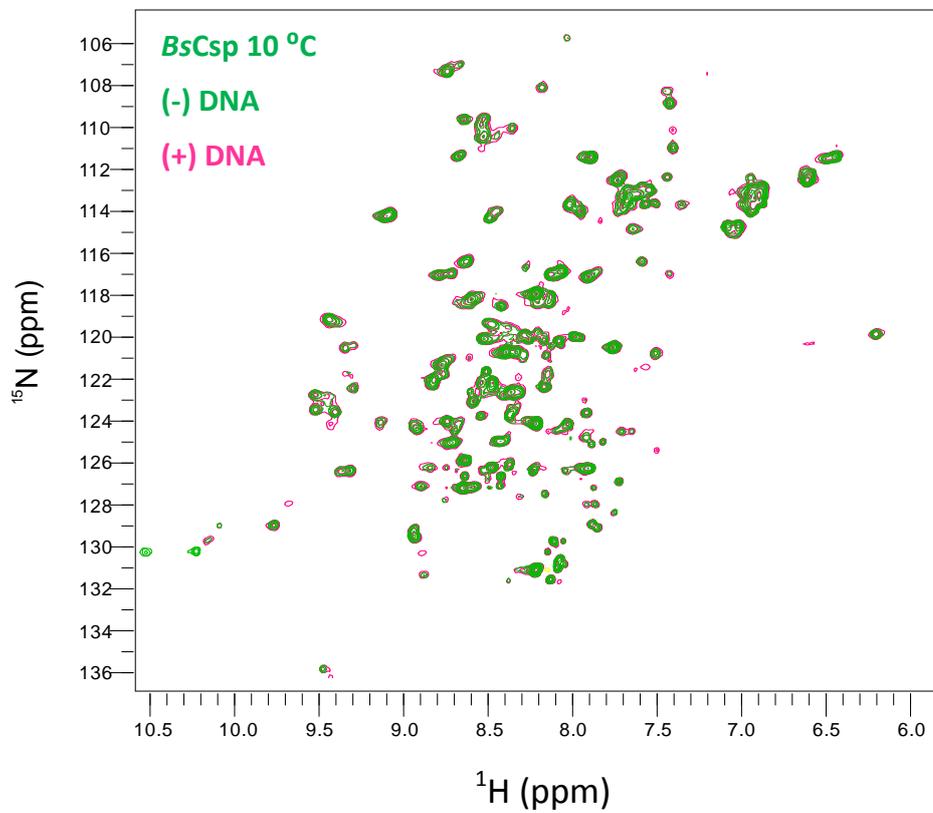


Figure 4.16. HSQC spectra showing the changes to PB6 Csp upon binding of DNA. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer. Blue represents without DNA and orange with DNA



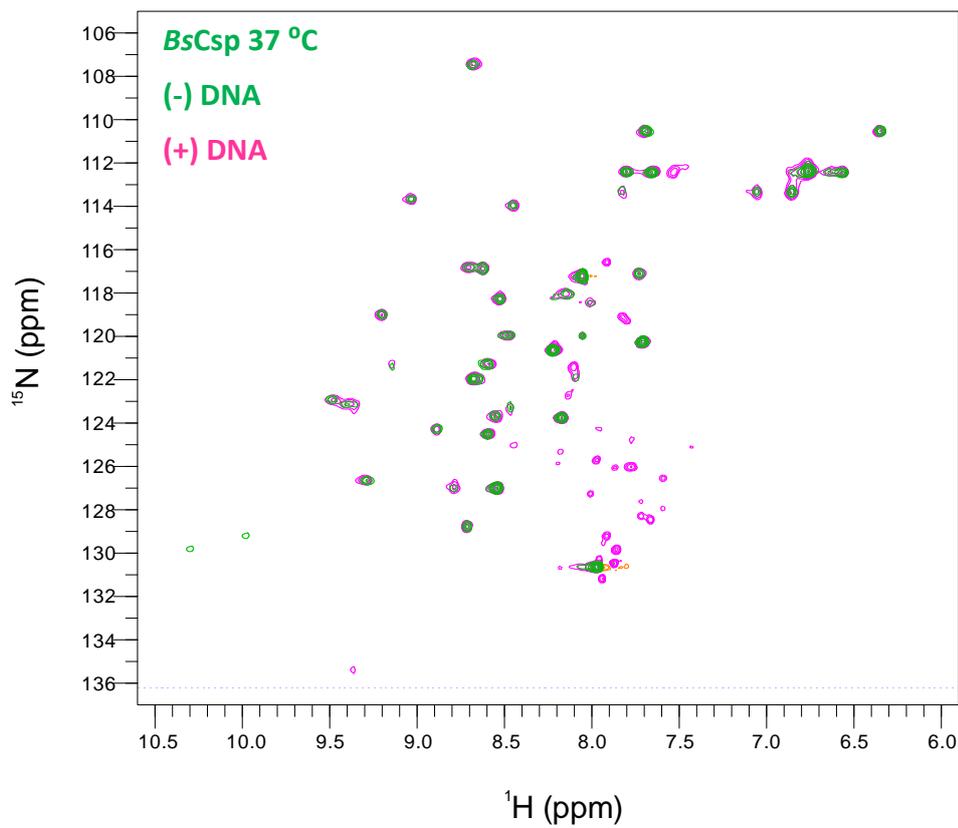
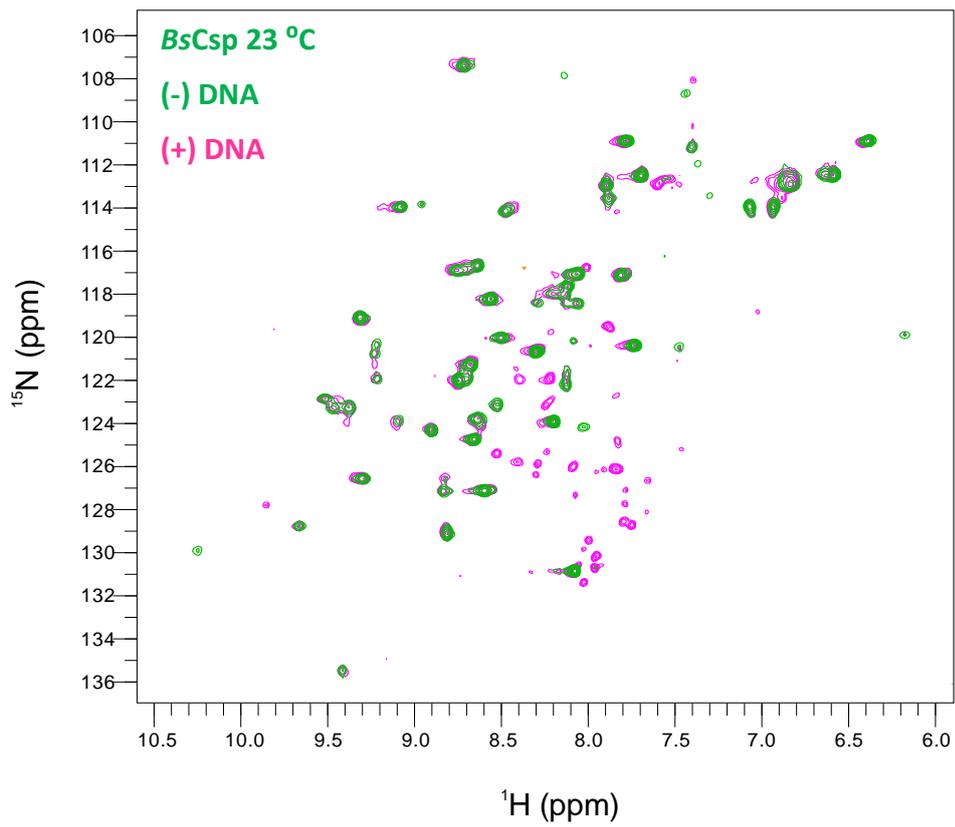


Figure 4.17. HSQC spectra showing the changes to *BsCsp* upon binding of DNA. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer. Green represents without DNA and magenta with DNA

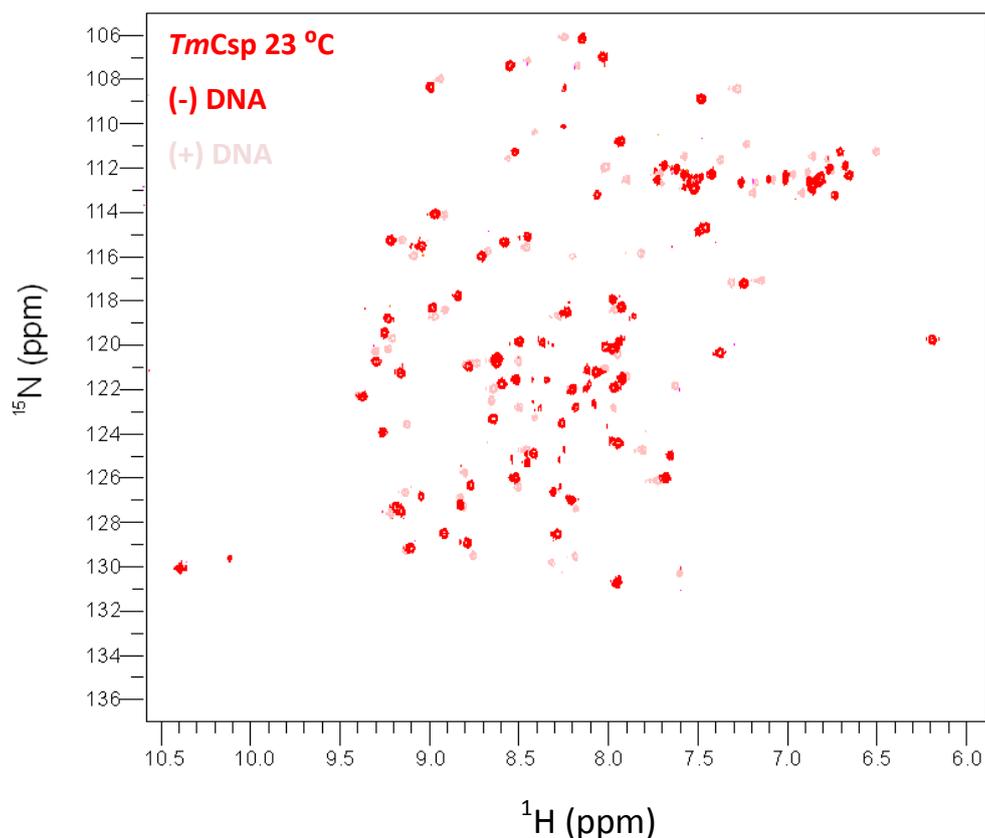


Figure 4.18. HSQC spectra showing the changes to *TmCsp* upon binding of DNA. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz spectrometer. Red represents without DNA and pink with DNA

The mesophilic *BsCsp* exhibits a very different pattern with the chemical shifts of the peaks between DNA-bound and unbound forms appearing almost identical (figure 4.18), suggesting that the DNA-free conformation and DNA-bound conformation are the same. What is interesting is that as temperature was increased (Figure 4.15) a number of signals were seen to disappear from the DNA free *BsCsp* spectra. When DNA is added these signals are restored at the higher temperatures implying DNA binding stabilizes this region. When the *BsCsp* spectrum was compared to a previously assigned *BsCsp* spectrum it was possible to identify the region that this corresponded to. The region comprises the second proposed RNA binding motif on the 3rd β -strand of the Csp and the start of the large loop which connects this to β -strand 4 (237-238) (figure 4.19).

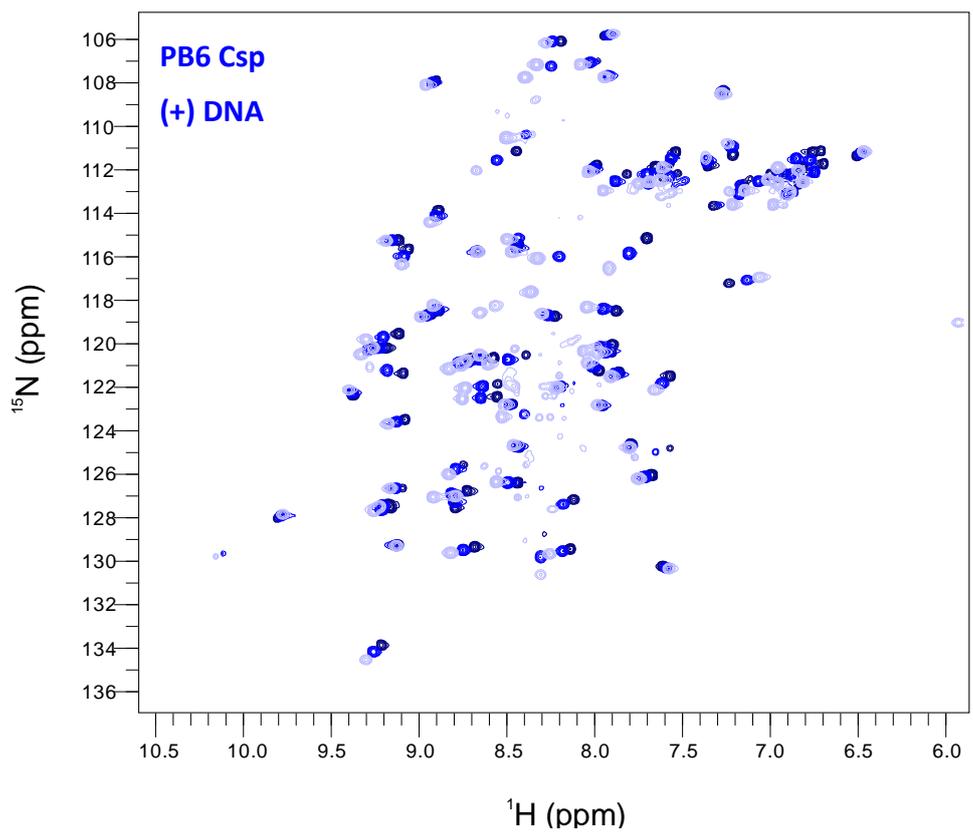


Figure 4.20: Temperature dependence of the HSQC spectra of DNA-bound PB6 Csp. Experiments performed in 63 mM sodium phosphate buffer on a 750 MHz instrument. Lightest shade 10 °C, middle shade 23 °C and darkest shade 37 °C. All spectra scaled relative to position of D₂O signal.

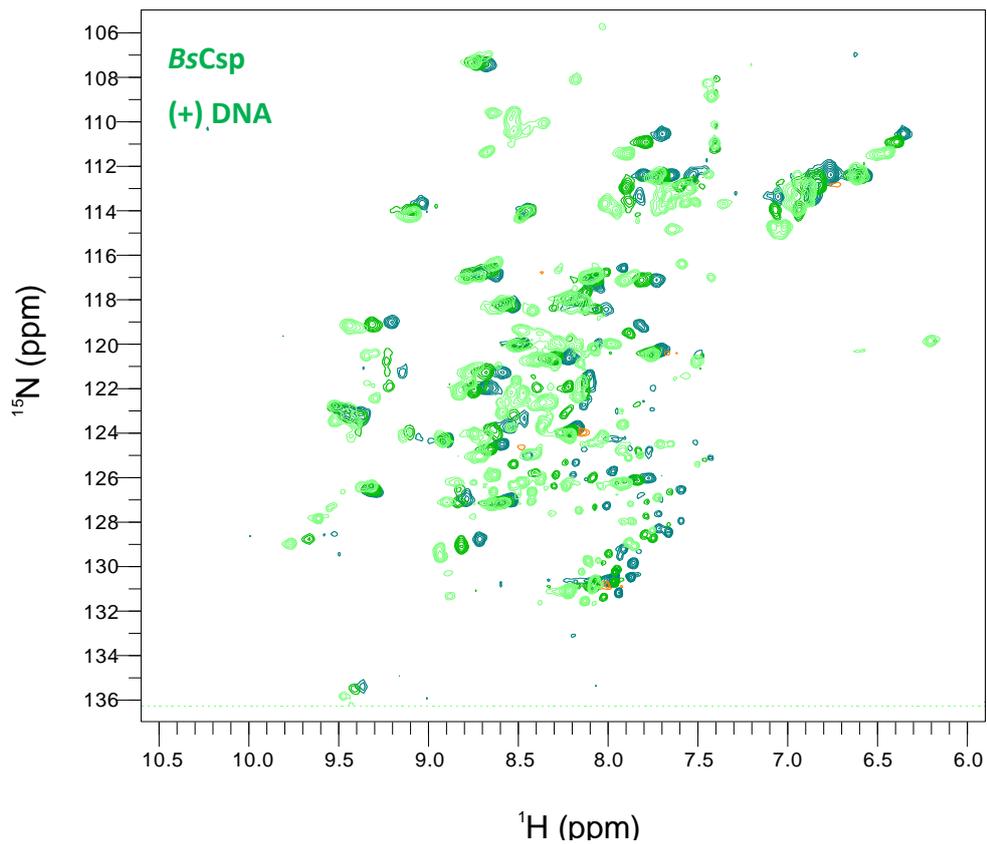


Figure 4.21: Temperature dependence of the HSQC spectra of DNA-bound *BsCsp*. Experiments performed in 63 mM sodium phosphate buffer on a 750 MHz instrument. Lightest shade 10 °C, middle shade 23 °C and darkest shade 37 °C. All spectra scaled relative to position of D₂O signal.

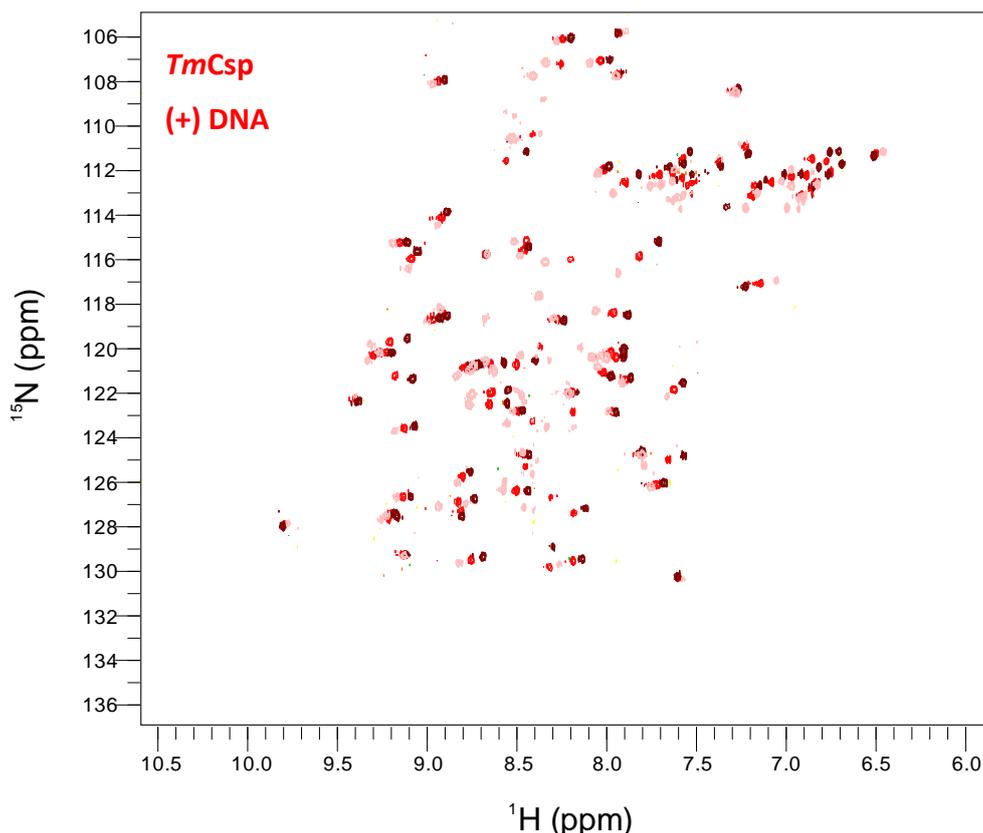


Figure 4.22: Temperature dependence of the HSQC spectra of DNA-bound *TmCsp*. Experiments performed in 63 mM sodium phosphate buffer on a 600 MHz instrument. Lightest shade 10 °C, middle shade 23 °C and darkest shade 37 °C. All spectra scaled relative to position of D₂O signal.

4.5 Conclusions

Kinetic studies showed that cold shock proteins derived from organisms growing at vastly different temperatures actually exhibit very similar equilibrium folding rate constants. This implies that the folding of Csp has been fully conserved during evolution. The unfolding rates reflect the relative thermostabilities of the Csp with a strengthening of the C-terminal region in *TmCsp* suggested in the literature to mediate the differences. Values of ΔG^{F-U} obtained from the stopped flow data were in agreement with those obtained by chemical denaturation.

Using minimal media it was possible to obtain ¹⁵N labelled Csp monomers however for PB6 Csp an alternative method involving expression of PB6 Csp as part of a

fusion protein with MBP was required to provide sufficient amounts of the Csp. The PB6 Csp could be cleaved from the MBP and separated by size-exclusion chromatography. ^{15}N - ^1H HSQC spectra were obtained for *TmCsp*, *BsCsp* and PB6 Csp. There were only relatively minor changes with temperature for *TmCsp* and PB6 Csp suggesting they maintain a single conformation over the range surveyed. Surprisingly from this very preliminary insight into dynamics no obvious changes in temperature sensitivity were seen between *TmCsp* and PB6 Csp. This may bring into question the idea that PB6 Csp has evolved greater flexibility. In *BsCsp* some peaks disappear at the higher temperatures around the largest loop in the Csp structure however these are recovered by the addition of DNA or a decrease in temperature. This suggests that DNA binding stabilizes this part of the protein which makes sense as it is close to a DNA/RNA binding site position. PB6 Csp and *TmCsp* both show substantial changes upon the binding of DNA either suggesting that DNA perturbs the structure of the Csp by an induced fit style mechanism or that it alters the energy levels of conformations so a different conformation is more energetically favoured. It must be noted that while interesting these findings are preliminary and limited to qualitative and global assessments of dynamics. Full assignments of the HSQC spectra would be needed to obtain quantitative data and data specific to regions of the protein. Assessments of rigidity were made here on a qualitative basis but if full assignments were obtained it would be possible to perform t_1 and t_2 relaxation experiments to obtain more information as to which parts of the protein differ in flexibilities.

5. Comparing the mechanical stabilities of cold shock proteins using atomic force microscopy

5.1 Introduction

The previous sections investigated the relative thermodynamic stability of the cold shock proteins. One emerging aspect in the field of protein stability is resistance to mechanical forces and how the elastic properties of biological polymers may be tuned. For medical uses such as tissue restoration, biological polymers are advantageous as they are biocompatible, may better mimic the elasticity of natural tissues and do not have any potential toxicity from metal ions (241). Design of bionanomaterials has been considered for developing treatments for back pain which cause around \$100 billion dollars of productivity losses in the USA each year (242). Such applications would rely on the mechanical properties of these biopolymers at body temperature (242). Crucial to any explorations of proteins for use in bionanomaterials is a greater understanding of how the mechanical properties of proteins relate to their thermostability properties.

One method of investigating the relative mechanical stabilities of proteins is to probe their resistance to applied forces using an atomic force microscope (AFM) to mechanically unfold individual protein molecules. The technique is called single molecule force spectroscopy (SMFS) and has emerged as a powerful tool for studying the mechanical properties of proteins at the single molecule level (136-137)(243). Mechanical unfolding of proteins is perhaps more biologically relevant to the normal working of proteins than chemical or thermal denaturation as it resembles protein unfolding for degradation by proteasomes and how proteins are transported through membranes in the endoplasmic reticulum by the translocon (244). Many proteins also have a function directly related to mediation of forces such as muscle or detection of forces such as force-mediated ion-channels (245).

Applied mechanical forces have the effect of ‘tilting’ the energy landscape by decreasing the Gibbs free energy level of the unfolded state U, from U to U* (figure 5.1). This reduces the energy barrier ΔG^{F-TS} between the folded state F and the unfolding transition state TS allowing this to be more easily overcome by thermal fluctuations (246).

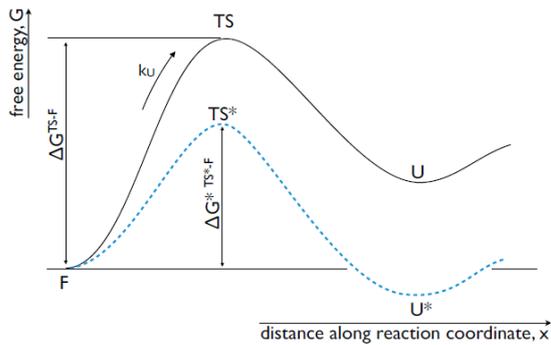


Figure 5.1: Free energy profile showing the effect of force on the energy levels of the folded (F), transition state (TS) and unfolded (U) protein states in a two-state transition. The black line represents the protein in the absence of force and the blue dotted line when force is applied. Taken from (244).

5.1.1 Probing the mechanical stabilities of proteins using AFM

SMFS experiments involve the stretching of proteins by increasing the distance between a protein attached at one end, typically to a gold surface and at the other end to a silicon-nitride cantilever. A schematic is shown in figure 5.2. Protein binding to a gold surface occurs between the side chain of a cysteine residue and the gold through a thiol linkage. C-terminal cysteine residues are genetically engineered onto the C-terminus of the proteins and other cysteine residues often removed to give a defined attachment point to the gold surface in each protein.

The gold surface and cantilever are brought into close proximity so the protein can adhere to the tip. The surface is then withdrawn from the tip at a constant velocity increasing the distance between the gold surface and the cantilever tip, applying an increasing force across the protein. This initially decreases the entropy in the

protein and eventually leads to unfolding of an individual protein domain which then relieves the tension on the protein chain. Further increases in extension leads to the force again increasing until all protein domains are unfolded and a final detachment of the protein chain from the cantilever is observed (247). The typical level of force required to break the non-covalent interactions in a protein domain to cause unfolding is between 20 and 300 pN depending on domain topology (248). The method by which the protein adheres to the silicon nitride cantilever is not clear although there is some suggestion the interaction is electrostatic in nature (249). The magnitude of this force must be more than that of the unfolding of protein domains (250).

The changes in forces applied to the protein trigger deflections in the cantilever holding the silicon nitride tip. These can be monitored by changes in the angle of the deflection of a laser beam focused on to the top of the cantilever tip and reflected onto a photodetector. When the cantilever bends it moves the position of the laser spot on the photodetector. This deflection can be converted to a voltage using a photomultiplier and the applied force deduced from this using the spring constant of the cantilever. For details of the calibration procedure see section 2.10.2.

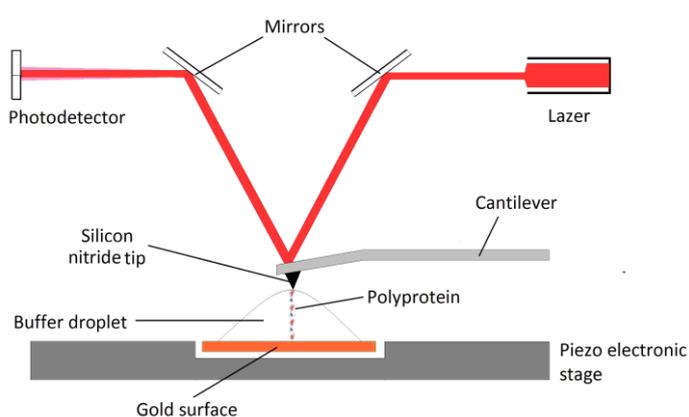


Figure 5.2: Schematic of an AFM microscope being used for single molecule unfolding of a protein. Not to scale.

5.1.1.1 Modelling the stretching of proteins using the worm-like chain model

The stretching of a protein chain can be characterised using a model of polymer elasticity called the worm-like chain (WLC) model (152). As a protein is unfolded the work done to overcome the decrease in entropy generates a restoring force characterised in equation 5.1. x represents the extension of the polypeptide, p represents the persistence length which reflects the distances over which correlations in the polypeptide chain are lost. p can be used as a measure of the stiffness of a polypeptide chain and in a protein typical values are 0.35 - 0.4 nm which is around the length of one amino acid. L_c is the extended length of the polypeptide and k_B the Boltzmann constant.

$$\text{Eqn (5.1): } F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} \right]$$

5.1.1.2 Studies of protein mechanical stabilities using AFM

The first AFM was developed in 1986 by Binnig, Quate and Gerber and since then AFM instruments have become commercially available (251). They were initially developed as a tool for high-resolution biological imaging but are now also regularly used for mechanical studies alongside other techniques such as optical tweezers (243). In 1996 *Mitsui et al.* reported the first mechanical study of protein unfolding using an AFM to investigate the mechanical stability of α_2 -macroglobulin (252). The technique has since been used to investigate the mechanical stabilities of a wide range of proteins and other biological interactions like those present between strands of DNA (252-253). Currently single molecule force spectroscopy (SMFS) using AFM is capable of monitoring unfolding forces from as high as nN to as low as 10 pN shown with studies on the α -helical calcium binding protein calmodulin (243)(248)(254-255).

Early SMFS studies using AFM centred on muscle and other proteins with a defined biological mechanical role. Studies on an RNase enzyme called barnase (2KF3) which

has no known mechanical function showed that proteins to be studied by SMFS did not need to be restricted to only those with a mechanical function. Non-mechanical proteins are likely to show reduced mechanical unfolding forces as was true for barnase (70 pN at 300 nm/s pulling speed) but are still generally within the detection sensitivity of the technique (256).

Since the first study of a protein with no mechanical function, a wide range of other proteins have been studied including notch1, GFP and catenin (257)(278). The mechanical properties of proteins have been found to be greatly affected by residues that exist a long way from each other in the primary sequence but form interactions such as electrostatic and hydrophobic interactions (136). Other experimental studies show that proteins with a high proportion of β -sheet content are generally more resistant to mechanical stresses than those mainly consisting of α -helices due to different hydrogen bonding patterns and impact of the direction of applied force relative to bond geometry (244)(259-263). Proteins with similar 3D folds such as Ubiquitin and protein L can exhibit similar mechanical stabilities which may help in predictions of mechanical stabilities of uncharacterized proteins and the identification of stabilizing motifs (264-265). The mechanical stability of β -sheet proteins can depend on whether they sequentially unzip or shear apart (265-266).

5.1.2 Using polyproteins in AFM studies

In early studies such as the initial investigation by *Mitsui et al.* it proved to be very difficult to distinguish between the unfolding of a single protein domain in the force extension traces and a plethora of low force interactions occurring between the tip and the surface or molecules in the buffer such as salts (244)(267). Unfolding of a single protein domain typically gives an increase in extension of only around 30 nm and as monomeric protein units have small diameters the tip needs to approach very close to the surface to interact with the protein (268). Such close proximity increases the occurrence of non-specific interactions between the surface and

cantilever tip. These interactions may be falsely identified as protein unfolding events and could significantly skew the average unfolding forces reported (252).

One way to address this problem is through the use of polyproteins which contain multiple copies of a target protein to be studied within a single polypeptide chain. The distance between the position of adherence of the tip and the gold surface is further apart in polyproteins than for a single domain and there is a greater extension in length upon unfolding of domains in the construct. This separates the domain unfolding forces from the low extension region highly populated by non-specific events (268-269). Upon approach of the gold surface to the tip, attachment of the polyprotein to the tip can occur. As the surface retracts from the cantilever tip at a constant velocity the entropy of the protein chain reduces and the force across the domains increases until the mechanically weakest domain unfolds, releasing the tension. The process repeats until all domains unfold and finally the protein detaches from the silicon nitride tip. Unfolding of domains occurs sequentially in order of their mechanical stability from lowest to highest producing a characteristic sawtooth pattern of domain unfolding events which can serve as a molecular fingerprint for the unfolding of the construct (243)(247-248) (256)(259)(299). This order is maintained irrespective of the order of domains within the construct (243). An example is seen for a model protein in figure 5.3. For more information on the conversion of force to a voltage signal see 2.10.2.

A consistent spacing in extension is seen between unfolding events of the same protein domain as the same number of amino acids are released from structural elements in each unfolding event leading to the same extension in polyprotein length. This means the distance between the same force on adjacent peaks termed the interpeak distance P_2P can be used as a signature for protein identity.

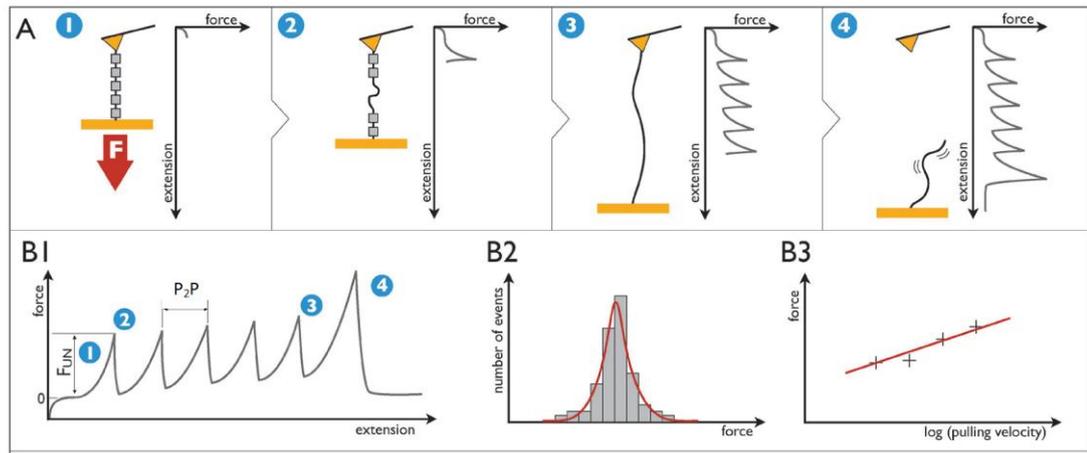


Figure 5.3: Schematic of the methodology of unfolding a polyprotein by AFM including analysis method. (Part A) Stretching of a polyprotein chain containing 5 identical domains by an AFM in force-extension mode. 1. shows approach of the surface to the tip and pickup of a polyprotein, 2 shows the unfolding of a 1st domain, 3 unfolding of further domains in order of mechanical stability and 4 the detachment of the protein chain from the tip. (Part B) B1 shows a model of the sawtooth pattern which gives a mechanical signature for the polyprotein being unfolded. The height of packs represents the value of unfolding force for the domain F_U . The distance between the same position on each peak is termed the interpeak distance. B2 shows that as data for many events is accumulated a bell-shaped distribution of forces is produced which can be fitted with a Gaussian distribution. B3 shows that unfolding forces can be plotted against the log of the pulling velocity.

Domains of the same protein may be hypothesised to unfold at the same force however small differences are present in the mechanical stabilities even of identical domains means simultaneous unfolding of protein domains does not occur. The order of unfolding of equivalent domains within a polyprotein is random and there are other effects on unfolding forces which relate to the folded status of other domains with the polyprotein (270). As each protein domain unfolds the probability of further domains unfolding at a given force decreases as these domains were not mechanically weak enough to unfold previously and there are fewer possible domains remaining to unfold (270-271). There is however a second competing effect where the unfolding of concurrent protein domains increases the length of unfolded unstructured protein chain present. This means compliance increases with each domain unfolded increasing the extension for a given force which decreases the unfolding force with greater event number (270-272). These two effects

compete against each other and can mean, for example that in a polyprotein of 5 repeat domains the 2nd or 3rd unfolding event rather than the 1st event often exhibit the lowest unfolding force. While it is possible for unfolded domains to refold this is extremely unlikely in the timescale of a polyprotein unfolding.

5.1.2.1 Naturally occurring polyproteins

Examples have been found of naturally occurring polyproteins and early studies of protein mechanical unfolding focused on sections of repeating protein domains found in muscle proteins. The first mechanical studies of proteins using AFM concentrated on sections of the extremely large cardiac muscle protein titin which is composed of a large number of fibronectin type III (FNIII) domains and a set of similar immunoglobulin domains (273). Individual titin domains may unfold under physiological stresses and can also refold (274). AFM experiments were then extended to studies of other proteins that occur naturally as tandem repeats of similar domains, such as tenascin and spectrin (275). All of these are thought to unfold by the 'beads on a string model' where domains unfold sequentially from mechanically weakest to strongest. A potential problem when using proteins containing natural repeats is that while the separate domains have similar folds, their mechanical stabilities are not identical. This heterogeneity makes the analysis complex and it becomes very difficult to assign which unfolding events represent the unfolding of which domains.

5.1.2.2 Using genetic engineering to create polyproteins

The solution to issues with the heterogeneity of natural polyproteins has been the generation of engineered polyproteins which contain multiple identical domains of the same protein (276). Typically between 5 and 8 protein domains are present within the chain with a double cysteine genetically engineered at the C-terminal to allow adherence to a fixed gold surface at one end through a thiol linkage (244). This means the geometry of unfolding is from the N to C terminus of the protein.

If multiple target protein copies are included in the construct then a single pickup may yield one or multiple target protein unfolding events (271). Obtaining multiple events from a single pickup is useful when pickup events typically represent 0.5% – 10% of overall traces in an SMFS experiment (277)(299). Cases where multiple polyproteins attach to the tip can be spotted due to the presence of extra peaks or unfolding forces significantly higher than those from other traces, allowing them to be discounted from analysis.

5.1.2.3 Generating polyproteins by chemically coupling monomers

One possible method of generating polyproteins is to first express individual monomer proteins and then couple these individual units into extended chains through cross-linking. Protein cross-linking can be performed through the formation of disulphide bridges between cysteine residues or alternatively bi-maleimide linkers can be used to couple sulphhydryl groups to produce a linkage less dependent on oxidation states (247)(269). The linking of cysteine residues has been used to study the mechanical stability of the protein lysozyme though an additional requirement of this method is that other cysteine residues must be substituted which could have a small effect on native stability (267). A standard protocol for this process is given in (279). With chemical cross-linking it can prove difficult to optimise the correct levels of cross linker and control the orientation of units (300). Whilst the cross-linking reaction is performed in a single step the process generates a highly heterogeneous mixture of different length multimers (269)(279).

5.1.2.4 Generating polyproteins using recombinant DNA technology

An alternate mechanism is to build up the full length construct within the DNA sequence using recombinant DNA technology and then express the whole construct as a single polypeptide chain (280). This provides the advantage that the number of each domain within the construct as well as the order and orientation of each domain within the polyprotein chain can be specifically defined with one defined

product produced. Developing such constructs can prove difficult as the DNA sequences of equivalent protein domains are identical so carefully chosen flanking sequences are required to ensure that each domain has a defined position within the gene sequence.

One way of achieving this is to use a series of unique restriction sites interspaced between each domain so that each domain becomes a defined cassette with a specified position within the construct (136)(244)(246)(264)(267-268). The unique restriction sites can be added on to the 5' and 3' ends of each cassette sequence by PCR. The plasmid and insertion sequence are both cut with the same restriction enzyme to generate complimentary sticky ends with 5' phosphates removed from the vector to prevent re-ligation (248). The insert is ligated into the plasmid, then further cycles conducted to build up the full construct with care taken to ensure all the sequence is kept within the same reading frame. This approach has advantages in terms of flexibility in that individual domains or sets of adjacent domains can be substituted in a single cycle without having to rebuild significant amounts of the full plasmid (268). It also allows the order and orientations of domains to be controlled. With this method however, changes to multiple non-adjacent cassettes require each individual cassette to be modified making it highly time intensive. The restriction digest method also requires considerable amounts of time to generate the initial construct as each replacement requires another cycle of digest and ligation (281). A recent modification to this method has been the use of different codon preferences for repeating sequences to allow more specific binding of primers within protein domains which would otherwise be difficult with the large sections of identical sequences (298). This could make sequencing and modifications such as site directed mutagenesis more straightforward on this type of polyprotein.

5.1.2.5 The Gibson assembly method of rapidly generating polyproteins

The main difficulty with producing polyproteins using a restriction enzyme based approach is that each cassette must be added separately. A way to resolve this is a recently developed strategy called Gibson assembly developed by *Gibson et al.* which is designed to generate a construct comprised of multiple cassettes in a single 'one pot' synthesis reaction (282). The Gibson assembly reaction can dramatically decrease the amount of time required to generate complex constructs relative to repeated rounds of restriction digests but still achieves the advantages of a defined order of domains (283). *Hoffman et al.* showed the potential of this technique for constructing polyproteins for SMFS. In the Gibson assembly reaction each domain initially has flanking sequences added by PCR. The flanking sequences are designed to be unique to specify the location within the polyprotein chain and contain an overlap with the sequence designed for the adjacent domain (281). An exonuclease is used to remove bases from the 5' end of each sequence to produce complimentary ends. These can be joined together by a DNA ligase and a DNA polymerase fills in any of the gaps. This means that one reaction is required rather than a whole series of cycles as with the restriction digestion approach. Two significant difficulties are that the regions near the boundaries of the domains are prone to increased levels of mutations and that a wider range of different products are produced so further screening is required to identify the correct products. Making changes to completed constructs may again prove complex as PCR primers can only reliably be used to amplify sequences from the unique linker regions meaning that adjustments within the middle of protein domains that may be desired for mutational studies are very difficult.

5.1.2.6 Using marker proteins of known mechanical stability within polyproteins

Polyproteins may be made as homopolyproteins comprising solely of repeating copies of the target protein domain or heteropolyproteins which contain different domains. In heteropolyproteins a domain of known mechanical stability can be included to act as a positive control for a polyprotein attaching to the tip and a reference marker to help verify reported target protein unfolding forces (248)(298).

The domains are separated by short linker regions which prevents interactions between domains interfering with the folding of the individual domains (298). A commonly used marker protein is the 27th Immunoglobulin domain of titin I27 as it is more chemically and mechanically stable than most Ig domains (244). I27 is an 89 amino acid β -sandwich like fold has high mechanical strength, elasticity and well characterised mechanical unfolding patterns (255). Using polyprotein constructs which contain I27 gives a clear unambiguous reference that increases certainty in forces obtained for unknown proteins and allows results from different studies to be directly compared (244). Other alternative marker proteins such as immunoglobulin G domain GB1 and fibronectin have been used though I27 will be used in this study (284)(298). A further advantage of using marker proteins is that some proteins such as barnase do not express well as a homopolyprotein but express well when interspaced with marker domains in a heteropolyprotein (256).

5.1.2.7 Effects of the incorporation of protein domains into polyproteins

One key question was if incorporation of the monomers into polyprotein chains had a significant effect on the stability of the monomer units. To investigate this *Best et al.* compared I27 monomers, I27 homopolyproteins, barnase monomers and a polyprotein with both I27 and barnase (256). They used NMR to show that the dynamics of the domains were unaffected by incorporation into a polyprotein. The relative thermostabilities of monomer domains and domains incorporated into a polyprotein were investigated with chemical equilibrium denaturation for barnase and I27. The Gibbs free energy of unfolding value ΔG^{F-U} was within 10% for the free and polyprotein incorporated forms for both types of domain. A study with T4 lysozyme however did show that catalytic activity is reduced 10-fold upon incorporation into polyproteins suggesting some restriction of protein movements (267).

5.1.3 Differences between unfolding proteins by chemical denaturants and applied forces

One of the conclusions of the study on barnase by *Best et al.* generated some debate. The group determined rate constants of mechanical unfolding k_u at various forces and extrapolated this to obtain a rate constant value at zero applied force (256). They showed that it matched closely with an extrapolation of a chevron plot generated with different concentrations of chemical denaturant (256). The hypothesis was proposed that mechanical and chemical unfolding occurred by the same mechanism. More recent comparisons performed on mutants of I27 however showed a 5-fold difference between mechanical and chemical rate constants (285). Molecular dynamics simulations on wild type I27 indicated that mechanical unfolding proceeds initially by detachment of β -strand A from I27 to make an intermediate (258). Mutations on this strand markedly affect the thermodynamic stability of I27 but do not appear to affect the mechanical stability. The AFM data appears to resemble unfolding of this intermediate as a rate limiting step while the chemical denaturation value appears to be from the native state to unfolded state. This suggests that the two pathways and rate constants are not analogous and it is now known that the AFM unfolding data represents a measure of kinetic rather than thermodynamic stability. The earlier finding that the values appeared the same for barnase and I27 is suggested to be because in both cases the transition states of unfolding are positioned near the native state (286). There may be differences between the pathway of unfolding of proteins under extreme force and refolding under no force due to hysteresis that are not observed with chemical unfolding and refolding (256).

The unfolded states produced when a protein is unfolded mechanically and chemically may also not be equivalent. Using NMR *Shortle et al.* were able to uncover that when *staphylococcal* nuclease was denatured by 8 M urea significant levels of secondary structure remained (287). Molecular dynamics simulations of the unfolded states of ubiquitin generated by force and chemical denaturation suggest that a fully extended protein chain was produced by mechanical unfolding

and a more globular structure remained with chemical denaturation (288). They do suggest however, that many of the observed secondary structure elements they predict in the chemically unfolded ubiquitin appeared non-native. Studies of ubiquitin unfolding have also shown that pulling direction can have a large effect on unfolding forces (289). This directionality of unfolding forces is not observed with chemical and thermal stabilities. The geometry of hydrogen and ionic bonds may be crucial to protein mechanical stability and set regions of ionic bonds can form so-called mechanical clamps which may mediate these differences between chemical and mechanical stability (244)(290). Mechanical clamp regions help reduce stretching when a force is applied. This means that mechanical stability information cannot directly be inferred from chemical data though strong trends may exist across both.

No distinctive motifs of amino acid sequences are associated with temperature adaption of mechanical stability however regions of stabilising salt bridges termed mechanical clamps may increase mechanical stability of proteins (292). The types of interactions regulating protein thermal and mechanical stability are similar such as hydrogen and ionic bonds and well as hydrophobic interactions though their relative importance may differ. It must also be noted that mechanical unfolding is dependent on kinetics rather than thermodynamics (136)(190). A previous study by *Li et al.* has showed while that I27 exhibits significantly greater thermal stability than I28, I28 exhibits greater mechanical stability raising some questions over the extent which it is possible to make inferences about mechanical stability (291).

5.1.4 Previous AFM studies on Csps

In a previous study Tych and Hoffman were able to construct a polyprotein containing three copies of the cold shock protein from the hyperthermophilic bacteria *Thermotoga maritima* (*TmCsp*) interspaced by I27 domains (I27-*TmCsp*)₃-I27. They found that *TmCsp* has a defined mechanical fingerprint with an unfolding

force of 70 (± 3) pN at a pulling speed of 200 nm/s and a contour length ΔL_c of 23.5 nm. This is shorter than 28 nm observed for I27 though this is expected due to the small size of Csps. They also looked at the temperature dependence of the mechanical stability of *TmCsp* using a custom built AFM (136). Work by *Schonfelder et al.* however, showed that when the hyperthermophilic homologue *TmCsp* was unfolded using force clamp SMFS very slowly, mechanical unfolding proceeds by heterogeneous unfolding pathways with some containing up to 4 individual steps (284). Under the high pulling velocities applied in this investigation a single two-state unfolded model will be applied.

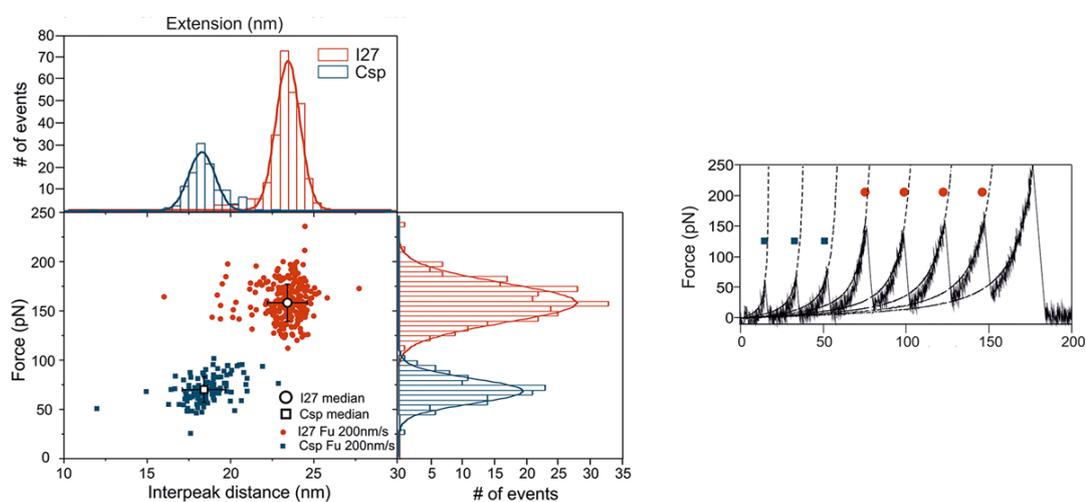


Figure 5.4: Unfolding data for $(I27-TmCsp)_3-I27$. Left: Examples of a scatterplot displaying unfolding force and interpeak distance data for samples of $(I27-TmCsp)_3-I27$ polyprotein pulled at 200 nm/s and a temperature of 23 °C. I27 peaks are indicated in orange and *TmCsp* peaks indicated in blue. Right: Force-extension trace displaying the mechanical fingerprint of $(I27-TmCsp)_3-I27$. Taken from (136)

While the thermostabilities of different proteins have been compared no comparisons of the mechanical stabilities of different temperature adapted homologues have been performed. A mechanical clamp is seen in the hyperthermophilic homologue *TmCsp* but not in the mesophilic or psychrotrophic homologues so this may be important in their relative thermostabilities (137). The polyproteins $(I27-TmCsp)_3-I27$ whose mechanical stability is shown in figure 5.4 and $(I27-BsCsp)_3-I27$ were generated by Dr Toni Hoffmann.

5.1.5 Aims

The mechanical stability of a cold-adapted protein had not been previously investigated. Of the 3 cold-adapted Csps (section 3.2.1) the cold adapted Csp with the most reduced thermostability which was PB6 Csp from the psychrotroph *Psychrobacter* Sp6 was selected for AFM analysis (see section 3.4.7). This csp may potentially give the greatest difference in mechanical strength from the mesophilic *BsCsp*. The first objective was to attempt to generate an analogous polyprotein (I27-PB6 Csp)₃-I27 to those used for the other Csps which contained PB6 Csp for comparison. It was then important to verify if PB6 Csp had a defined mechanical signature.

A large number of force-extension experiments were performed to attempt to verify if the lower thermostability of PB6 Csp reflected in a lower mechanical stability. It was observed previously in section 3.4.8 that the arbitrary selection of room temperature may hide underlying properties of the mechanical stability landscape. AFM experiments are therefore performed at a series of different temperatures on a custom built AFM with an inbuilt temperature control and pump (136). This allows force extension traces to be obtained at temperatures between 5 °C and 40 °C with the limitations due to electronic sensitivities and droplet freezing or evaporation. Performing the experiments at different temperatures improves the reliability of comparisons between the properties of the Csps and can also allow differences in how sensitive the mechanical stabilities of the Csps are to temperature.

Overlaying the temperature range which could be investigated on to the thermostability curves from section 3.4.7 shows that *TmCsp* remains highly thermostable throughout the temperature range exhibiting only small differences in thermal ΔG^{F-U} (see figure 5.5). With *BsCsp* and PB6 Csp there is a substantial decrease from the maximal ΔG^{F-U} within this range. If the thermodynamic and

mechanical stabilities do correlate due to the similarities of the Csp topologies then there could be differences in the temperature sensitivity of the Csp.

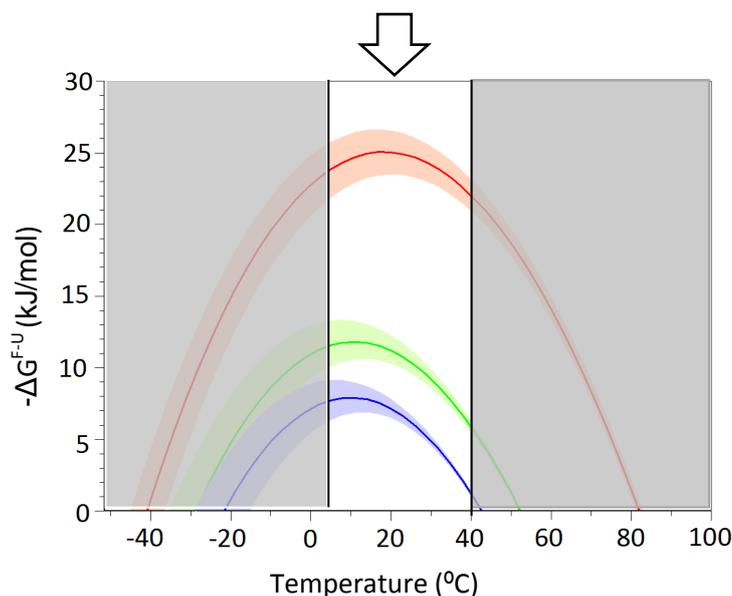


Figure 5.5: The range of the thermostability profile surveyed by AFM experiments between the two black lines indicated by arrow above.

This section of the thesis was performed in conjunction with Dr Toni Hoffmann who generated the (I27-*BsCsp*)₃-I27 and (I27-*TmCsp*)₃-I27 constructs and Dr Kasia Tych who performed the AFM unfolding experiments for these two constructs at 5 °C, 23 °C and 40 °C. The next sections will detail production of the (I27-PB6 Csp)₃-I27 polyprotein construct and SMFS studies for (I27-PB6 Csp)₃-I27 performed at 5 °C, 10 °C, 23 °C and comparison with all 3 (I27-PB6 Csp)₃-I27 constructs at both 14 °C and 32 °C.

5.2 Results

5.2.1 Production of the *TmCsp* homopolymer

Two methods for the production of polyproteins were used during this study. Firstly a restriction enzyme based approach was used to produce a DNA construct comprised of 5 identical *TmCsp* domains. The second polyprotein was a

heteropolyprotein containing 7 domains alternating between the marker protein I27 and the Csp sequence from the psychrothroph *Psychrobacter* PB6 Csp using the Gibson assembly mechanism (section 5.1.2.5).

The first strategy attempted was to create homopentameric Csp polyproteins which contained 5 copies of the respective Csp starting with *TmCsp*. *TmCsp* had already been expressed in the polyprotein (I27-*TmCsp*)₃-I27 but had not been expressed as *TmCsp*₅. A pET3a plasmid containing the DNA sequence for I27-*TmCsp*₃-I27-*TmCsp*-I27 was obtained from Dr Toni Hoffman. To obtain the *TmCsp* homopentamer required the substitution of position 1 and positions 5-7 as indicated in figure 5.6.

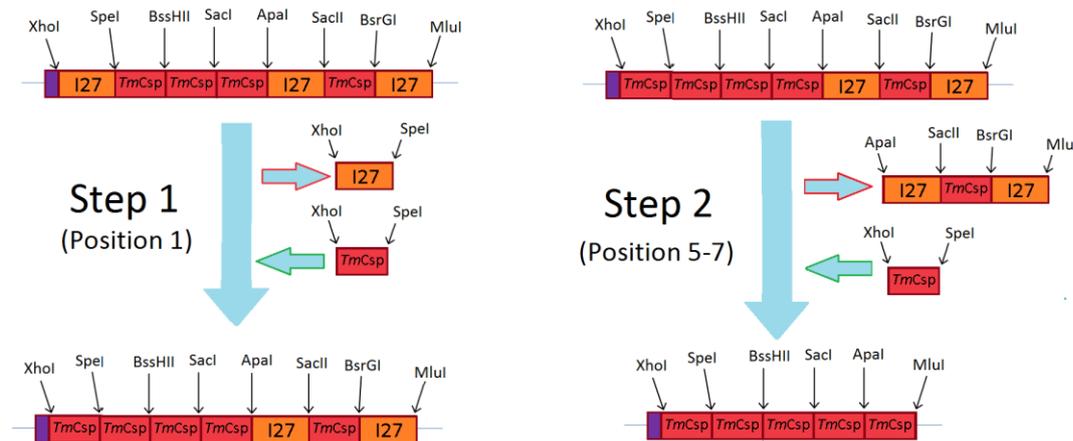


Figure 5.6: Schematic representing the two-step molecular biology approach involved in the production of the *TmCsp* homopentamer *TmCsp*₅. The DNA sequences for I27 domains are shown in orange, *TmCsp* domains in red and an affinity purification tag shown in purple. The location of restriction enzymes sites is also indicated. Step 1 substitutes the I27 in position 1 of the DNA sequence provided by Dr Toni Hoffman (I27-*TmCsp*₃-I27-*TmCsp*-I27) for a *TmCsp* sequence. Step 2 then replaces positions 5-7 with a *TmCsp* sequence. Sections of DNA removed in each step are shown with a red outlined arrow and those inserted with a green outline arrow. The rest of the pET3a plasmid is not shown.

The process required generation of *TmCsp* cassette sequences for insertion at positions 1 and 5. The correct flanking restriction enzyme cut-sites were added to the end of a *TmCsp* sequence using specific primers (see methods 2.2.1) and the

sequence amplified by PCR. Amplification of the correct region was shown by bands on an agarose gel (figure 5.7).

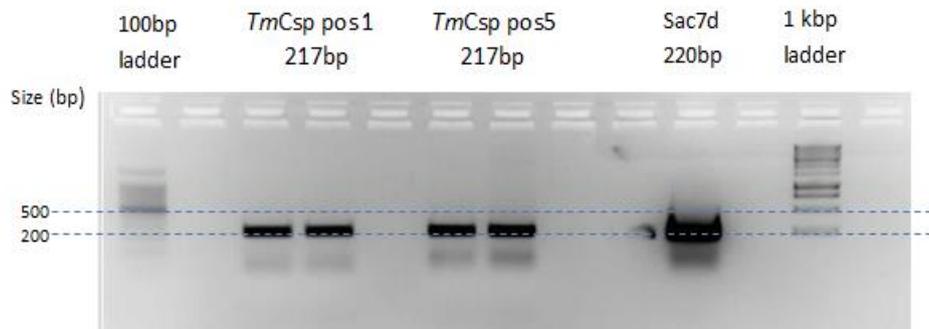


Figure 5.7: 2% Agarose gel displaying amplification of PCR templates of *TmCsp* cassettes for position 1 and 5 in duplicate. A reference protein of similar size to *TmCsp* (*Sac7d*) is included as a positive control

In the first step pET3a I27-*TmCsp*₃-I27-*TmCsp*-I27 was cleaved with *Xho*I and *Spe*I restriction enzymes to remove the I27 monomer for position 1. The products were run on an agarose gel and this showed a band of the correct size for I27 and a much larger band corresponding to the remainder of the pET3a plasmid. The position 1 insert of *TmCsp* was ligated into the pET3a plasmid, transformed and amplified in *E. coli* SURE2 cells. Polyprotein sequences are highly repetitive so it was important to choose a strain such as SURE2 which is recombinase deficient to minimise the chance of DNA rearrangements. The pET3a *TmCsp*₄-I27-*TmCsp*-I27 DNA was amplified and extracted for the second step. The plasmid was then digested with *Apa*I and *Mlu*I restriction enzymes. This released a DNA segment of I27-*TmCsp*-I27 which can be seen to be significantly larger than the smaller band present in figure 5.8 left side. Ligation of the *TmCsp* sequence for position 5 was performed as before and the plasmid sent for sequencing. This was performed from sites 5' upstream of the insertion region on the coding strand and reverse sequencing performed on the non-coding strand from the 3' end of the polyprotein sequence to cover the full construct. The sequencing confirmed that the correct *TmCsp*₅ DNA sequence had been produced and the full construct was in-frame.

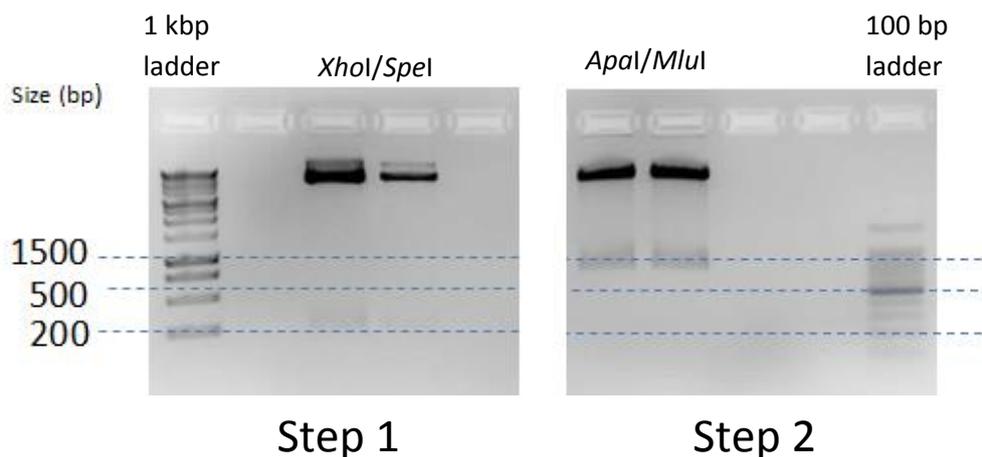


Figure 5.8: Agarose gel displaying restriction digest products used to generate *TmCsp₅* plasmid. Removal of I27 from position 1 of I27-*TmCsp₃*-I27-*TmCsp*-I27 shown left and removal of I27-*TmCsp*-I27 shown right from the pET3a *TmCsp₄*-I27-*TmCsp*-I27 plasmid. The restriction enzymes used are indicated above the lane with each digest performed in duplicate.

5.2.2 Expression of the *TmCsp* homopolymer

The *TmCsp₅* construct was expressed in *E. coli* using IPTG and a T7 viral promoter. Significant perturbation of growth rates were seen in the bacteria with this construct. It was not possible to express this construct so an additional construct I27-*TmCsp₅*-I27 was trialed. It was hoped that the presence of I27 would increase expression of the polyprotein. After pull-down assays on the expressed proteins an unusual result was noticed. The expected size of construct was 66 kDa however the most pronounced product observed by SDS-PAGE gel was below 10 kDa (figure 5.9). Sequencing results revealed that the inserted sequence in the plasmid had been modified by the bacteria to just a single I27 domain. This was unexpected as the *E. coli* strain BLR (DE3) is recombinase deficient however it is likely that any bacteria that lost Csp sequences through mutation would grow faster than other bacteria out-competing those with the full construct. The stringency of the system was increased by using increased supplements of glucose to prevent endogenous protein expression prior to induction. No protein of the expected size of the polyprotein was then observed to be over-expressed with either IPTG induced LB or auto-induction media. Previous work had shown *TmCsp* to express in polyprotein

constructs so it appeared that the expression issue was with adjoining Csps. In the next sections all constructs contain Csp sequences interspaced with I27 domains.

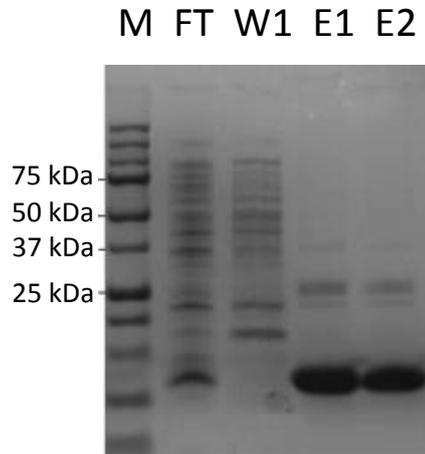


Figure 5.9: Nickel pull-down assay of I27-*TmCsp*₅-I27 expression with IPTG in *E. coli* BLR [DE3] pLysS cells. Lane FT represents endogenous *E. coli* proteins which did not bind the Ni beads. W1 shows the 1st wash step of the beads to remove non-specifically bound proteins. E1 and E2 show proteins eluted by two successive imidazole washes.

5.2.3 Using Gibson assembly to generate (I27-PB6 Csp)₃-I27 polyprotein

The construct (I27-*TmCsp*)₃-I27 has already been mechanically characterised so the aim of this section was to generate an equivalent construct for the psychrotrophic PB6 Csp characterised in sections 3-4 with the same polyprotein scaffold (I27-PB6 Csp)₃-I27. To accelerate this, the alternative method of Gibson assembly was used to generate this polyprotein as shown in the schematic (Figure 5.10). Each PB6 Csp and I27 domain was separated by unique linker regions (for sequence see appendix S2).

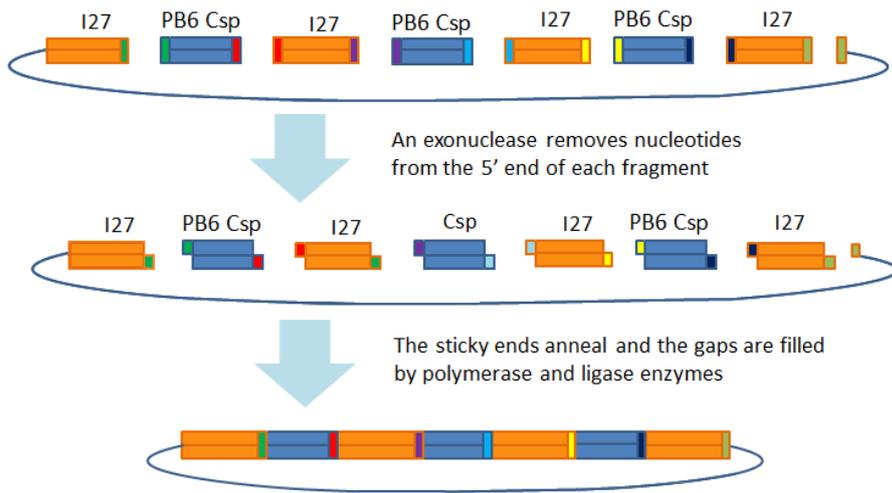


Figure 5.10: Schematic showing production of an (I27-PB6 Csp)₃-I27 polyprotein DNA construct in a pET3a plasmid using the Gibson reaction: I27 is shown in orange and PB6 Csp in blue. The other colours denote specific linker regions.

Specific flanking sequences were added to each individual domain and the backbone using unique pairs of PCR primers. The amplification of the cassettes was confirmed with PCR and a gel extraction kit used to remove enzymes and left over nucleotides (see figure 5.11). The concentrations of cassettes obtained were between 20 ng/ μ l to 75 ng/ μ l. Volumes were calculated to add an approximately equal molecular ratio of each cassette to the final mixture. See section 2.2.9 for more details of the Gibson assembly procedure.

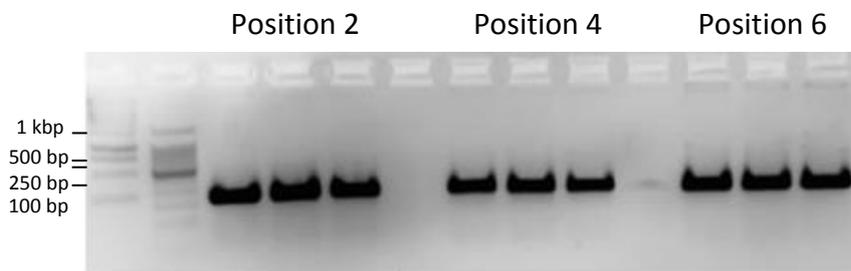


Figure 5.11: 1.5 % Agarose gel displaying amplification of PB6 Csp cassettes by PCR for the different sites within the polyprotein

The Gibson product was transformed into DH5 α cells and a range of colonies selected from the agar plate. As colonies each originate from the same cell, homogeneity of the plasmid within each colony would be expected. Colony PCR using sequences from either end of the polyprotein region of the plasmid were used to amplify the polyprotein section. This Gibson assembly reaction was relatively complex in containing a backbone of pET3a-I27 as well as a further 5 domains (2 of I27 and 3 of PB6 Csp). The vast majority of colonies grown from cells transformed with the Gibson reaction products did not contain the full construct. The size of the construct produced could be determined by colony PCR which first involved extracting the DNA from cell samples and then amplifying the construct DNA from primer sites flanking the insertion region. Those containing only re-ligated vector or incomplete polyprotein could be distinguished by size using SDS-PAGE and those with the full construct distinguished and linked back to their host colony (see methods 2.2.10). From figure 5.12 lane 7 and 17 were selected for commercial sequencing as they contained bands at approximately the desired size of 3 kbp as indicated by the positive control in the lane marked + on the right.

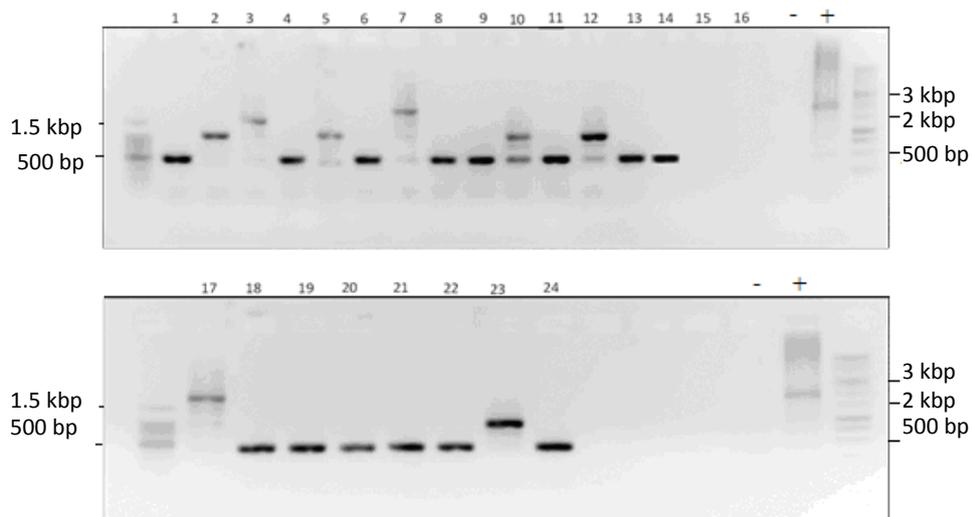


Figure 5.12: Products of the amplification region of the Gibson reaction shown on a 1.5% agarose gel. Each lane represents the DNA product from colony-PCR of a separate colony. The lane marked with a + symbol is a positive control showing a different polyprotein (I27-*TmCsp*)₃-I27 of the same desired size. The – symbol indicates a negative control with no construct.

During the Gibson assembly reaction an exonuclease removes nucleotide from the 5' end of each cassette. This however gives a high risk of mutations arising around the joints between the individual domains where the polymerase fills in the gaps. Of the 2 plasmids found to contain full constructs (figure 5.12) both had mutations but at different locations. One contained a mutation in the first Csp sequence and the other in the last I27 sequence. These constructs however did contain most of the correct sequence and the region of position 2 to position 6 which was found to be correct could be amplified as a single unit. This could be combined with a fresh unit for position 7 which did not contain the mutation (figure 5.13). A further 24 transformed colonies were selected from the plated transformed colonies and colony PCR performed on the products. One construct was found (lane 1) which contained the full correct sequence (figure 5.14).

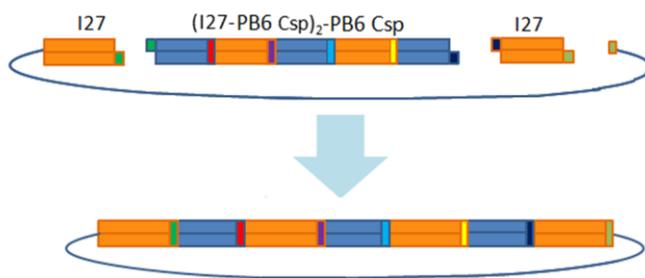


Figure 5.13: Using a second stage to the Gibson reaction to replace the final I27 domain containing a mutant with a fresh I27 domain. I27 in orange, Csp in red and unique linkers in other colours

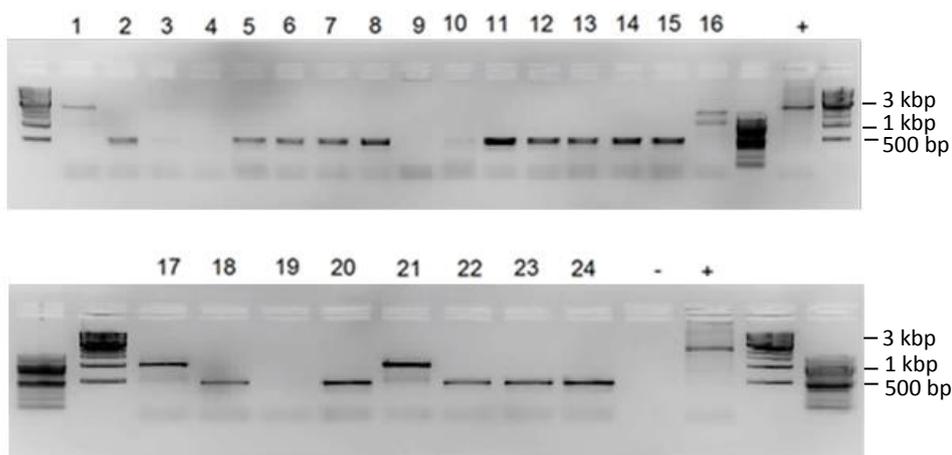


Figure 5.14: Agarose gel of colony PCR products displaying the sizes of the DNA sections within the polyprotein region. + indicates a polyprotein construct of $(I27-TmCsp)_3-I27$ which is approximately the same size as the desired polyprotein.

5.2.4 Expression and purification of the (I27-PB6 Csp)₃-I27 polyprotein

The pET3a plasmid containing (I27-PB6 Csp)₃-I27 was transformed into BLR(DE3) pLysS cells. Small scale expression trials revealed that the polyprotein is expressed strongly between 12 and 22 hours after IPTG induction (figure 5.15). A period of 16 hours was used for the polyprotein expression as this gave strong expression and high cell density. The cells grew to a greater density than for Csp monomer expression suggesting that the I27 was stabilising or the polyprotein was less toxic to cells.

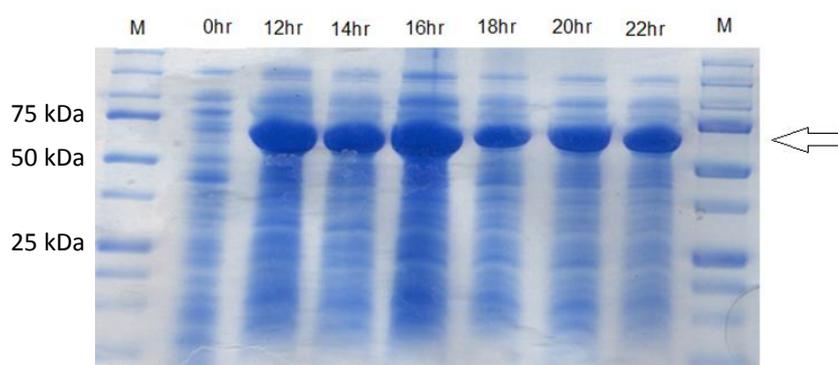


Figure 5.15: Expression trial of the (I27-PB6 Csp)₃-I27 polyprotein in *E. coli* BLR [DE3] PLYS cells. The arrow indicates the expected size of polyprotein which is 66 kDa

The (I27-PB6 Csp)₃-I27 polyprotein was expressed in 5L LB medium and purified by a similar 3 step process as the monomer (sections 2.3 and 2.4). Early attempts showed some degradation of the polyprotein so the method was adjusted to perform the full purification procedure including dialysis steps within 5 days to reduce this. The initial Ni-NTA purification was performed using a batch process as per the small scale pull-down experiments with all the volumes increased by 10 times (methods 2.3.4). A large band was observed at the correct size of 66 kDa.

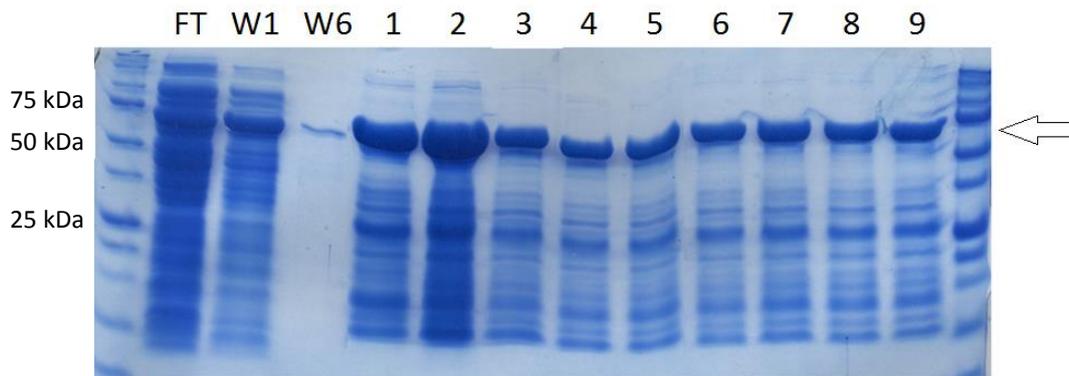


Figure 5.16: SDS-PAGE of Ni-NTA separation of the proteins from large scale expression of $(I27-PB6\ Csp)_3$ -I27 polyprotein. FT indicates flow-through, W1 and 6 the 1st and second wash steps and the remaining lanes show elution with imidazole. Arrow indicates expected size of 66 kDa

Csps are DNA/RNA binding proteins so as with the monomers, it remained important to remove the DNA/RNA from the sample by ion-exchange. The binding of ligands has been shown to affect protein stiffness and increase unfolding forces through studies of small ubiquitin-like modifier whose unfolding forces increased by 30% when a complementary ligand was added (214). Large DNA/RNA peaks were observed in the ion-exchange purification step as each polyprotein has 3 potential polynucleotide binding sites (Figure 5.17).

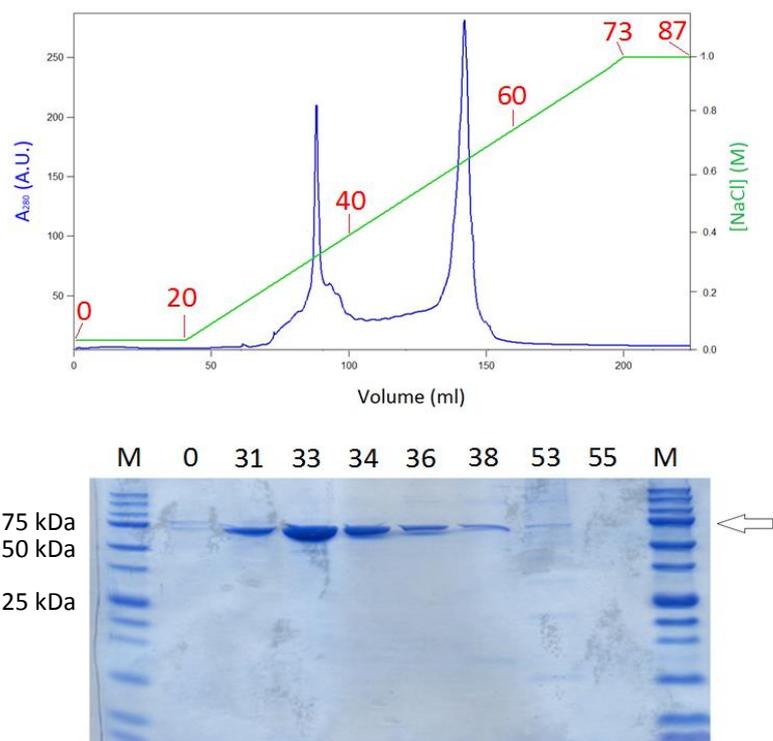


Figure 5.17: Ion exchange separation of $(I27-PB6\ Csp)_3-I27$ and SDS-PAGE showing fractions eluted from a resource Q column. Top: Blue line and left axis represents the absorbance at 280 nm. Green line represents the NaCl concentration applied to the column. Elution fractions were 2 ml in volume for the first 40 ml and 3 ml subsequent to that with fraction numbers indicated on graph. Bottom: SDS-PAGE gel of selected fractions with arrow indicating expected polyprotein mass

Size exclusion chromatography is particularly useful in the purification of polyproteins as it can separate truncated proteins which may still have maintained a hexa-histidine tag and potentially a similar charge balance to the complete construct (293). Size exclusion chromatography of $(I27-PB6\ Csp)_3-I27$ gave a large peak which was confirmed to be the correct size by SDS-PAGE (figure 5.18).

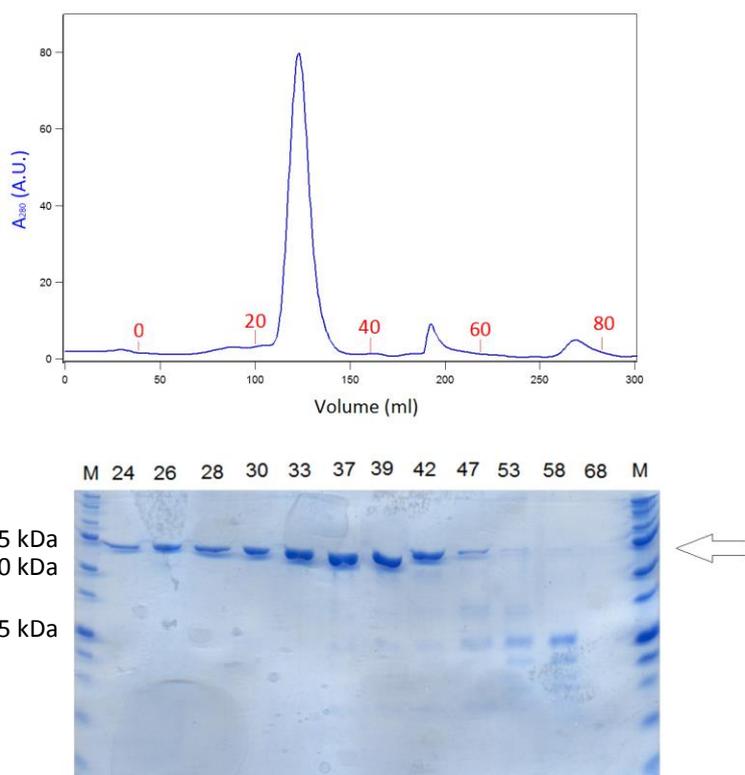


Figure 5.18: Size exclusion purification of $(I27-PB6\ Csp)_3-I27$ using a Sup75 column. Top: Absorbance profile with blue line representing the absorbance at 280 nm. Bottom: SDS-PAGE showing relevant fractions. Fractions were 3 ml in volume throughout. Arrow indicates expected mass of 66 kDa.

5.2.5 Confirmation of generation of the $(I27-PB6\ Csp)_3-I27$ polyprotein

Production of the correct polyprotein was confirmed by MS-MS which gave a mass of 66 kDa matching that predicted based on the primary sequence. The polyprotein was analysed by CD to confirm that domains within the polyprotein were folded. The ellipticity spectra at 4 °C showed peaks around 200 nm and 220 nm which were lost upon heating of the protein sample to 85 °C (figure 5.19). The polyprotein was not refoldable after thermal denaturation though this may be due to aggregation of unfolded polyprotein. During heating of the polyprotein construct using thermal ramping, 2 separate unfolding transitions could clearly be seen which fits with a combination of the two different T_m values of the two respective domains (Figure 5.20). Both I27 and PB6 Csp showed a slightly lower T_m value than their respective monomers although this has been observed in other polyproteins (Table 5.1)(256).

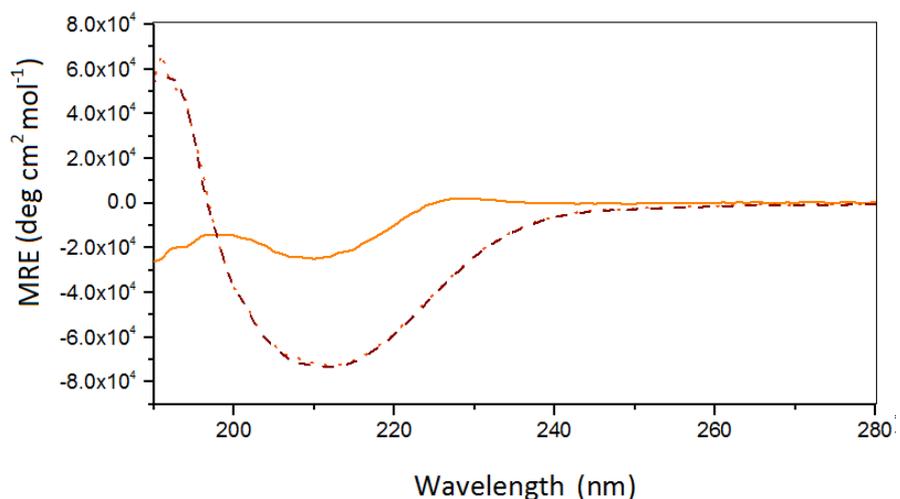


Figure 5.19: CD spectra of (I27-PB6 Csp)₃-I27. CD trace of (I27-PB6 Csp)₃-I27 at 5 °C solid gold line, 85 °C dashed brown line and 5 °C after heating to 85 °C dotted dark orange line. MRE – mean residue ellipticity

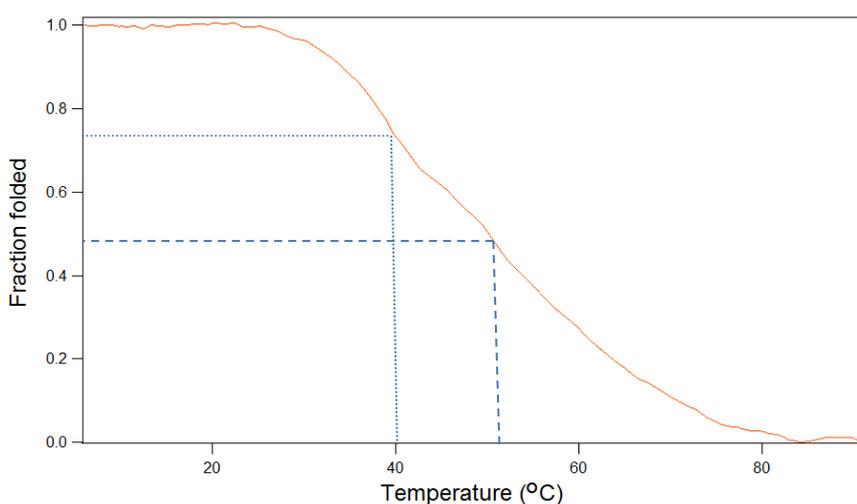


Figure 5.20: Thermal denaturation of (I27-PB6 Csp)₃-I27 with dotted line for PB6 Csp unfolding and dashed line for I27 unfolding

Table 5.1: Comparison of T_m values for PB6 Csp and I27 as monomers and within (I27-PB6 Csp)₃-I27

	Monomer T_m	Polyprotein T_m
I27	54.8	52.2
PB6 Csp	43.2	40.5

5.3 AFM mechanical unfolding

As per (136) all experiments in this section were performed in 63 mM sodium phosphate buffer. Each fresh cantilever tip was calibrated by the thermal noise method (as detailed in section 2.10.2) and the spring constants were consistently between 31 and 34 pN/nm. As all the Csp constructs are composed of the same domains with the same location of the C-terminal cysteine residues, there should be no issues over differences in pulling geometry between the constructs.

The velocity at which proteins are extended affects the observed pulling force. When a molecule is pulled slower there is more potential for thermally activated unfolding but less opportunity for this with higher pulling speeds. This means that a higher pulling speed produces greater unfolding forces. The dependence of unfolding forces on pulling velocity highlights how mechanical stability is a kinetic rather than thermodynamic phenomenon (259). To enable the extraction of energy level parameters the unfolding experiments were performed at different pulling speeds to generate a force-pulling speed dependence. Pulling speed ranges used in previous force extension studies fall in the range of 10 nm/s and 10,000 nm/s (265). The previous studies of *TmCsp* and *BsCsp* used pulling speeds of 100 nm/s, 200 nm/s, 600 nm/s and 2000 nm/s for *TmCsp* and *BsCsp*. These pulling speeds were chosen to make the natural log of the pulling speeds equidistant.

5.3.1 Identification of a mechanical signature for PB6 Csp

As a polyprotein chain is stretched, domains unfold in order of their mechanical resistance and as Csp domains are mechanically weaker they are expected to unfold before I27 domains as is seen with polyproteins containing I27 and *TmCsp* (136)(242). With the polyproteins studied the characteristic sawtooth pattern was produced (see figure 5.21).

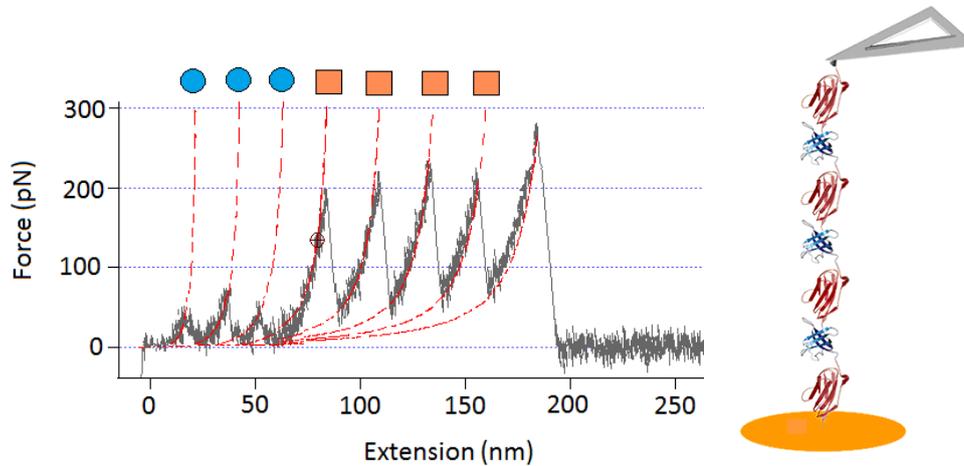


Figure 5.21: Mechanical unfolding of $(I27-PB6\ Csp)_3-I27$. Left: A schematic of an $(I27-PB6\ Csp)_3-I27$ unfolding trace with the peaks corresponding to PB6 Csp indicated by blue circles and I27 unfolding events indicated by orange squares. Worm-like chain fits to the trace are shown in red. Right: Schematic of the $(I27-PB6\ Csp)_3-I27$ showing stretching of the polyprotein between the gold surface and cantilever tip. Adapted from (137)

In many of the polyprotein traces clear peaks which were distinct from background noise were observed at lower extensions than the I27 domains (see figure 5.22). This suggested that it was possible to obtain a mechanical signature for PB6 Csp and it appears to unfold thermally by a two-state process agreeing with the earlier comparison of CD and fluorescence data (section 3.4.4). The location at which the polyprotein chain is picked up shows significant variation so in traces where polyprotein unfolding is observed anywhere between 0 and 3 Csp unfolding events can be seen. Smaller numbers of unfolding events are more commonly seen and it is rare that unfolding of the full construct is observed.

Preliminary experiments showed that PB6 Csp was mechanically weaker than *BsCsp* and *TmCsp*. Below 25 pN it becomes very difficult to identify peaks from non-specific interactions so a modified set of pulling speeds was used to keep the unfolding forces above thermal noise. The speeds chosen were still selected to be equally spaced in terms of the natural log of pulling speeds. For PB6 Csp the pulling speeds used were 330 nm/s, 600 nm/s, 1100 nm/s and 2000 nm/s. A selection of representative traces obtained at 14 °C at different pulling velocities from 330 nm/s to 2000 nm/s is presented in Figure 5.22. The magnitude of the unfolding forces and the levels of base thermal noise increases with higher pulling velocities. Two clearly distinct peak heights can be observed with small peaks corresponding to PB6 Csp and larger peaks I27. The final peak in each trace represents detachment of the polyprotein from the AFM tip and is not included in peak analysis.

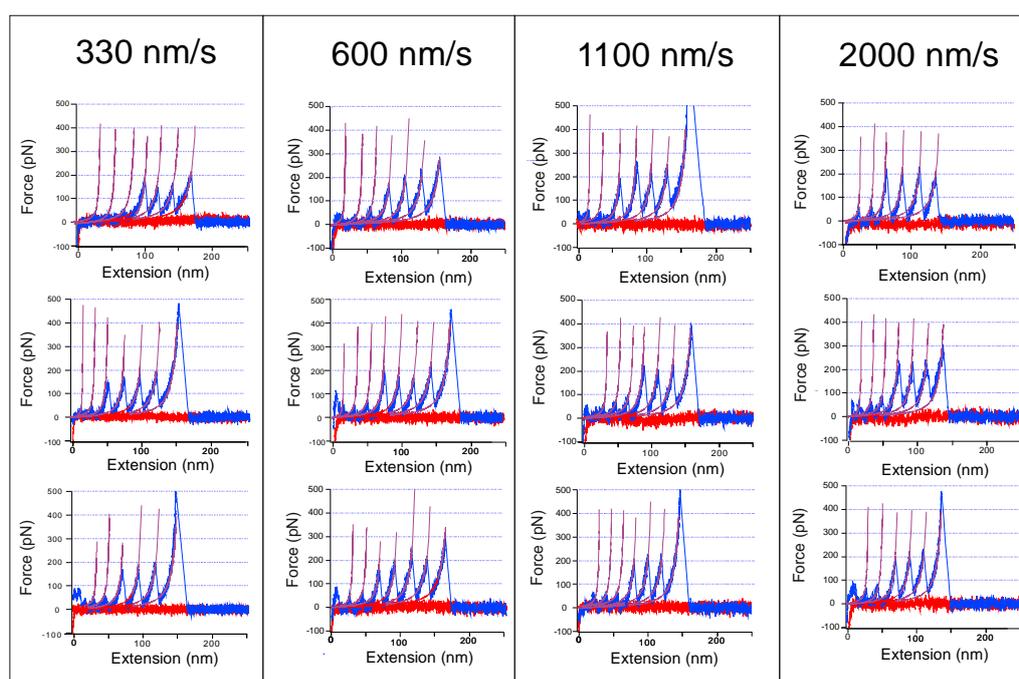


Figure 5.22: Representative AFM unfolding traces for $(I27-PB6\ Csp)_3-I27$ performed at 14 °C in 63 mM sodium phosphate buffer at a series of pulling speeds and fitted with the worm-like chain model in purple. Each trace was obtained from a separate experiment. The red trace represents approach of the AFM tip and the blue trace the retraction of the tip.

To discover more about the size of the mechanical domains within the sawtooth pattern the peaks are fitted to the worm like chain (WLC) model. Through fitting of the WLC to observed peaks, the change in contour length ΔL_c between folded and unfolded protein domains was consistently found to be 28 nm for I27 and 23.5 nm for PB6 Csp. These values applied for all of the Csp homologues and agree with previously published data (136)(190)(307). An approximate distance between N and C termini in the folded state is known for I27 and an estimate for the unfolded protein length can be made by multiplying the average length of an amino acid by the number of amino acids in the protein domain. The difference between these two figures gives a predicted value of ΔL_c of 29.6 nm for I27 (136)(190). The WLC fits aided in distinguishing low-force peaks from non-specific events and the identification of multiple polyprotein pickup events. Fitting of the WLC model to the peaks consistently gave a persistence value of 0.35 nm - 0.4 nm.

5.3.2 Filtering of AFM traces

Best et al. previously highlighted the need for stringent selection criteria when analysing AFM force extension traces (246). The traces are filtered to include only those containing at least 1 clearly distinguishable Csp unfolding event, 2 I27 unfolding events and a detachment peak. It is possible to distinguish where multiple constructs have been picked up by the presence of more unfolding peaks than the number of domains present within one polyprotein or abnormally large unfolding forces. Such traces were discounted from the analysis. Unfolding traces showing large peaks prior to the first Csp event, significant peak splitting or curvature in the baseline after detachment were also discounted. A maximum limit of 300pN for unfolding events was set as a selection criteria although final detachment was allowed to exceed this. Finally peaks which showed significant deviance from WLC fits were discounted. Typically a rate of around 0.2 - 0.4% of the traces obtained for the unfolding of Csp and I27 heteropolyprotein constructs fit within all the selection criteria for analysis.

5.3.3 Obtaining and analysing (I27-PB6 Csp)₃-I27 mechanical unfolding data at 14 °C

The unfolding forces observed in different proteins show a degree of variation and so to be able to obtain reliable estimates of the median unfolding force (F_U) of the Csp, a large number of unfolding events must be obtained. Before attempting this, initial trials were carried out to optimise the concentration of polyprotein applied to the surface. If this is too low, a reduced pick-up rate will be obtained and if too high this can lead to aggregation or a coating of protein across the gold surface which may produce an increase in non-specific events (294). In line with previous protocols, 30 minutes was allowed for the C-terminal cysteine residues to bond to the gold surface by thiol linkages (266).

Performing many force-extension experiments on a polyprotein under the same conditions generates a distribution of unfolding forces for each constituent domain. Using at least 15 Csp events means histograms can be constructed and fitted with Gaussian distributions allowing the most probable unfolding force to be seen for each construct and the standard distribution of unfolding forces to be assessed. The mean and median averages typically align to within 1-2 pN. Outlying events with extremely high or low unfolding forces have greater effect at skewing the mean average so the median unfolding force is used throughout this study. To improve reliability 3 sets of data obtained from 3 separate experiments were obtained for each pulling speed, minimising differences from variation between the cantilevers and allowing outlying data sets to be identified. The variation between values of F_U for both I27 and Csp conformed to within 10 pN for each set of repeats performed at the same temperature and pulling velocity. Using a fresh cantilever for each data-set avoids too much build-up of protein on the cantilever which could interfere with fresh protein-cantilever interactions. In all of the histograms there was no obvious drop-off at the lower forces suggesting there was no skewing of data due to difficulty identifying smaller force unfolding events from thermal noise.

A further confirmation of unfolding event identity is the interpeak distance (P_2P). With all the polyproteins a P_2P distance of 19 nm was seen for the Csp's and 23 nm for the I27 domains regardless of temperature or the Csp used. Histograms were plotted to verify a normal distribution of unfolding forces and P_2P distances with a bin width of 10 pN selected for unfolding force graphs and 1 nm bin width for interpeak distance. Separate histograms were produced for each of the 3 experiments at each pulling speed. Each histogram shows two separate distributions for Csp and I27 which are fitted with a Gaussian distribution (figure 5.23). As pulling speed was increased the unfolding forces increased and also showed a broader distribution. A broader spread of P_2P distances was also observed though only a very small increase in P_2P appeared across all the Csp's (Table 5.2). The I27 unfolding force and P_2P median values agreed with those from previous studies at the same pulling force and temperature (136-137).

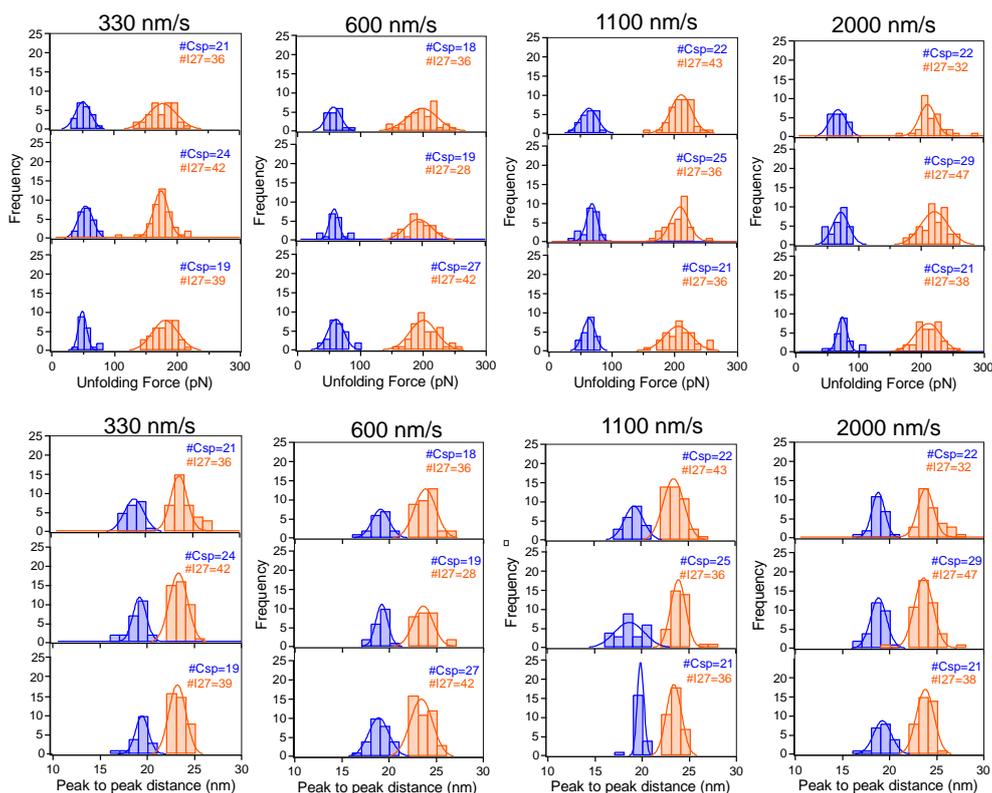


Figure 5.23: Histograms showing the normal distribution of F_U and interpeak distance for unfolding of $(I27-PB6\ Csp)_3-I27$ in 63 mM sodium phosphate buffer with PB6 Csp events in blue and I27 events in orange. The top half shows F_U data and bottom half interpeak distance for the same experimental runs. The three graphs for each pulling-speed correspond to the triplicate experimental repeats and the number of each unfolding events within each set is indicated at the top right.

Table 5.2: Unfolding forces and interpeak distances obtained at 14 °C for (I27-PB6 Csp)₃-I27

pulling speed	Expt Repeat	PB6 Csp peaks	Median force F_U PB6 Csp (pN)	Median p_2p PB6 Csp (nm)	I27 peaks	median force F_U I27 (pN)	Median p_2p I27 (nm)
330 nm/s	1	21	50.3 ± 9.6	18.5 ± 1.0	36	178.8 ± 19.9	23.8 ± 1.2
	2	24	55.2 ± 13.5	19.1 ± 1.1	42	174.3 ± 20.0	23.3 ± 0.7
	3	19	48.8 ± 9.1	19.3 ± 1.0	39	181.6 ± 17.2	23.3 ± 0.8
	Average		51.4 ± 3.3	18.9 ± 0.4		178.2 ± 3.7	23.5 ± 0.3
600 nm/s	1	18	57.0 ± 9.9	19.0 ± 1.1	36	195.5 ± 24.0	24.0 ± 1.0
	2	19	54.5 ± 12.8	19.1 ± 0.8	28	191.4 ± 20.0	23.5 ± 1.1
	3	27	60.6 ± 13.5	18.8 ± 1.0	45	199.3 ± 21.3	23.3 ± 0.9
	Average		57.4 ± 3.1	19.0 ± 0.2		195.4 ± 4.0	23.6 ± 0.4
1100 nm/s	1	22	63.2 ± 12.8	19.4 ± 1.0	43	211.9 ± 19.3	23.6 ± 1.0
	2	25	66.6 ± 11.4	18.8 ± 1.3	36	209.5 ± 17.9	23.9 ± 0.9
	3	21	64.3 ± 9.4	19.6 ± 0.7	36	206.4 ± 22.7	23.6 ± 0.7
	Average		64.7 ± 1.8	19.2 ± 0.4		209.3 ± 2.8	23.7 ± 0.2
2000 nm/s	1	22	67.0 ± 11.1	18.9 ± 0.8	32	215.0 ± 21.5	24.0 ± 1.2
	2	29	70.2 ± 13.2	18.6 ± 0.9	47	218.8 ± 20.3	23.5 ± 1.2
	3	21	73.8 ± 12.4	19.0 ± 1.1	39	210.1 ± 19.1	23.8 ± 0.8
	Average		70.3 ± 3.4	18.8 ± 0.4		214.7 ± 4.4	23.8 ± 0.2

5.4 Comparing the mechanical unfolding of the different Csp at 14 °C

Additional unfolding experiments were performed at the same temperature of 14 °C for the analogous polyproteins (I27-BsCsp)₃-I27 and I27-TmCsp)₃-I27 which were expressed and purified by the same strategy. They showed similar levels of variation in P_2P and F_U with data again showing Gaussian distributions (see appendix section). Both show increased values of F_U for their respective Csp but similar values for the marker protein I27, as expected. As can be seen in the histograms for all 3 Csp, there is minimal overlap between the unfolding forces or interpeak distances between the Csp and I27. Plotting the pairs of P_2P and F_U values for each unfolding event as a scatterplot shows the I27 and Csp peaks can clearly be differentiated

(figure 5.24). Even the most stable Csp, *TmCsp* has F_U values still clearly separate from I27 forces.

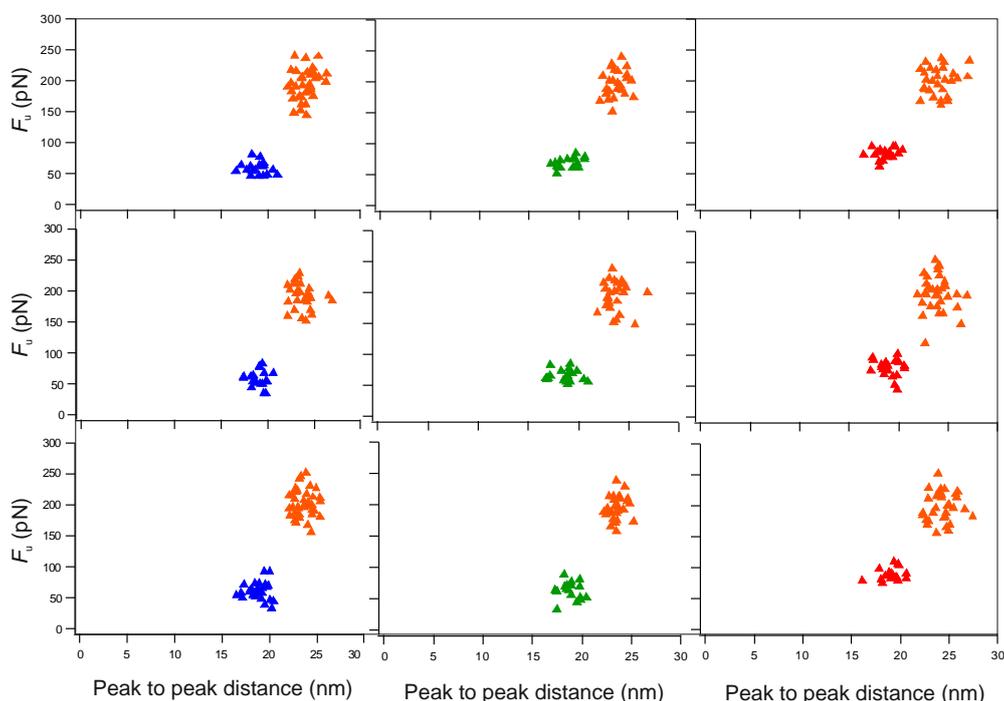


Figure 5.24: Scatterplots for $(I27-Csp)_3-I27$ unfolding in 63 mM sodium phosphate buffer at a pulling speed of 600 nm/s and temperature 14 °C showing the unfolding force and interpeak distance for I27 and the respective Csp. I27 is indicated in orange, PB6 Csp in blue (left), *BsCsp* in green (centre) and *TmCsp* (right) in red. A darker colour diamond shape indicates the median value for each data set. A separate scatterplot is shown for each of the 3 repeat experiments at 600 nm/s

The hierarchy of Csp unfolding forces F_U proceeds in the order $TmCsp > BsCsp > PB6$ Csp at all pulling velocities examined with a larger gap between *TmCsp* and *BsCsp* than the difference between *BsCsp* and PB6 Csp (example at 600 nm/s shown in figure 5.25). Therefore the measured mechanical hierarchy for the Csp is in the same order as the thermostability hierarchy in section 3.4.7.

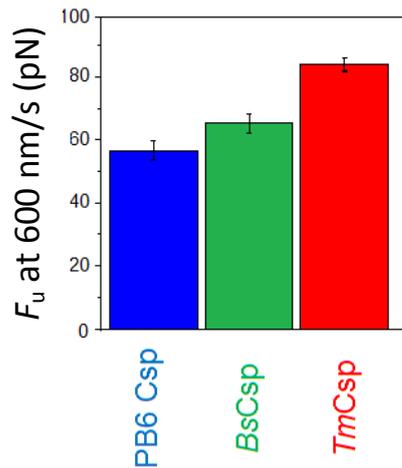


Figure 5.25: Median unfolding forces across triplicate data sets for the Csp within (I27-Csp)₃-I27 constructs performed in 63 mM sodium phosphate buffer at 600 nm/s pulling velocity at 14 °C.

5.5 Mechanical unfolding forces of Csp at different temperatures

To determine more about the temperature dependence of Csp mechanical stability, mechanical unfolding experiments were performed at different temperatures. Examples of temperature studies have shown that with I27 (295), the filamin protein ddFLN4 (296) and *TmCsp* a temperature softening effect occurs as temperature increases. This shifts the position of the transition state of mechanical unfolding further away from the native state resulting in a larger value of Δx_u (296). The amount that different proteins show mechanical softening seems substantially different with filamin protein ddFLN4 showing a much bigger temperature softening effect than I27. This may help stabilize proteins at high temperatures as a larger reaction co-ordinate distance to the transition state broadens the energy well minima. No significant increase in mechanical unfolding rate constants is seen for these two proteins with temperature and in fact a decrease is observed for I27 (295). It is hypothesised that the mechanical softening is due to a gradual decreased role of electrostatic contributions to mechanical stability and an increase in hydrophobic contributions (296). For a comparison of the F_U values at 23 °C for a variety of proteins see a review by *Hoffman et al.* (265).

Similar experiments to those performed at 14 °C were performed at 5 °C , 23 °C and 32 °C. The normal distributions of unfolding force F_U and P_2P were preserved across all temperatures examined for both Csp and I27. Comparison of F_U values of the marker I27 domains in the different polyprotein constructs show that they maintain very similar median unfolding forces (within 5 pN). This is observed at all pulling speeds and temperatures within the three different constructs. The mechanical stabilities of the 3 different Csp show a consistent pattern with $TmCsp$ the most mechanically stable and PB6 Csp the least over all temperatures and pulling speeds. As expected, values of F_U however show substantial decrease between 5 °C and 32 °C.

In a previous report there was a clear decrease in the spread of F_U at high temperatures for ddFLN4 though only a minor decrease was observed in this study for I27 and Csp (296). It may be that the simpler energy landscape of I27 and Csp decreases this effect while ddFLN4 has multiple unfolding intermediates whose proportions may be affected by temperature changes. The ΔL_c values from the WLC fits appeared to show no dependence on temperature as seen in previous studies (295).

Previous unfolding studies have shown increasing F_U at higher pulling velocities across all surveyed proteins (265). Generally more mechanically stable proteins show a greater sensitivity to the pulling velocity applied as is observed with the steeper gradient of I27 as against PB6 Csp in figure 5.26. A plot of how the force pulling speed dependence varies for the components of the (I27-PB6 Csp)₃-I27 polyprotein is shown in figure 5.27. F_U and gradient of the force-pulling speed dependence both decrease as the temperature is increased.

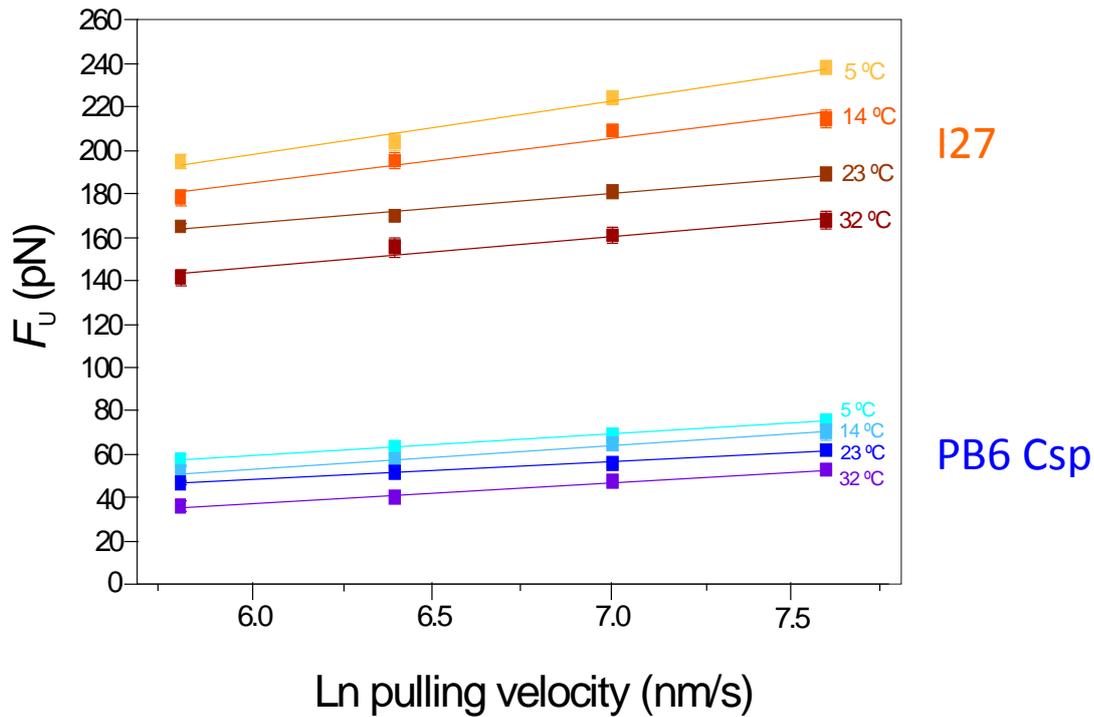


Figure 5.26: Comparison of the unfolding force pulling speed dependence of I27 and PB6 Csp in the (I27-PB6 Csp)₃-I27 polyprotein construct performed at a series of different temperatures. Orange tones are used for I27 and blue for PB6 Csp.

At each temperature the dependence of F_U on pulling speed exhibited similar gradients for all 3 Csp with no increase seen for *TmCsp* (Figure 5.26). I27 appears to show a stronger correlation between F_U and pulling speed but again the gradient becomes shallower with increased temperature. The gradients of *TmCsp*, *BsCsp* and PB6 Csp with pulling speed are similar at each temperature.

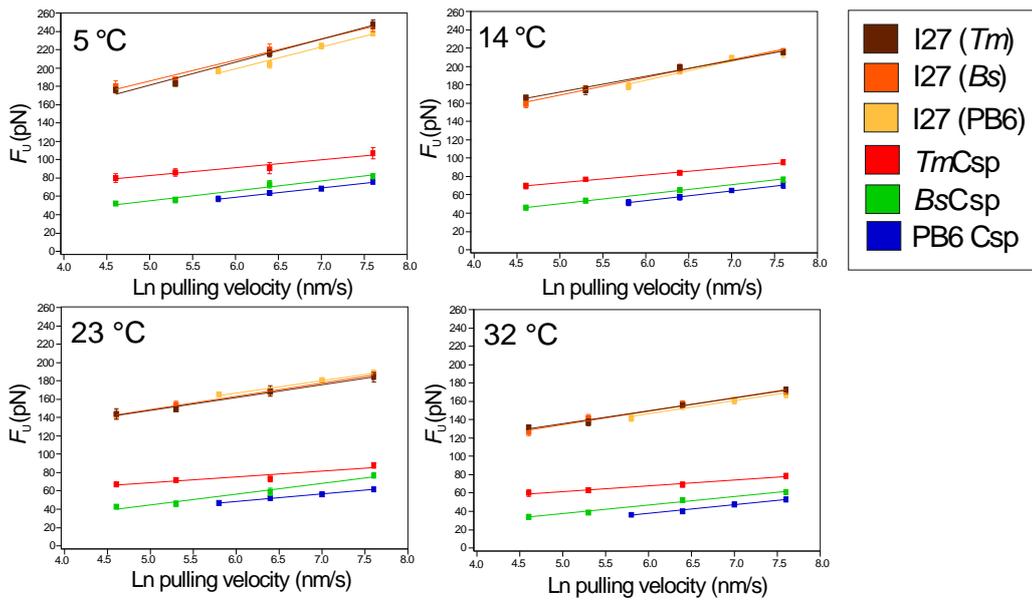


Figure 5.27: Comparison of the unfolding force pulling speed dependence of the I27 marker protein and Csp for all 3 constructs at a range of temperatures. *TmCsp* and *BsCsp* at 5 °C and 23 °C performed by K. M. Tych

When temperature is increased the unfolding forces decrease for all the Csp's. The change in unfolding force with temperature shown in figure 5.28 appears linear. The effect which makes the thermostability of proteins exhibit curvature is the hydrophobic effect. The lack of curvature in Force dependences may suggest that electrostatics rather than the hydrophobic effect is of greater importance to protein mechanical stability.

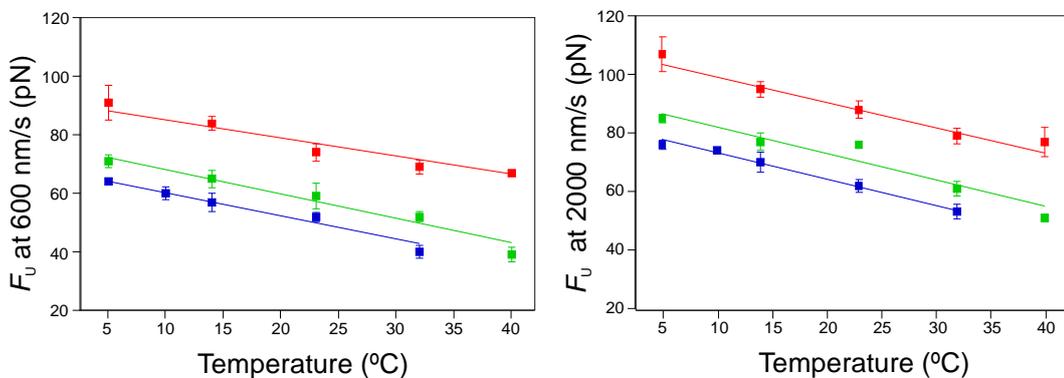


Figure 5.28: A comparison of the unfolding force temperature dependence of the Csp's at 600 nm/s left and 2000 nm/s. *TmCsp* is shown in red, *BsCsp* in green and PB6 Csp in Blue. Data at 600nm/s shown left and at 2000 nm/s shown right.

Due to temperature limitations of the AFM instrument and the barriers of thermal noise it was not possible to investigate higher temperatures to see if any deviation from linearity occurs as the T_m value for each protein is approached. To probe what happens at low temperature an additional data set was obtained at the point of maximal thermal stability for PB6 Csp which is 10 °C. No clear deviation was seen in the linearity of the unfolding force against temperature plot between 5 °C and 10 °C compared to the rest of the data set though this does not probe sub-zero temperatures where any effects would be more pronounced.

The F_U decreases at a similar rate for all 3 Csp when the temperature is increased (Figure 5.28). This suggests that all the Csp have similar responses in mechanical stability to changes in temperature which fits with their very similar 3 dimensional structures. When the temperature is raised from 5 °C the F_U for the 3 Csp all decrease by a similar proportion (Figure 5.29). This suggests that all the Csp show similar sensitivity to temperature and does not agree the hypothesis from the thermostability data (Figure 5.5) that PB6 Csp would be more sensitive to temperature. The value at 32 °C for PB6 Csp does appear lower than those for the other Csp but this may be occurring as the thermal mid-point T_m temperature for PB6 Csp is approached.

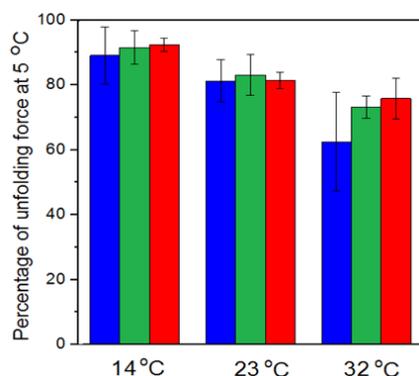


Figure 5.29: Expression of the median unfolding force of the Csp with a pulling speed of 600 nm/s at specified temperatures, as a proportion of their unfolding force at 5 °C

The estimated F_U at the optimum temperature of the host cell organisms are around 51 pN for PB6 Csp $T_{opt} = 22\text{ }^\circ\text{C}$, 45 pN for *BsCsp* $T_{opt} = 37\text{ }^\circ\text{C}$ and 42 pN $T_{opt} = 80\text{ }^\circ\text{C}$ for *TmCsp* assuming linearity was maintained beyond the temperature range of these studies. With regards to thermal stability a clear connection between thermostability of the protein and the optimum growth temperature of the host bacteria was seen in section 3.4.9. No such correlation was seen with regards to mechanical stability although this is not unsurprising as Csps have no direct mechanical function. The F_U of the Csps though are relatively similar at the respective optimal growth temperatures of the host bacteria.

5.6 Determination of mechanical unfolding energy barriers

5.6.1 Monte Carlo simulations

To determine further details of the energy barrier of mechanical unfolding the Zhurchov-Bell model can be applied to the force-pulling speed dependence (269)(297). This yields two interesting parameters of the unfolding energy landscape: the mechanical unfolding distance from the folded protein state to the transition state termed Δx_u and the rate constant of unfolding when zero force is applied, k_u (153)(246)(298). As the Δx_u and k_u parameters for each of the two different types of domains within a polyprotein are not independent of each other, the parameters were extracted using Monte Carlo (MC) simulations. These simulations assume that the unfolding of protein domains occur on a two-state, all or nothing basis. MC simulations utilise variation in defined parameters such as linker length, persistence, folded protein lengths, unfolded protein lengths and the Δx_u and k_u for each type of domain to generate simulated data that is used to match up to the experimentally determined data (see section 2.10.6) (136-137). The Monte Carlo simulation assumes a number of unfolding events between 3 and 7 for the polyproteins. The simulations initiate with all the domains being designated as folded and use comparison between defined parameters and randomly generated numbers to determine to the unfolding forces (for more detail see methods 2.10.6) (246)(259). The proportion of simulations containing each number of events may

not correspond fully with the number of events in experimental unfolding traces however MC simulations of (I27)₅ showed that the variations caused by this were small (271).

To improve the reliability of the MC output values in the study, a bootstrapping method was used to fit the steepest and shallowest gradient within the error of the force-pulling speed dependence as detailed in section 2.10.6. The MC simulations were conducted for each polyprotein construct at each temperature studied. This generated a series of pairs of Δx_u and k_u values.

5.6.2 Interpreting Δx_u and k_u values obtained from the Monte Carlo simulations

Δx_u is a measure of the degree to which the native state can deform without unfolding of the protein with a greater Δx_u implying a softer protein. It also represents the dependence of unfolding force on pulling velocity (291). Reported Δx_u values for the array of different proteins investigated are in the range 0.1 - 2 nm (300). The Δx_u values for the Csp homologues are relatively similar at 5 °C (0.39 nm for PB6 Csp to 0.47 nm for *TmCsp*) suggesting they have similar rigidity at this temperature. As the temperature is increased the position of the transition state of all the Csps moves more towards the unfolded state indicated by increasing Δx_u values (figure 5.30 and table 5.3). This shows increasing Csp softness allowing a greater level of deformation to occur before full Csp unfolding. The linear increase in Δx_u with temperature for all Csps implies a gradually increasing mechanical softness. The gradient of this increase is similar for PB6 and *BsCsp* but is steeper for *TmCsp*. This implies that the hyperthermophilic *TmCsp* becomes increasingly softer than its temperate and psychrotrophic homologues at high temperatures. *TmCsp* contains a greater number of salt bridges and these could allow *TmCsp* to stretch more at high temperatures without unfolding, providing increased mechanical stability. This may be an advantage to prevent unfolding at high temperatures where electrostatic interactions especially hydrogen bonds are relatively weaker.

The Δx_u values for I27 are lower than those for the Csp's (0.186 at 5 °C – 0.325 at 32 °C) suggesting a more rigid structure and were also in agreement with those values previously reported (295). Another β -barrel protein, GFP has been found to have a lower Δx_u value of 0.28 at 23 °C. The values of Δx_u for the Csp's are amongst the highest amongst all β -sheet proteins but below those reported for many α -helical proteins. A previous correlation identified that proteins with a lower mechanical stability (lower F_U values) generally have greater Δx_u values (265). This seems to be the opposite pattern to what is observed across the Csp homologues. A high Δx_u value could prevent mechanically weak proteins for unfolding under low forces.

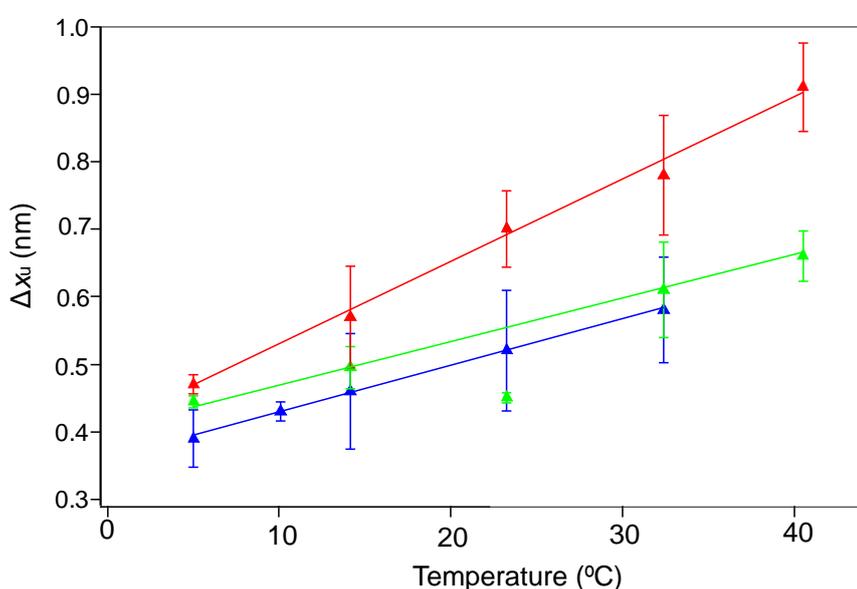


Figure 5.30: Comparison of the variation of Δx_u with temperature for the Csp's. *TmCsp* in red, *BsCsp* in green and PB6 Csp in blue. Error bars were determined from the bootstrapping method.

The value of the unfolding rate constant k_u remains very similar at different temperatures for each Csp. The values of k_u reflect the relative thermostabilities of the Csp with PB6 Csp having the highest unfolding rate constant and *TmCsp* having a significantly lower unfolding rate constant than *BsCsp* (figure 5.31 and table 5.3). There appears to be a small decrease in the k_u of *TmCsp* over the range and that was similar to a pattern previously reported for I27 (295).

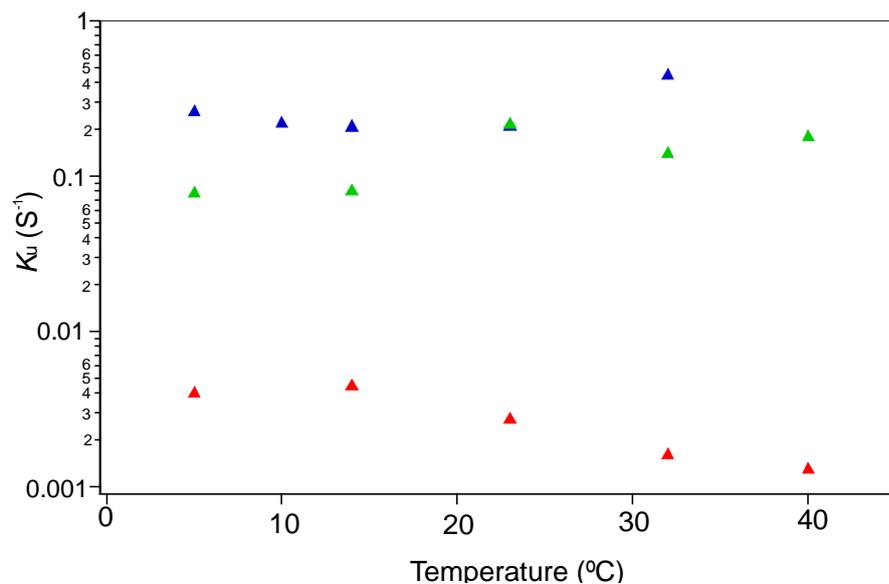


Figure 5.31: variation of k_u with temperature for the Csp with PB6 Csp plotted in blue, BsCsp plotted in green and TmCsp red

Table 5.3: Mechanical energy barrier properties derived from Monte Carlo simulations and the Arrhenius equation

Temp	PB6 Δx_u (nm)	PB6 k_u (S^{-1})	PB6 ΔG^{F-TS} (kJ/mol)	I27 Δx_u (nm)	I27 k_u (S^{-1})	I27 ΔG^{F-TS} (kJ/mol)
5	0.39 ± 0.04	0.256 ± 0.128	35.2 ± 1.6	0.19 ± 0.09	0.0053 ± 0.0010	44.1 ± 2.1
10	0.43 ± 0.06	0.218 ± 0.072	36.3 ± 1.0	0.19 ± 0.02	0.0055 ± 0.0035	45.0 ± 2.7
14	0.46 ± 0.09	0.205 ± 0.172	37.5 ± 0.5	0.24 ± 0.02	0.0022 ± 0.0006	47.8 ± 2.5
23	0.52 ± 0.09	0.208 ± 0.068	38.1 ± 1.0	0.31 ± 0.03	0.0029 ± 0.0006	48.7 ± 2.7
32	0.58 ± 0.08	0.440 ± 0.305	37.2 ± 2.3	0.33 ± 0.11	0.0053 ± 0.0040	48.5 ± 3.4

	BsCsp Δx_u (nm)	BsCsp k_u (S^{-1})	BsCsp ΔG^{F-TS} (kJ/mol)	I27 Δx_u (nm)	I27 k_u (S^{-1})	I27 ΔG^{F-TS} (kJ/mol)
14	0.49 ± 0.04	0.079 ± 0.025	39.4 ± 1.1	0.22 ± 0.01	0.0035 ± 0.0016	46.9 ± 3.1
32	0.61 ± 0.07	0.139 ± 0.002	40.2 ± 3.4	0.33 ± 0.04	0.0021 ± 0.0012	50.9 ± 3.9

	<i>TmCsp</i> Δx_u (nm)	<i>TmCsp</i> k_u (S^{-1})	<i>TmCsp</i> ΔG^{F-TS} (kJ/mol)	I27 Δx_u (nm)	I27 k_u (S^{-1})	I27 ΔG^{F-TS} (kJ/mol)
14	0.57 ± 0.11	0.0034 ± 0.0021	47.0 ± 3.3	0.25 ± 0.01	0.0011 ± 0.0002	49.7 ± 3.6
32	0.78 ± 0.09	0.0014 ± 0.009	52.0 ± 2.9	0.33 ± 0.02	0.0014 ± 0.0007	51.8 ± 2.1

5.6.3 Obtaining unfolding energy landscapes

The Arrhenius equation can be used to convert the unfolding rate constant to an estimate of the height of the energy barrier of unfolding ΔG^{F-TS} . While different values have been quoted for the pre-exponential factor, 10^6 is commonly used in the literature so this was used to allow easier comparisons (271)(301). The energy barriers for the 3 different Csp and I27 have similar heights at different temperatures within experimental error (figure 5.32 and 5.33).

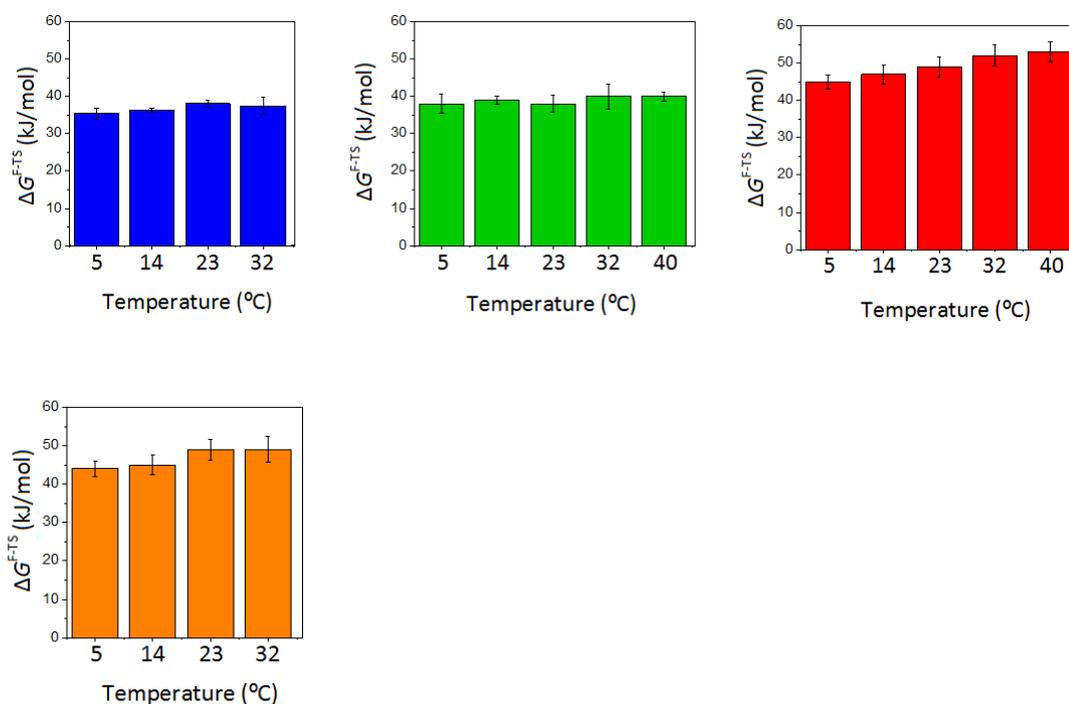


Figure 5.32: Variation in mechanical energy barrier heights with temperature. *TmCsp* is shown in red, *BsCsp* in green, *PB6 Csp* in blue and I27 taken from $(I27-PB6 Csp)_3-I27$ construct shown in orange.

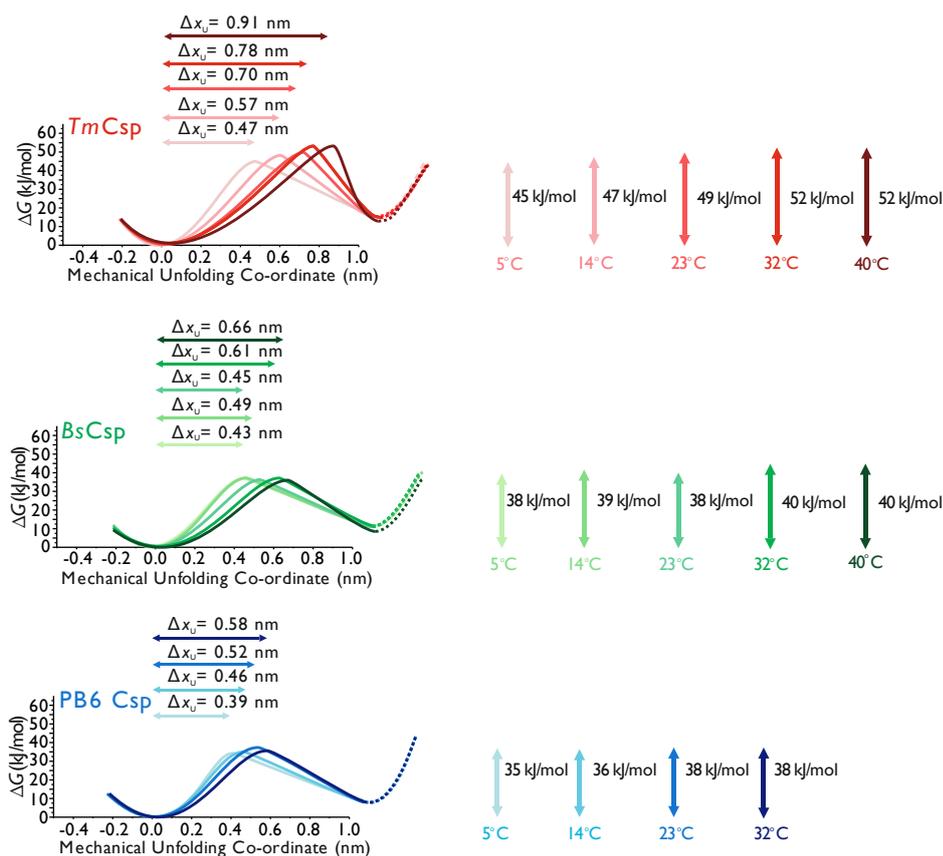


Figure 5.33: Comparison of the unfolding energy landscapes for the Csp's with *TmCsp* in red, *BsCsp* in green and PB6 Csp in blue at a series of different temperatures

At each individual temperature the unfolding barrier for *BsCsp* is slightly higher than that of PB6 Csp with *TmCsp* displaying a significantly greater energy barrier height (Figure 5.35). The energy barrier heights positively correlate with differences in Δx_u with *TmCsp* showing both greatest barrier height and largest Δx_u value. The I27 energy barrier heights are similar across all the polyprotein constructs, providing confidence in the differences observed in the Csp's. Figure 5.34 again highlights the increase in mechanical softness of *TmCsp* especially at the higher temperatures.

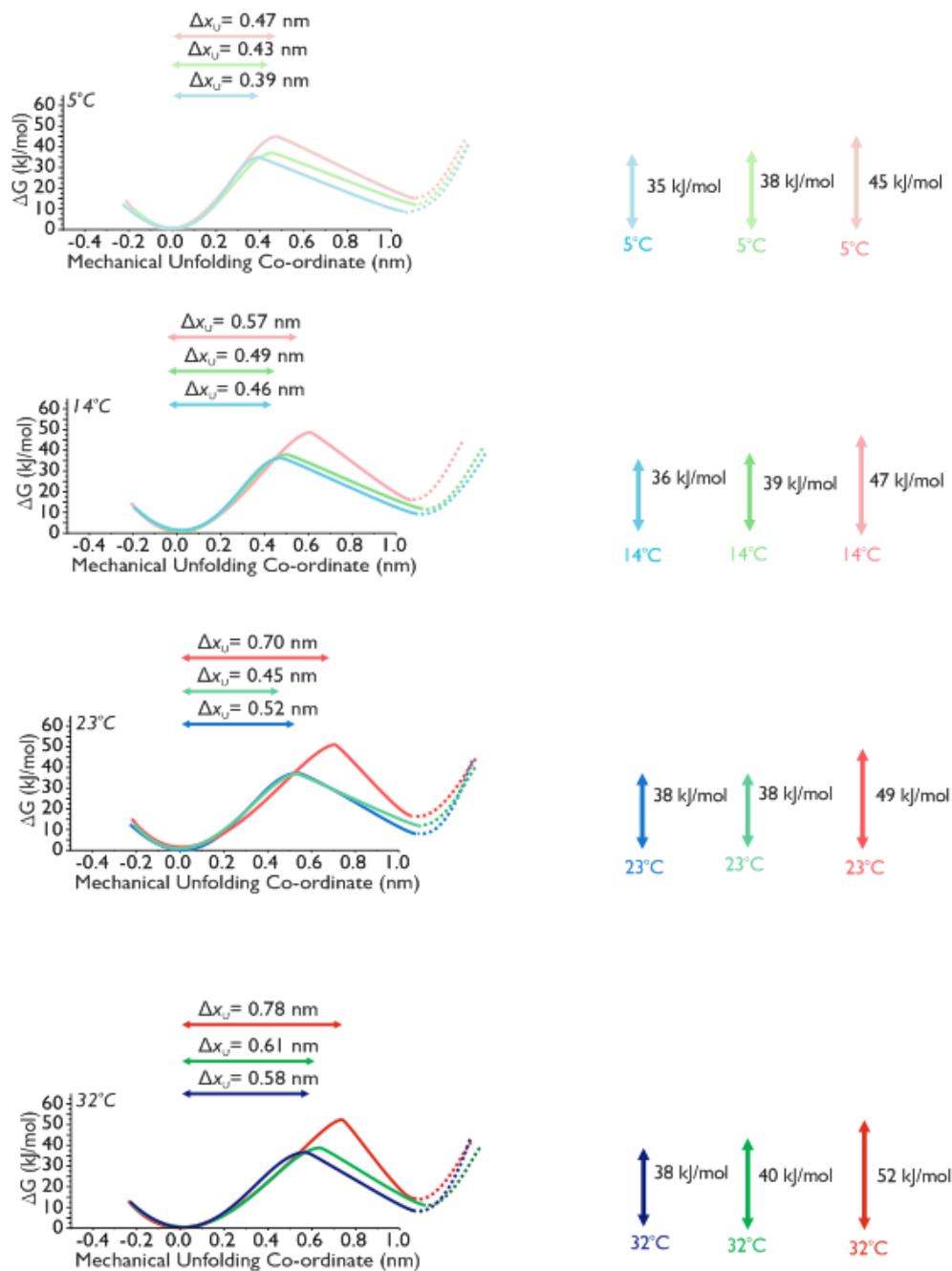


Figure 5.34: Comparison of the unfolding energy landscapes for the 3 different Csp at different temperatures with $TmCsp$ shown in red, $BsCsp$ shown in green and $PB6 Csp$ shown in blue

The differences in Δx_u become larger with temperature suggesting that the differences in softness between the Csp would become much more pronounced approaching the temperatures that $TmCsp$ would be found. If the distance between the Δx_u values continues to increase linearly with temperature the value of $TmCsp$

at the optimum temperature of *Thermotoga maritima* would be expected to be around 1.35 nm. If *BsCsp* were stable at 80 °C it would only have a Δx_u value of around 0.9 nm which is the same as *TmCsp* at 40 °C. If we compare the proteins at the optimum temperature of their host organisms 1.35 nm for *TmCsp* ($T_{opt} = 80$ °C) is far higher than 0.67 nm for *BsCsp* ($T_{opt} = 37$ °C) and 0.52 nm for PB6 Csp ($T_{opt} = 22$ °C). The proposed energy diagrams for each of the Csp at the optimal growth temperature of their respective bacteria are shown in figure 5.35. While Csp are expressed primarily below these temperatures this still provides an estimate of the mechanical energy barriers occurring in nature for each Csp.

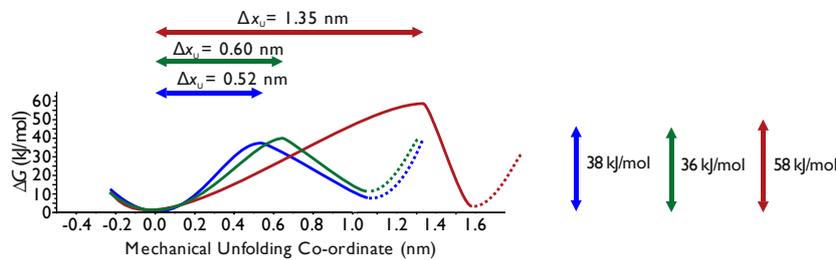


Figure 5.35: Comparison of the unfolding energy landscape of the Csp at the respective optimal growth temperatures of the organisms they derive from with *TmCsp* red $T_{opt} = 80$ °C, *BsCsp* green $T_{opt} = 37$ °C and PB6 Csp $T_{opt} = 22$ °C

5.7 Discussion

The hierarchy of Csp stabilities observed under chemical and thermal denaturation are mirrored upon mechanical perturbation. The psychrotrophic Csp PB6, as predicted, showed a reduced unfolding energy barrier ΔG^{F-TS} , relative to the mesophilic *BsCsp* and hyperthermophilic *TmCsp*. These figures correlate with differences in Δx_u and k_u values implying the stronger Csp were also softer. This trend was matched across different temperatures and pulling speeds. This is likely to be due to their highly conserved 3 dimensional structures. The sensitivity of the Csp unfolding forces to pulling velocity and temperature appeared similar for all three Csp. This suggests that there is no intrinsically different mechanical characteristic of extremophilic adaption. The linearity of the temperature dependence

of the csps suggests that electrostatics may be more important and the hydrophobic effect less important to mechanical stability than to thermostability.

Hyperthermophilic proteins are typically thought to be highly rigid to maintain stability at high temperatures so it seems surprising that *TmCsp* shows a greater degree of mechanical softness than its homologues. At 5 °C, the Δx_u values were relatively similar for the Csp's but an increase in *TmCsp* softness relative to its homologues becomes more pronounced with increasing temperature making the *TmCsp* more resistant to mechanical stresses (266). This highlights that rigidity is affected by temperature and that it is important to consider more specifically the temperature at which comparisons are being made. While projections suggest that *TmCsp* would be very mechanically soft at the optimal temperature for *Thermotoga maritima* of 80 °C ($\Delta x_u = 1.35$) it is difficult to know if the dynamic properties near the point of thermal unfolding may weaken the mechanical unfolding force of *TmCsp* below linear projections.

The mechanical adaptations of *TmCsp* may exist as a consequence of thermal adaptations rather than evolving due to selective pressure specifically for resistance to mechanical forces. Many of the features supporting thermal stability such as an increased network of ionic bonds and stronger hydrophobic interactions could also help to increase mechanical stability. While additional non-covalent interactions may allow *TmCsp* to deform significantly before unfolding, such deformation could unfold *BsCsp* and PB6 Csp as they have a reduced number of interactions. Hydrophobic interactions are proposed to become more significant to protein stability at higher temperatures (288). This would agree with a greater mechanical stability of *TmCsp* than *BsCsp* as *TmCsp* like other hyperthermophiles has a more optimised hydrophobic core (136-137). The importance of this core was highlighted by the large effect mutations in the region have on thermal stability (section 3.5.2).

The mechanical hierarchy of the Csp's in this study and the insights gained into their mechanical softness demonstrates the potential of studying extremophilic proteins to gain insights towards engineering proteins with specific mechanical properties. By starting rational design approaches with proteins that have higher or lower T_m values it may be possible to more easily tune how mechanically flexible biomaterials are and what degree of mechanical force they can resist. This may help in the future towards the aim of developing synthetic muscle-like proteins. Single point mutations performed on *TmCsp* and *BsCsp* in references (190-191) showed substantial changes in unfolding barrier height ΔG^{F-TS} but only very small changes in Δx_u . These parameters appeared to be coupled in the natural Csp homologues which have evolved interactions through evolution but not when individual mutations are made (190-191). This highlights that selective design of proteins may produce different effects to evolution as it does not optimise surrounding residues and the mechanical properties may be a product of a vast array of different interactions. This may mean that selecting extremophilic proteins as starting points for design studies could save time on initiating studies with mesophilic proteins.

6. Conclusions

This study examines the relative thermodynamic and mechanical stabilities of cold shock protein homologues derived from bacteria growing at vastly different temperatures. The Csp family of proteins exhibit many trends of temperature adaption, such as increased numbers of charged residues in hot-adapted variants and a glycine cluster within a psychrotrophic variant. Structural and functional residues are highly conserved across the family but the aromatic residues in the more hot-adapted forms contain a higher proportion of tyrosine and tryptophan residues which have a greater surface area potentially to intercalate more tightly with DNA. Csp sequences from a hyperthermophile, a mesophile, two psychrotrophs and a psychrophile were expressed in *E. coli* and purified.

Use of the Gibbs-Helmholtz equation and parameters obtained from thermal denaturation experiments allowed the temperature dependences of Csp thermostabilities to be determined. Projected values of the Gibbs free energy of unfolding (ΔG^{F-U}) agreed with those obtained from chemical equilibrium experiments. The cold-adapted Csps showed lower thermostabilities at all temperatures and are thermostable over a narrower range of temperatures. These differences were relatively mild compared to the dramatic differences between the thermostability curve of the hyperthermophilic *TmCsp* relative to the mesophilic *BsCsp*. The differences in the thermostabilities of the Csps are mediated by the different ΔH values of the homologues. Despite the large differences in T_m values for the Csps, the temperature of maximum thermostability was very similar across the hyperthermophilic, mesophilic and cold-adapted homologues.

The maximum thermostabilities of the Csps appear to reflect the optimum growth temperatures of their respective organisms except *PsiCsp* from the extreme psychrophile *Psychromonas ingrahamii* which showed slightly higher thermostability than the psychrotrophic PB6 Csp. Cold-adapted proteins still need to maintain a certain level of thermostability to avoid transient unfolding.

Traditional comparisons of protein thermostabilities at room temperature may hide key features of thermostability adaptations. When compared at 8 °C below the optimal growth temperatures of the bacteria, the ΔG^{F-U} values of the Csps were almost identical. This suggests that the very different ΔG^{F-U} values for Csps at room temperature are a consequence of optimisation of the Csps to a similar ΔG^{F-U} value at their respective operating temperatures. With so little available data on other protein families it is not clear how widely applicable this effect is.

An ideal hyperthermophilic protein is suggested to exhibit increased values of ΔH , T_s and reduced ΔC_p . What has been less well considered are the properties of extreme psychrophilic proteins. The opposite modifications to thermophilic adaptations (lower ΔH , lower T_s and increased ΔC_p) gives reduced thermostability but also a smaller range of temperatures over which the protein remains folded. Extreme low temperature thermostability may not be possible using such changes to these parameters due to cold denaturation (Figure 6, top left). An alternative strategy for cold adaption shown in a psychrophilic dihydrofolate reductase is a reduction in ΔC_p value which decreases the sensitivity of a protein to temperature, broadening the thermostability curve (Figure 6, top right). This type of strategy is used by some hyperthermophilic proteins to cope with high temperatures so it is surprising to see it also used in low temperature protein adaption. Psychrophilic protein have been typically considered to exhibit lower thermostabilities than mesophilic protein, For extreme psychrophiles maintenance of a higher level of tyhermostability may protect against the pressures of cold denaturation at extreme low temperatures (Figure 6, bottom left). Therefore the thermostability pofile of extreme psychrophilic proteins may resemble mesophilic proteins although there may also be additional optimisation of localised flexibility or reduced binding affinities to increase activity at low temperatures.

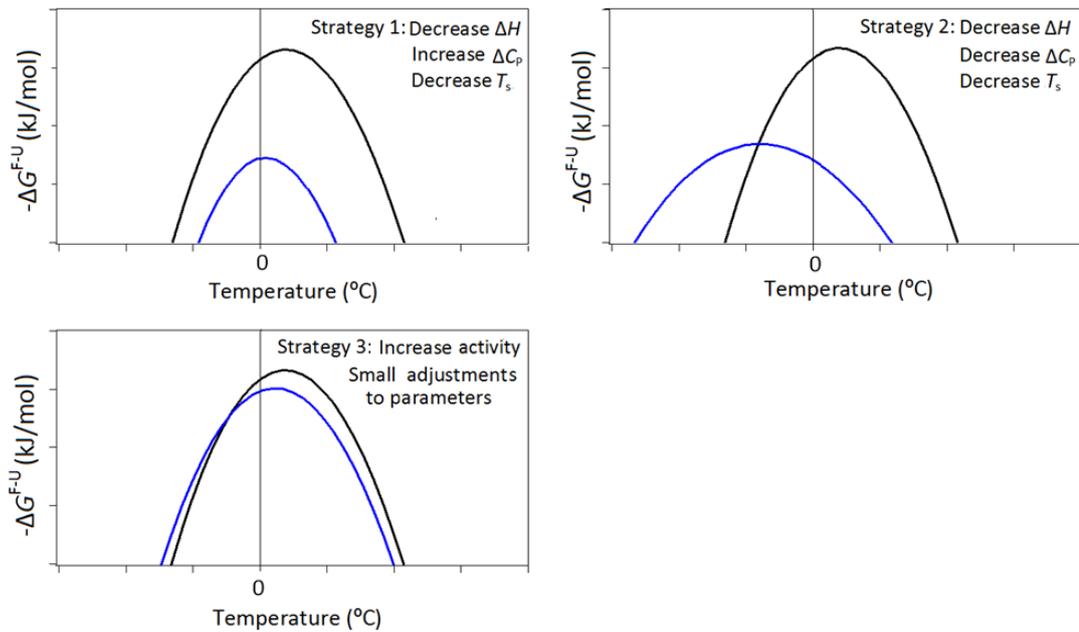


Figure 6: Methods of adaption of cold adapted proteins with proposed psychrophilic forms shown in blue relative to a temperate form in black

Mutational studies were performed to learn more about the thermostabilisation of the hyperthermophilic Csp from *Thermotoga maritima*. An attempt to substitute 3 salt-bridge forming residues from *TmCsp* into the mesophilic *BsCsp* yielded an increase in T_m but only a small increase in maximum stability though this may be due to unexpected electrostatic repulsion. This highlights the complexities of trying to increase protein thermostability as many thermostabilising effects rely not just on the interactions forming a salt-bridge but also a wider range of interactions with surrounding residues. Extremophilic proteins have already evolved many of these more subtle modifications. Decreasing the size of two side chains from *TmCsp* dramatically reduced the thermostability to levels close to the mesophilic Csp, highlighting how crucial hydrophobic packing is to thermostability. Only a small number of sequence combinations can generate hyperthermophilic levels of protein stability and this shows why further optimisation of thermostability is so difficult. A far larger number of mutations decrease protein stability and this may suggest why psychrophilic proteins could exhibit greater genetic diversity.

It was possible to obtain a mechanical signature for PB6 Csp by incorporating it into a polyprotein with I27, representing the first AFM mechanical characterisation of a cold-adapted protein. Using force-extension experiments comparisons of the mechanical stabilities of a hyperthermophilic, mesophilic and psychrotrophic Csp could be made. The similarity of I27 domain unfolding forces in each construct provided additional validation. The mechanical stabilities of the Csps showed the same hierarchy as the thermostabilities (PB6 Csp < *BsCsp* < *TmCsp*). The linear dependence of unfolding force on temperature suggests that the hydrophobic effect is not the crucial factor in mechanical stability so electrostatics may be more significant. The mechanical stability of the Csps shows similar sensitivity to changes in temperature suggesting that PB6 Csp does not have any special mechanical adaptations. An unusual finding was an increase in mechanical softness at high temperatures for *TmCsp* which may allow it to maintain mechanical stability.

These studies use a single molecule approach but applications for proteins in nanomaterials are more likely to use larger fibres, many protein molecules long and wide. This presents a further area of study as the properties of protein fibres and larger biomaterials may not match those of monomers. There are challenges regarding the linkage of protein chains and relative cross linking densities. There is still much work to do to apply extremophilic proteins to creating biomaterials with defined mechanical properties.

Stopped flow kinetic experiments have shown that despite vast differences in thermostability between the Csps, they exhibit similar rapid folding rate constants showing how the folding of the protein has been conserved since an early ancestor. The Csps however, exhibit different unfolding rate constants corresponding to their relative thermostabilities. These differences imply that the temperature adaptation is based around different levels of thermodynamic stabilisation. Using NMR, ^{15}N - ^1H HSQC spectra were obtained for the different Csps, however further assignments of the spectra would be required to make quantitative flexibility comparisons. It could

however be seen that a large conformational shift occurs in the cold adapted PB6 Csp upon DNA binding. Both the DNA bound and unbound PB6 Csp forms however showed little change in dynamics with temperature between 10 °C and 37 °C.

The lack of change in PB6 Csp dynamics with temperature in the HSQC spectra and the similarities of how the mechanical properties of PB6 Csp, *BsCsp* and *TmCsp* varied with temperature, suggest that PB6 Csp may not have evolved unique strategies. Most mutations arising by chance reduce thermostability so this may suggest that the cold-adapted Csps are simply a product of genetic drift occurring due to lower selection pressure for thermostability. There are common features in psychrophilic proteins such as changes in the proportions of certain amino acids, more glycine clusters, changes in loop lengths and lower substrate binding affinities. A lower selection pressure for thermostability may allow more freedom for the protein to evolve these strategies. At low temperatures the selection pressure for higher activity is increased driving the evolution of modifications.

It is unclear how much of the increase in activity reported in psychrophilic enzymes relative to mesophilic homologues is mediated through increases in flexibility and to what extent this relates to changes in either global or local dynamics. The modifications may be specific to the protein family involved. This project yielded a large amount of information on how Csps adapt to extreme temperatures, helping to shed more light on the properties of psychrophilic proteins. This knowledge may be useful towards the aim of modifying current proteins to exploit the higher activity of psychrophilic proteins at low temperatures.

Appendix

PB6 Csp sequence

MHHHHHHSSSDKVEGTVKWFNEAKGFGFIAQDNGGQDVFAHYSAIQGGGFKTLAEGQKVSFILGDGK
KGPQAEQIEA*

BsCsp sequence

MHHHHHHSSLEGKVKWFNSEKGFIEVEGQDDVVFHFSAIQGGGFKTLEEGQAVSFEIVEGNRGPQA
ANVTKEA*

PsiCsp sequence

MHHHHHHSSKVQGTVKWFNESHKGFIEQASGPDVFAHFSASSEGFKTLAEGQKVEFTITQGQKGP
AENIVAI*

SheCsp sequence

MHHHHHHSSNKTGLVKWFNEDKGFITPDNGGADVVFHFRSITSEGFKTLAEGQKVSFDVEQGQK
GPQAANVVAV*

TmCsp sequence

MHHHHHHSSRGKVKWFDSKKGYGFITKDEGGDVVFHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQA
AHVKVVE*

Figure S1: sequences for the monomer Csps

MHHHHHHSSLIEVEKPLYGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDIIEGKKHILILHN
SQLGMTGEVSFQAANAKSAANLKVKELLSVGATISDKVEGTVKWFNEAKGFGFIAQDNGGQDVFAHY
SAIQGGGFKTLAEGQKVSFILGDGKKGPPQAEQIEAITVIGLASLIEVEKPLYGVEVFGVGETAHFEIELSEPDV
HGQWKLKGQPLTASPDIIEGKKHILILHNSQLGMTGEVSFQAANAKSAANLKVKELALSQTIVSDKV
EGTVKWFNEAKGFGFIAQDNGGQDVFAHYSAIQGGGFKTLAEGQKVSFILGDGKKGPPQAEQIEAIVITG
SLALIEVEKPLYGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDIIEGKKHILILHNSQLGMTG
EVSFQAANAKSAANLKVKELLSALGIVTSDKVEGTVKWFNEAKGFGFIAQDNGGQDVFAHYSAIQGGG
FKTLAEGQKVSFILGDGKKGPPQAEQIEAITAGVSLLIEVEKPLYGVEVFGVGETAHFEIELSEPDVHGQWKLK
GQPLTASPDIIEGKKHILILHNSQLGMTGEVSFQAANAKSAANLKVKELCC*

Figure S2: Amino acid sequence of the (I27-PB6 Csp)₃-I27 polyprotein with PB6 in blue, I27 in orange, hexahistidine tag in purple and linkers in black

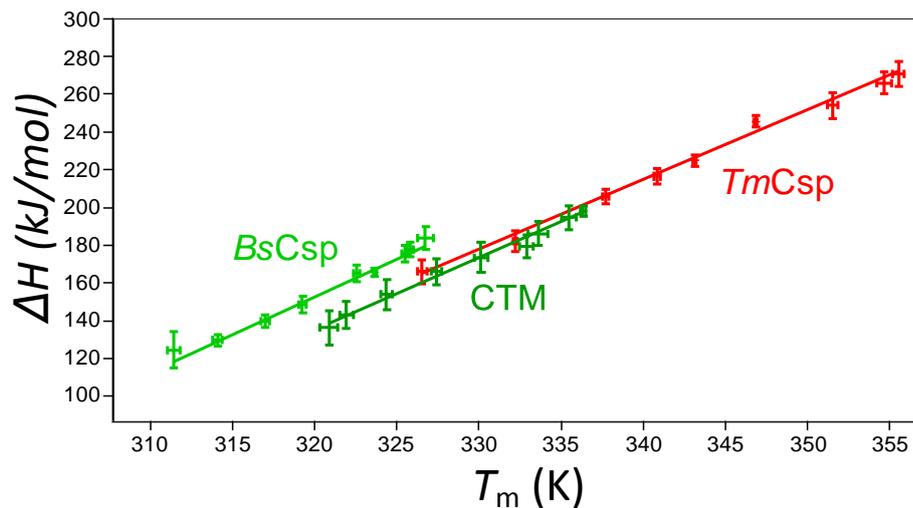


Figure S3: Plots of ΔH against T_m value pairs obtained for *BsCsp*, *TmCsp* and CTM with increasing concentrations of GdnHCl from right to left. Errors for each data point derive from uncertainties in the fitting of the denaturant curve. A weighted linear fit was applied to the data with the significance of each data point inversely proportional to the size of the uncertainty of data points in the series.

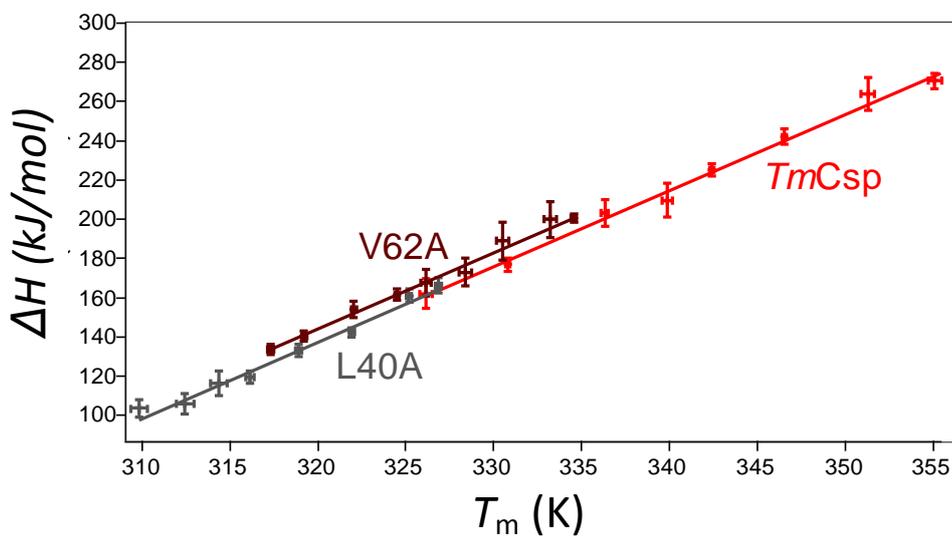
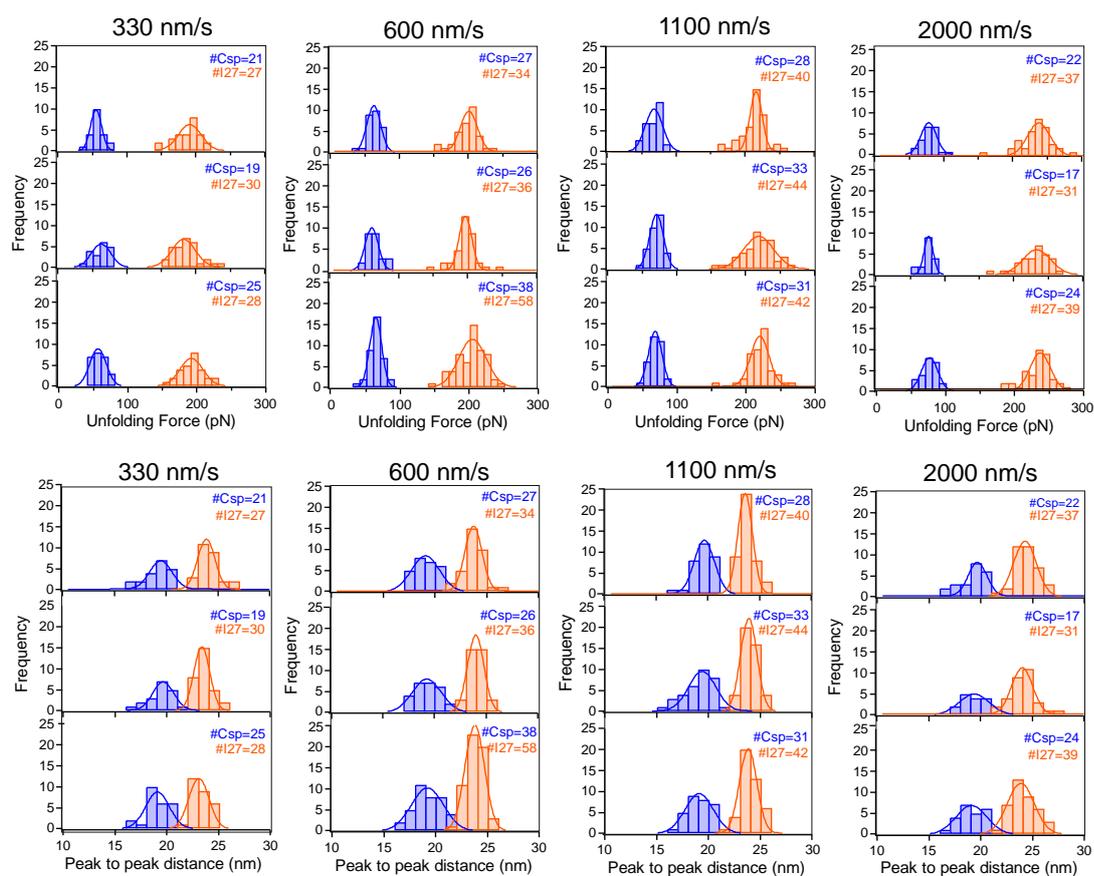


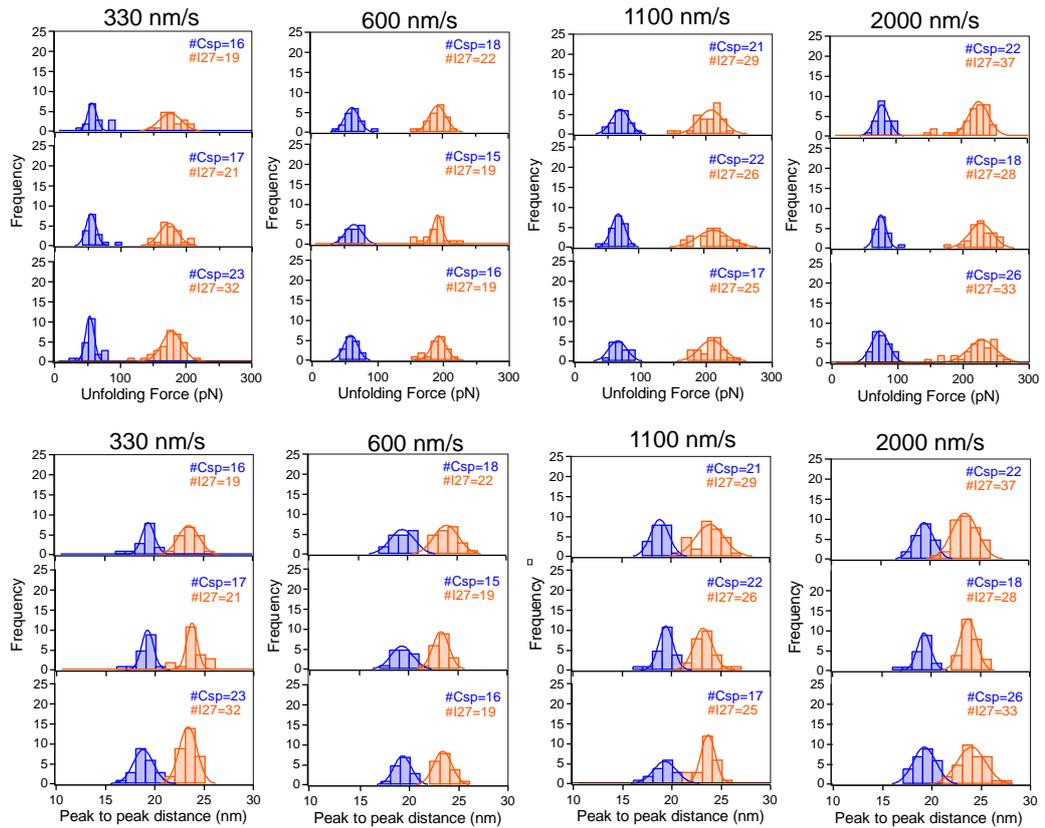
Figure S4: Plots of ΔH against T_m value pairs for *TmCsp*, V62 and L40A obtained with increasing concentrations of GdnHCl from right to left. Errors for each data point derive from uncertainties in the fitting of the denaturant curve. A weighted linear fit was applied to the data with the significance of each data point inversely proportional to the size of the uncertainty of data points in the series.

AFM Appendix – 5 °C PB6 Csp



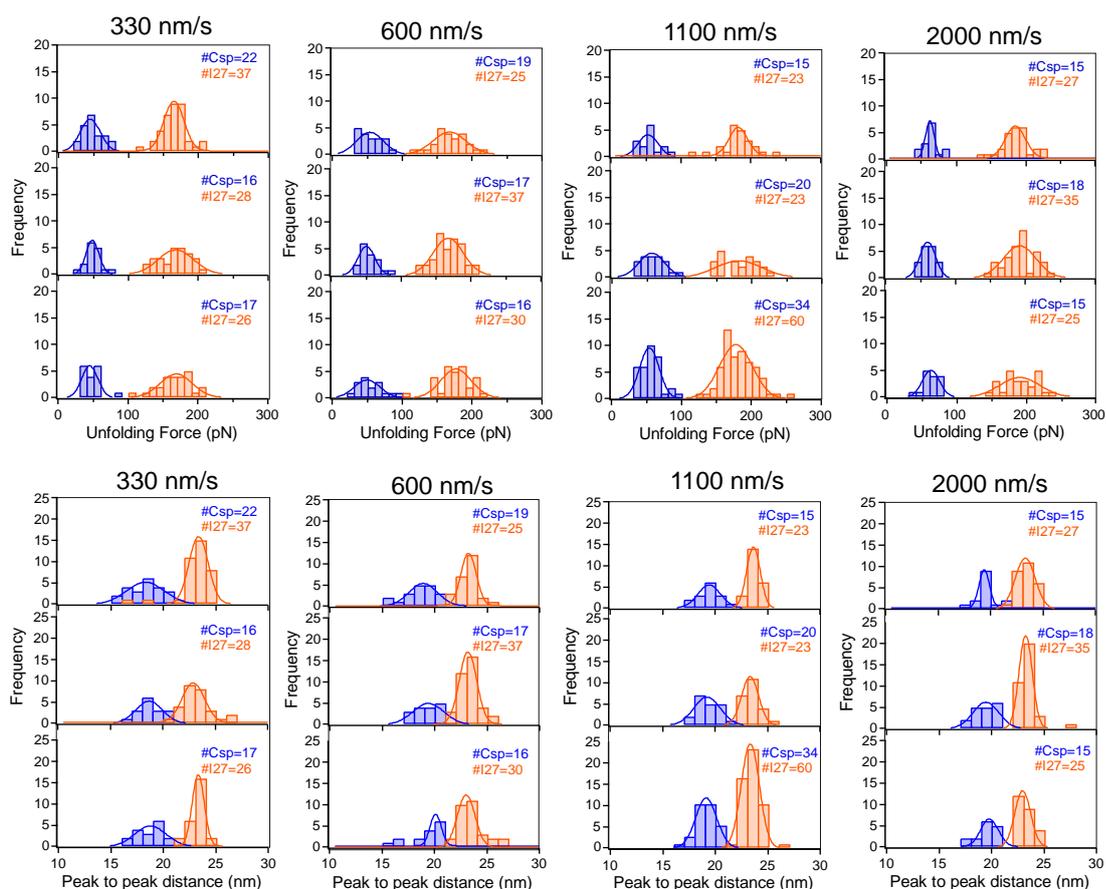
pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p _{2p} Csp (nm)	I27 peaks	median force I27 (pN)	Median p _{2p} I27 (nm)
330 nm/s	1	21	53.6 ± 9.7	19.3 ± 1.5	27	194.9 ± 19.0	24.0 ± 1.1
	2	19	61.5 ± 11.6	19.2 ± 1.2	30	189.9 ± 17.9	23.7 ± 1.0
	3	25	56.6 ± 11.1	18.9 ± 1.2	28	196.3 ± 18.3	23.2 ± 0.8
	Average		57.3 ± 3.0	19.1 ± 0.2		193.7 ± 3.4	23.6 ± 0.4
600 nm/s	1	27	63.8 ± 9.8	19.1 ± 1.2	34	204.6 ± 18.7	23.7 ± 1.0
	2	26	62.7 ± 11.6	19.5 ± 1.2	36	198.0 ± 18.0	24.0 ± 0.6
	3	38	64.4 ± 9.9	19.1 ± 1.5	58	205.6 ± 21.0	23.9 ± 0.8
	Average		63.6 ± 0.9	19.2 ± 0.5		202.7 ± 4.1	23.8 ± 0.1
1100 nm/s	1	28	68.7 ± 10.9	19.6 ± 1.0	40	226.0 ± 21.8	23.4 ± 0.8
	2	33	70.0 ± 9.4	19.2 ± 1.5	44	221.9 ± 24.1	23.8 ± 1.5
	3	31	66.5 ± 9.3	19.1 ± 1.5	42	225.6 ± 19.5	23.9 ± 1.5
	Average		68.6 ± 1.9	19.3 ± 0.3		224.53 ± 2.3	23.70 ± 0.2
2000 nm/s	1	21	77.7 ± 11.4	19.6 ± 1.5	33	238.6 ± 24.3	24.4 ± 1.1
	2	17	76.1 ± 9.7	19.7 ± 1.6	31	235.2 ± 21.6	24.0 ± 1.3
	3	24	74.5 ± 11.3	18.9 ± 1.5	39	239.3 ± 25.1	23.9 ± 1.2
	Average		76.1 ± 1.6	19.4 ± 0.5		237.7 ± 2.2	24.1 ± 0.2

10 °C PB6 Csp



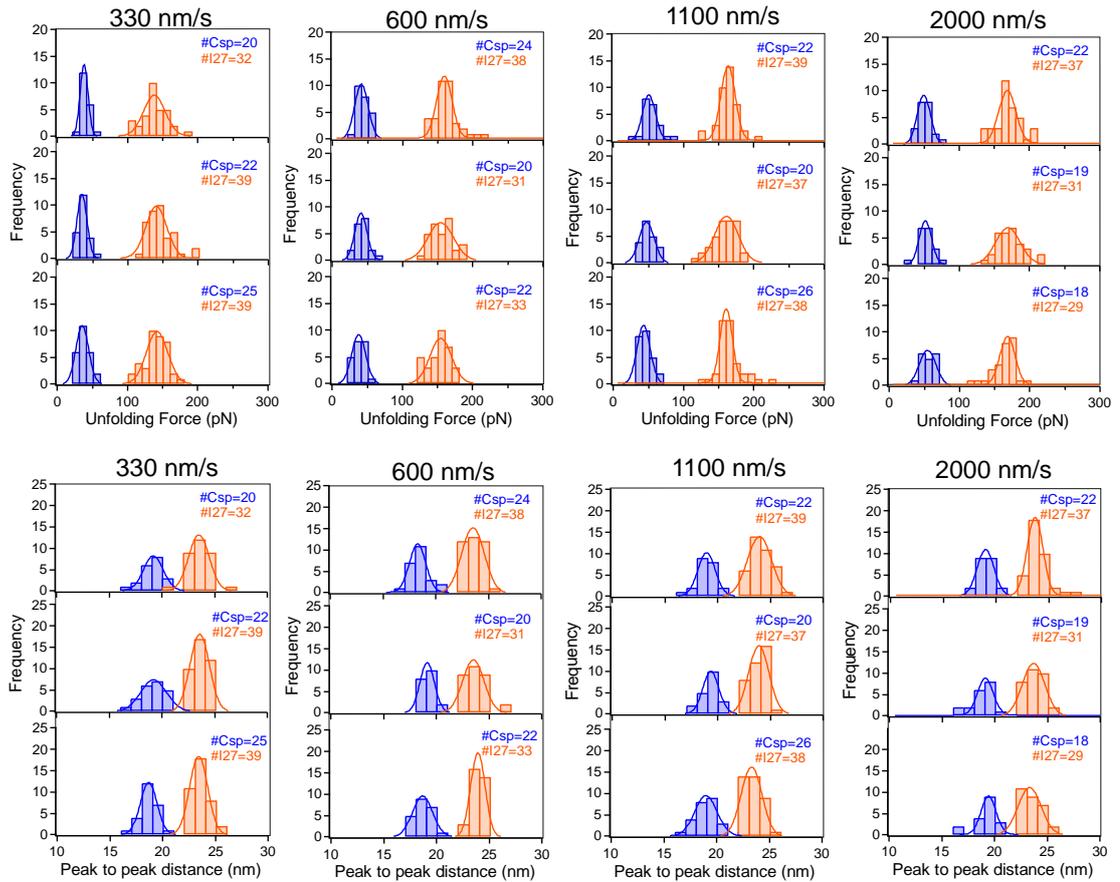
pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p _{2p} Csp (nm)	I27 peaks	median force I27 (pN)	Median p _{2p} I27 (nm)
330 nm/s	1	16	54.6 ± 11.9	19.3 ± 1.4	19	186.9 ± 16.5	23.3 ± 1.0
	2	17	57.0 ± 12.4	19.2 ± 1.2	21	180.6 ± 16.1	23.9 ± 1.1
	3	22	53.6 ± 12.3	18.5 ± 1.2	30	186.7 ± 15.6	23.3 ± 0.9
	Average		55.1 ± 1.7	19.0 ± 0.4		184.7 ± 3.6	23.5 ± 0.4
600 nm/s	1	18	60.6 ± 14.1	19.5 ± 1.0	22	194.0 ± 14.1	24.0 ± 1.3
	2	15	62.1 ± 11.1	19.1 ± 1.0	19	197.0 ± 16.6	23.2 ± 0.7
	3	16	57.8 ± 9.6	19.1 ± 0.9	19	198.7 ± 14.1	23.2 ± 0.9
	Average		60.2 ± 2.2	19.3 ± 0.2		196.6 ± 2.4	23.4 ± 0.5
1100 nm/s	1	21	69.5 ± 12.6	18.7 ± 0.9	29	211.4 ± 22.0	23.8 ± 1.4
	2	16	66.7 ± 11.8	19.2 ± 1.1	20	210.2 ± 24.1	23.2 ± 1.2
	3	17	62.1 ± 11.7	19.5 ± 1.1	25	208.4 ± 17.9	23.4 ± 1.0
	Average		66.1 ± 3.8	19.12 ± 0.4		210.0 ± 1.5	23.5 ± 0.3
2000 nm/s	1	22	76.0 ± 10.4	19.2 ± 1.2	37	224.9 ± 23.9	23.4 ± 1.2
	2	18	75.3 ± 8.9	19.2 ± 1.0	28	227.8 ± 19.3	23.9 ± 0.9
	3	26	73.7 ± 12.8	19.4 ± 1.1	33	225.5 ± 27.6	23.9 ± 1.4
	Average		75.0 ± 1.2	19.3 ± 0.1		226.0 ± 1.5	23.7 ± 0.3

23 °C PB6 Csp



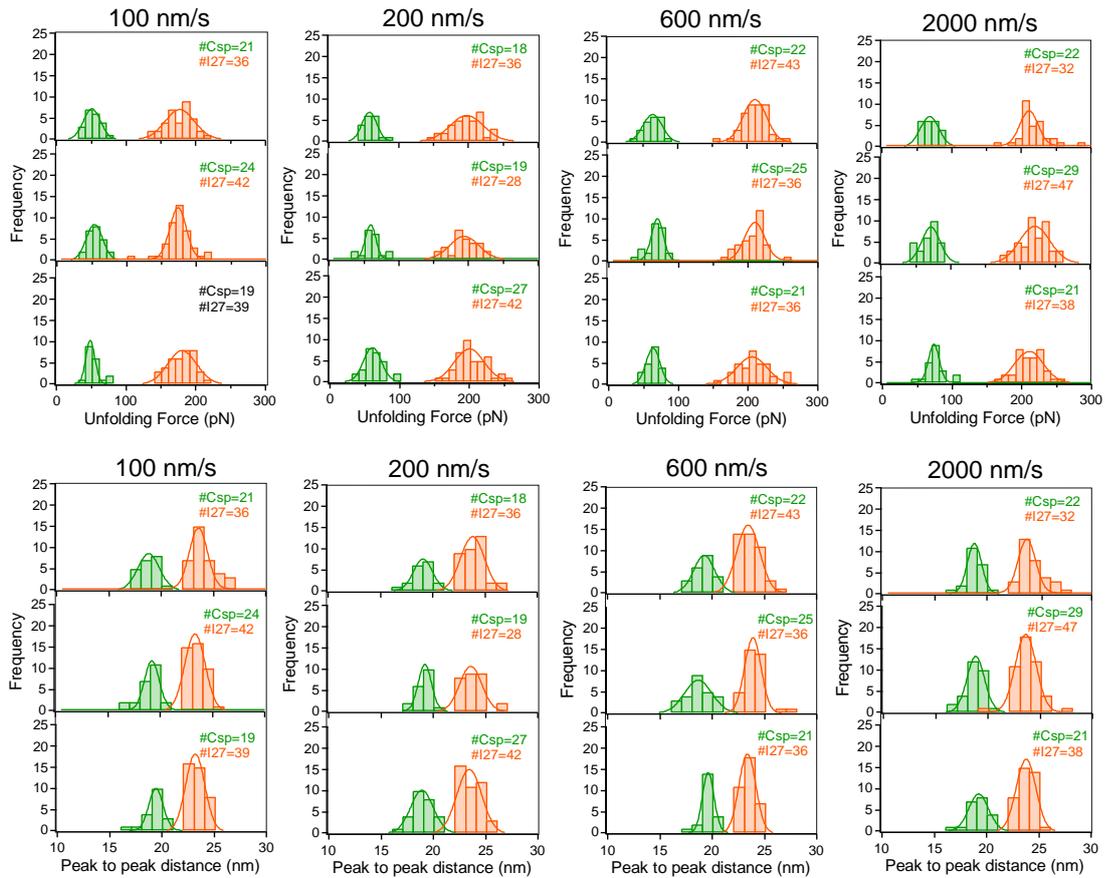
pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p ₂ p Csp (nm)	I27 peaks	median force I27 (pN)	Median p ₂ p I27 (nm)
330 nm/s	1	22	46.4 ± 13.7	18.4 ± 1.5	37	164.1 ± 17.6	23.3 ± 1.4
	2	16	48.7 ± 13.9	18.5 ± 1.2	46	164.7 ± 22.6	23.1 ± 1.3
	3	17	47.0 ± 12.3	18.9 ± 0.7	26	168.5 ± 22.6	23.6 ± 1.2
	Average		47.4 ± 1.2	18.6 ± 0.3		165.2 ± 1.4	23.2 ± 0.2
600 nm/s	1	19	53.7 ± 13.9	18.7 ± 2.0	25	167.8 ± 26.9	23.0 ± 0.9
	2	17	52.0 ± 13.5	19.2 ± 1.2	37	168.8 ± 19.9	23.1 ± 0.9
	3	16	50.7 ± 19.0	19.7 ± 1.7	30	173.0 ± 23.5	23.2 ± 1.2
	Average		52.1 ± 1.5	19.2 ± 0.5		169.9 ± 2.8	23.1 ± 0.1
1100 nm/s	1	15	53.7 ± 18.5	19.5 ± 1.2	23	181.0 ± 25.2	23.7 ± 0.6
	2	20	57.7 ± 16.3	19.1 ± 1.1	23	181.7 ± 24.5	23.5 ± 0.8
	3	34	56.1 ± 14.8	19.1 ± 1.0	60	179.0 ± 25.8	23.3 ± 0.8
	Average		55.8 ± 2.0	19.2 ± 0.2		180.6 ± 1.4	23.5 ± 0.2
2000 nm/s	1	15	64.4 ± 11.0	19.3 ± 1.0	27	187.4 ± 21.4	23.1 ± 0.8
	2	18	60.6 ± 8.5	19.5 ± 1.0	35	191.0 ± 22.1	23.3 ± 0.9
	3	15	60.4 ± 15.6	19.6 ± 0.8	23	189.0 ± 25.6	23.4 ± 0.7
	Average		61.8 ± 2.2	19.5 ± 0.2		189.1 ± 1.9	23.3 ± 0.2

32 °C PB6 Csp



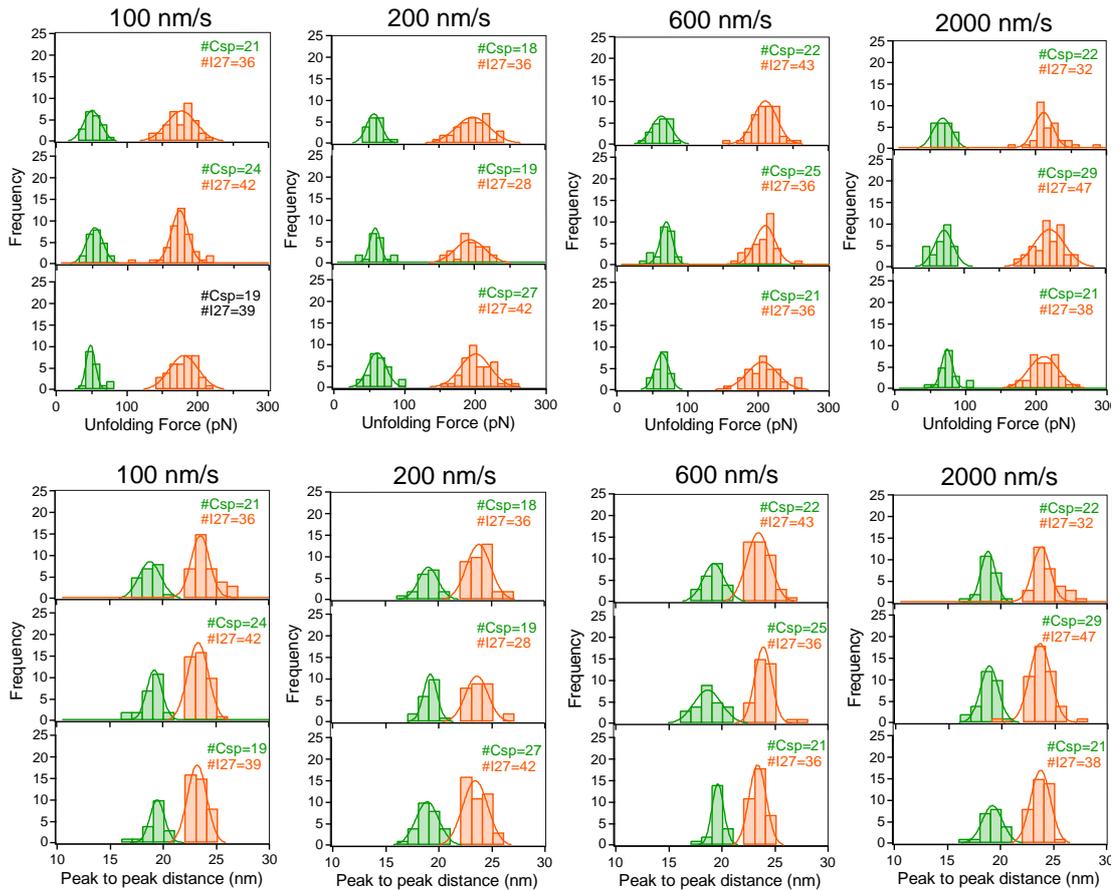
pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p _{2p} Csp (nm)	I27 peaks	median force I27 (pN)	Median p _{2p} I27 (nm)
330 nm/s	1	20	39.2 ± 8.1	19.4 ± 1.1	32	138.2 ± 18.1	23.7 ± 1.1
	2	22	34.1 ± 6.8	19.2 ± 1.0	39	145.0 ± 17.6	23.8 ± 0.8
	3	25	35.6 ± 8.0	18.2 ± 1.0	41	141.0 ± 17.4	23.5 ± 0.9
	Average		36.3 ± 2.7	18.9 ± 0.6		141.4 ± 3.4	23.7 ± 0.2
600 nm/s	1	24	42.3 ± 6.8	18.1 ± 0.9	38	160.5 ± 18.1	23.4 ± 0.9
	2	20	40.4 ± 9.2	19.1 ± 0.7	31	154.0 ± 15.2	23.6 ± 1.1
	3	22	37.9 ± 8.9	18.6 ± 0.8	35	151.9 ± 14.9	23.9 ± 0.7
	Average		40.2 ± 2.2	18.7 ± 0.3		155.5 ± 4.5	23.6 ± 0.3
1100 nm/s	1	22	50.3 ± 12.2	18.7 ± 0.9	39	162.5 ± 24.0	24.0 ± 1.1
	2	20	47.2 ± 9.0	19.4 ± 0.8	37	156.8 ± 16.6	23.9 ± 0.8
	3	26	45.5 ± 9.1	19.0 ± 1.4	38	163.6 ± 17.6	23.4 ± 0.9
	Average		47.7 ± 2.4	19.0 ± 0.3		161.0 ± 3.6	23.8 ± 0.4
2000 nm/s	1	22	50.5 ± 10.8	19.1 ± 0.8	37	167.4 ± 17.3	23.6 ± 1.1
	2	19	52.9 ± 9.7	18.8 ± 1.0	31	171.8 ± 18.4	23.8 ± 1.0
	3	18	55.6 ± 8.4	19.6 ± 1.1	29	164.2 ± 16.6	23.4 ± 1.0
	Average		53.0 ± 2.5	19.1 ± 0.3		167.8 ± 3.9	23.6 ± 0.2

14 °C BsCsp



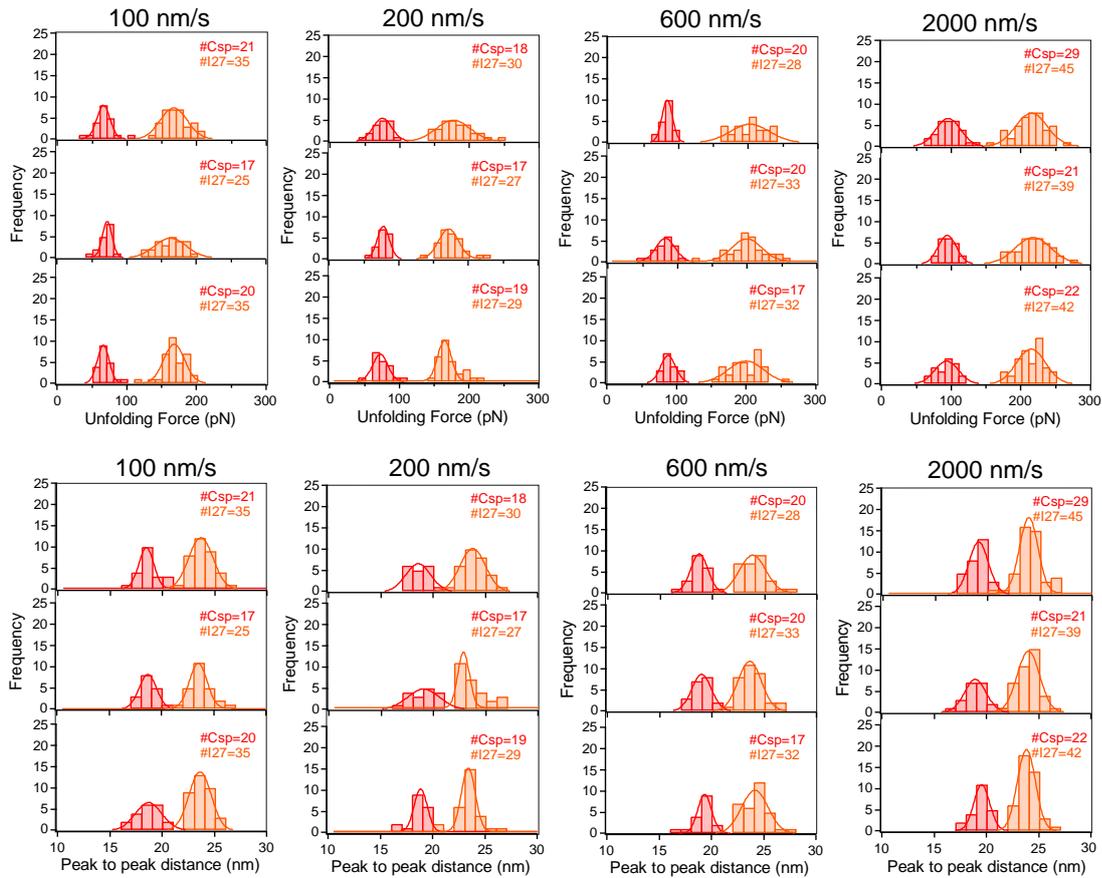
pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p _{2p} Csp (nm)	I27 peaks	median force I27 (pN)	Median p _{2p} I27 (nm)
100 nm/s	1	19	45.8 ± 8.0	18.8 ± 0.9	26	162.9 ± 18.4	23.3 ± 0.8
	2	22	45.1 ± 10.8	18.8 ± 0.9	33	156.1 ± 15.5	23.5 ± 0.9
	3	19	47.1 ± 10.4	18.9 ± 1.0	34	157.9 ± 16.2	23.7 ± 1.4
	Average		45.8 ± 2.5	18.8 ± 0.1		159.1 ± 3.7	23.5 ± 0.2
200 nm/s	1	16	50.8 ± 12.1	18.7 ± 0.9	32	173.3 ± 26.7	23.5 ± 1.1
	2	20	52.1 ± 10.7	19.1 ± 1.1	28	176.8 ± 16.8	23.5 ± 1.3
	3	16	55.4 ± 9.9	19.2 ± 1.1	22	178.6 ± 17.6	23.7 ± 1.6
	Average		53.7 ± 2.4	19.0 ± 0.3		176.2 ± 1.8	23.6 ± 0.1
600 nm/s	1	18	68.1 ± 8.0	19.4 ± 1.1	25	199.0 ± 21.4	23.6 ± 0.9
	2	19	62.2 ± 9.4	18.8 ± 1.3	25	201.1 ± 23.6	23.6 ± 1.1
	3	18	64.8 ± 13.4	18.8 ± 0.9	28	194.7 ± 18.3	23.5 ± 0.7
	Average		65.0 ± 3.0	19.0 ± 0.4		197.9 ± 4.5	23.6 ± 0.1
2000 nm/s	1	22	79.9 ± 10.9	18.65 ± 0.8	33	219.3 ± 20.6	23.75 ± 0.89
	2	18	74.2 ± 12.3	19.28 ± 1.2	25	217.6 ± 27.2	23.83 ± 0.82
	3	18	75.6 ± 11.9	19.27 ± 1.2	26	214.1 ± 23.1	24.27 ± 1.26
	Average		76.6 ± 3.0	19.1 ± 0.4		217.0 ± 2.7	24.0 ± 0.3

32 °C BsCsp



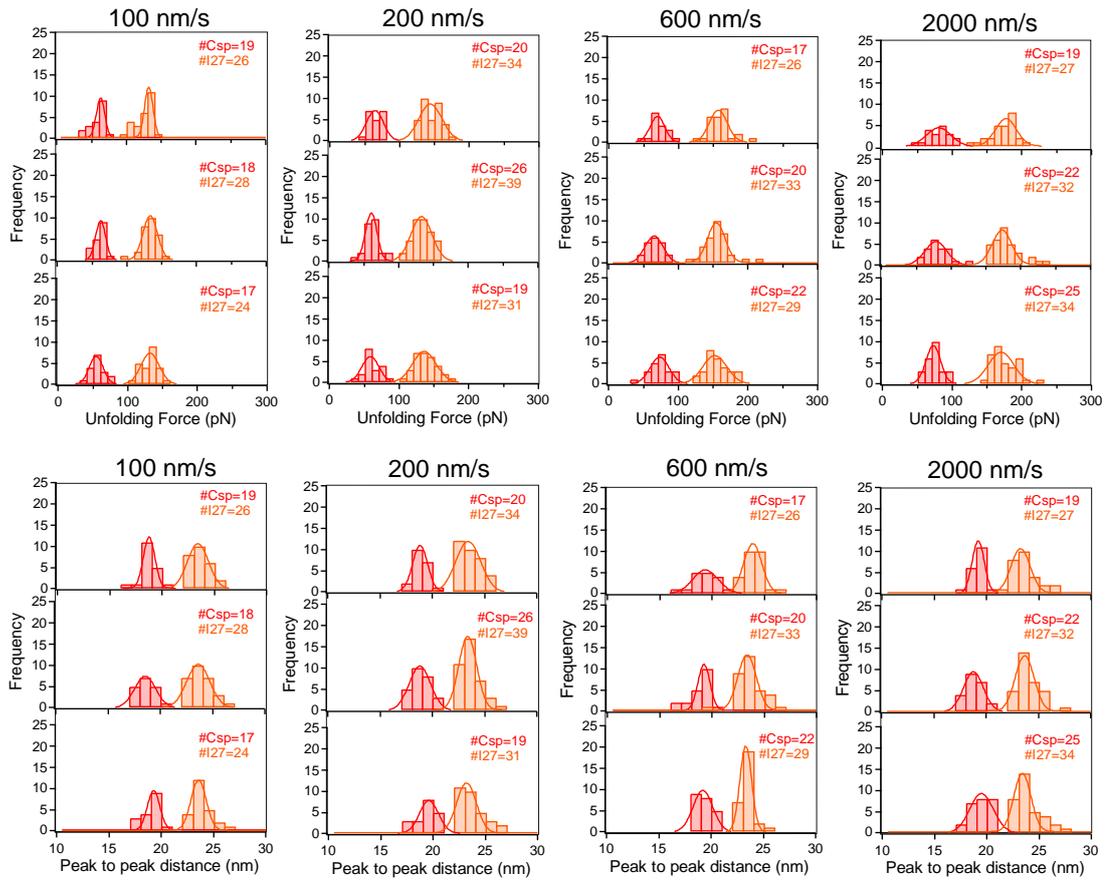
pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p ₂ p Csp (nm)	I27 peaks	median force I27 (pN)	Median p ₂ p I27 (nm)
100 nm/s	1	16	30.9 ± 6.9	19.3 ± 0.9	26	126.0 ± 16.2	23.4 ± 0.8
	2	18	35.3 ± 6.9	19.3 ± 1.2	28	129.5 ± 13.8	23.4 ± 1.0
	3	21	35.9 ± 8.5	18.6 ± 1.1	39	133.1 ± 16.5	23.6 ± 0.9
	Average		34.1 ± 2.8	19.1 ± 0.4		129.6 ± 3.5	23.5 ± 0.1
200 nm/s	1	21	37.3 ± 8.7	18.9 ± 1.0	38	137.8 ± 13.9	23.5 ± 1.3
	2	17	38.2 ± 6.8	18.5 ± 0.9	25	140.3 ± 13.4	23.2 ± 0.8
	3	22	41.2 ± 7.7	18.9 ± 1.2	44	146.0 ± 16.0	23.8 ± 0.8
	Average		38.9 ± 2.0	18.8 ± 0.3		141.3 ± 4.2	23.5 ± 0.3
600 nm/s	1	25	50.2 ± 8.9	18.9 ± 1.0	43	160.4 ± 22.5	23.6 ± 1.1
	2	21	53.0 ± 8.1	18.6 ± 1.1	30	153.4 ± 17.4	23.2 ± 0.9
	3	24	53.5 ± 12.6	19.2 ± 1.1	41	158.1 ± 18.7	23.7 ± 1.1
	Average		52.2 ± 1.8	18.87 ± 0.3		157.3 ± 3.6	23.5 ± 0.3
2000 nm/s	1	20	63.6 ± 9.3	18.6 ± 1.1	40	168.8 ± 19.5	23.9 ± 1.0
	2	22	58.4 ± 9.6	19.4 ± 1.0	42	170.9 ± 15.9	24.3 ± 1.0
	3	21	60.3 ± 11.9	18.6 ± 1.1	37	174.1 ± 23.0	23.5 ± 1.0
	Average		60.7 ± 2.6	18.9 ± 0.4		171.0 ± 2.9	23.9 ± 0.4

14 °C TmCsp



pulling speed	Repeat	Csp peaks	Median force Csp (pN)	Median p _{2p} Csp (nm)	I27 peaks	median force I27 (pN)	Median p _{2p} I27 (nm)
100 nm/s	1	21	68.2 ± 14.4	18.5 ± 0.9	35	167.6 ± 18.4	23.7 ± 1.1
	2	17	73.1 ± 9.8	18.7 ± 0.8	25	163.2 ± 20.2	23.8 ± 1.1
	3	20	67.6 ± 10.6	18.4 ± 1.2	35	166.7 ± 16.1	23.8 ± 1.0
	Average		69.6 ± 3.1	18.5 ± 0.1		166.1 ± 2.4	23.7 ± 0.1
200 nm/s	1	18	77.9 ± 13.1	18.3 ± 1.2	30	178.1 ± 25.5	23.8 ± 1.2
	2	17	78.1 ± 10.1	19.1 ± 1.3	27	175.9 ± 18.0	23.3 ± 1.4
	3	19	75.1 ± 14.1	18.7 ± 1.0	29	169.0 ± 15.5	23.4 ± 1.3
	Average		77.0 ± 1.7	18.7 ± 0.4		174.3 ± 4.8	23.5 ± 0.3
600 nm/s	1	20	83.6 ± 10.0	18.6 ± 0.9	28	202.1 ± 21.6	24.1 ± 1.3
	2	20	82.1 ± 13.9	19.2 ± 1.0	33	198.6 ± 23.3	23.9 ± 1.2
	3	17	86.6 ± 9.7	19.1 ± 1.2	32	196.3 ± 23.1	24.2 ± 1.2
	Average		84.1 ± 2.3	19.0 ± 0.3		199.0 ± 2.9	24.1 ± 0.2
2000 nm/s	1	29	97.9 ± 16.2	19.1 ± 0.9	45	214.9 ± 23.2	23.97 ± 1.25
	2	21	95.9 ± 12.2	18.8 ± 1.1	39	217.6 ± 23.7	24.17 ± 1.10
	3	22	92.4 ± 14.3	19.3 ± 1.2	42	214.4 ± 18.9	23.72 ± 0.93
	Average		95.4 ± 2.8	19.1 ± 0.3		215.7 ± 1.7	24.0 ± 0.2

32 °C TmCsp



pulling speed	Repeat	Csp peaks	Median force (pN)	Median p2p (nm)	I27 peaks	median force (pN)	Median p2p (nm)
	100 nm/s	1	19	60.7 ± 10.8	18.5 ± 0.8	26	129.0 ± 11.6
	2	18	62.7 ± 7.5	18.7 ± 1.0	28	134.1 ± 10.5	23.7 ± 1.2
	3	17	57.9 ± 10.3	19.3 ± 0.9	24	132.3 ± 12.8	23.6 ± 1.1
	Average		60.4 ± 2.4	18.8 ± 0.4		131.8 ± 2.6	23.6 ± 0.2
200 nm/s	1	20	64.8 ± 10.0	18.6 ± 0.7	34	140.8 ± 13.5	23.5 ± 1.0
	2	26	61.6 ± 11.8	18.5 ± 0.9	39	133.6 ± 13.4	23.6 ± 1.1
	3	19	59.6 ± 11.3	19.2 ± 1.0	31	136.7 ± 15.4	23.2 ± 1.2
	Average		62.0 ± 2.6	18.7 ± 0.4		137.0 ± 3.6	23.4 ± 0.2
600 nm/s	1	17	69.0 ± 11.6	19.3 ± 1.2	26	158.0 ± 16.7	24.0 ± 1.0
	2	20	66.6 ± 11.2	18.8 ± 1.3	33	154.7 ± 17.5	23.5 ± 1.3
	3	22	71.8 ± 15.4	19.3 ± 0.8	29	154.0 ± 15.7	23.6 ± 0.7
	Average		69.2 ± 2.6	19.1 ± 0.3		155.6 ± 2.1	23.7 ± 0.3
2000 nm/s	1	19	81.8 ± 16.2	19.1 ± 0.8	27	172.4 ± 18.9	23.4 ± 1.2
	2	22	78.0 ± 15.8	18.8 ± 0.8	32	175.7 ± 19.9	23.8 ± 1.1
	3	25	76.5 ± 11.0	19.6 ± 1.0	34	169.9 ± 18.1	23.7 ± 1.3
	Average		78.8 ± 2.7	19.2 ± 0.4		172.7 ± 2.9	23.6 ± 0.2

References

- 1) Onuchic J. N., Luthey-Schulten Z. and Wolynes P. G. (1997) Theory of protein folding: the energy landscape perspective. *Ann. Rev. Phys. Chem.* **48**: 545-600
- 2) Jenik M., Parra R. G., Radusky L. G., Turjanski A., Wolynes P. G. and Ferreiro D. U. (2012). Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* **40**: 348-351
- 3) Dunker A. K., Lawson J. D., Brown C. J., Williams R. M., Romero P., Oh J. S., Oldfield C. J., Campen A. M., Ratliff C. M., Hipps K. W., Ausio J., Nissen M. S., Reeves R., Kang C., Kissinger C. R., Bailey R. W., Griswold M. D., Chiu W., Garner E. C. and Obradovic Z. J. (2001) Intrinsically disordered protein. *Mol. Graphics Modell.* **19**(1) 26-59
- 4) Lodish H., Berk A., Zipursky S. L., Matsudaira P., Baltimore D., and Darnell J. (2000) *Molecular Cell Biology*. 4th edition. New York: Section 2.2, Noncovalent Bonds.
- 5) Graziano G. (2014) On the mechanism of cold denaturation. *Phys. Chem. Chem. Phys.* **16**: 21755-21767
- 6) Karan R., Capes M. and DasSarma S. (2012) Function and biotechnology of extremophilic enzymes in low water activity. *Aquat. Biosyst.* **8**(1): 4
- 7) Pace C. N., Treviño S., Prabhakaran E. and Scholtz, J. M. (2004). Protein structure, stability and solubility in water and other solvents. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **359**(1448): 1225-1235
- 8) Teilum K., Olsen J. G. and Kragelund B. B. (2009) Functional aspects of protein flexibility. *Cell. Mol. Life Sci.* **66**(14): 2231-2247
- 9) Koshland D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **44**(2): 98-104
- 10) Bennett W. S. and Steit, T. A. (1978). Glucose-induced conformational change in yeast hexokinase. *Proc. Natl. Acad. Sci. U.S.A.* **75**(10): 4848-4852

- 11) Changeux J. P. and Edelstein S. (2011). Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol. Rep.* **3**: 19
- 12) Sammond D. W., Kastelowitz N., Himmel M. E., Yin H., Crowley M. F., Bomble Y. J. (2016) Comparing residue clusters from thermophilic and mesophilic enzymes reveals adaptive mechanisms. *PLoS. One.* **11**(1).
- 13) Vihinen M. (1987) Relationship of protein flexibility to thermostability. *Protein Eng.* **1**(6): 477-480
- 14) Matthews, B. W., Nicholson, H., and Becktel W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **84**(19): 6663-6667
- 15) Shoichet B. K., Baase W A., Kuroki R. and Mathews B. W. (1995) A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **92**(2): 452-456
- 16) Kumar S., Tsai C.J. and Nussinov R. (2001) Thermodynamic differences among homologous thermophilic and mesophilic proteins. *Biochemistry* **40**(47): 14152-14165
- 17) Collins T., Meuwis M.A., Gerday C. and Feller G. (2003). Activity, stability and flexibility in glycosidases adapted to extreme thermal environments. *J. Mol. Biol.* **328**(2): 419-428
- 18) Perrier V., Burlacu-Miron S., Bourgeois S., Surewicz W. K. and Gilles A. M. (1998) Genetically engineered zinc-chelating adenylate kinase from *Escherichia coli* with enhanced thermal stability. *J. Biol. Chem.* **273**(30): 19097-19101
- 19) Struvay C. and Feller G. (2002) Optimization to low temperature activity in psychrophilic enzymes. *Int. J. Mol. Sci.* **13**(9): 11643-11665
- 20) Cavicchioli R., Siddiqui K. S., Andrews D. and Sowers K. R. (2002) Low-temperature extremophiles and their applications. *Curr. Opin. Biotechnol.* **13**(3): 253-261

- 21) Brock T. D. and Brock M. L. (1968) Relationship between environmental temperature and optimum temperature of bacteria along a hot spring thermal gradient. *J. Appl. Bacteriol.* **31**(1): 54-58
- 22) Brock T. D., Brock M. L., Bott T. L. and Edwards M. R. (1971). Microbial life at 90 C: the sulfur bacteria of Boulder Spring. *J. Bacteriol.* **107**(1): 303-314
- 23) Stetter K. O. (2011) History of discovery of extremophiles, Springer, Tokyo
- 24) Morita R.Y. (1975) Psychrophilic bacteria. *Bacteriol. Rev.* **39**(2): 144-167
- 25) Liebert M. A. and Kuhn E. (2012) Towards understanding life under subzero conditions: The significance of exploring psychrophilic “cold-shock” proteins. *Astrobiology* **12**(11): 1078-1086
- 26) DasSarma S. and DasSarma P. (2012) Halophiles. *eLS*
- 27) Fangemil J., Zhang L. and Bazylnski D. A. (2010) Deep-sea piezosphere and piezophiles: geomicrobiology and biogeochemistry. *Trends Microbiol.* **18**(9): 413-422
- 28) Baker-Austin C. and Dopson M. (2007) Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* **15**(4): 165-171
- 29) Horikoshi K. (1999) Alkaliphiles: Some applications of their products for biotechnology. *Microbiol. Mol. Biol. Rev.* **63**(4): 735-750
- 30) Rothschild L. J. and Mancinelli R. L. (2001) Life in extreme environments. *Nature* **409**: 1092-1101
- 31) Risso V. A., Gavira J. A. and Sanchez-Ruiz J. M. (2014) Thermostable and promiscuous precambrian proteins. *Environ. Microbiol.* **16**(6): 1485-1489
- 32). Gaucher E. A., Govindarajan S., and Ganesh O. K. (2008) Palaeotemperature trend for precambrian life inferred from resurrected proteins. *Nature* **451**(7179): 704–707

- 33) Di Giulio M. (2003) The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J. Theor. Biol.* **221**(3): 425-436
- 34) Hart K. M., Harms M. J., Schmidt B. H., Elya C., Thornton J. W. and Marqusee S. (2014) Thermodynamic system drift in protein evolution. *PLoS Biology*: **12**(11)
- 35) Wheeler L. C., Lim S. A., Marqusee S. and Harms M. J. (2016) The thermostability and specificity of ancient proteins. *Curr. Opin. Struc. Biol.* **38**: 37-43
- 36) Brock T. D. and Freeze H. (1969) *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile *J. Bacteriol.* **98**(1): 289-297
- 37) Saiki, R. K., Gelfand D. H., Stoffel S., Scharf S. J., Higuchi R., Horn G. T., Mullis K. B. and Erlich H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* **239**: 487-491
- 38) Schuliger J. W., Brown S. H., Baross J. A. and Kelly R. M. (1993) Purification and characterization of a novel amylolytic enzyme from ES4, a marine hyperthermophilic archaeum. *Mol. Mar. Biol. Biotech.* **2**: 76-87
- 39) Elleuche S., Schro C., Sahm K. and Antranikian G. (2014) Extremozymes – biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotech.* **29**: 116-123
- 40) Sterner R. and Liebl W. (2001) Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol.* **36**(1): 39-106
- 41) Eisenmesser E. Z., Millet O., Labeikovsky W., Korzhnev D. M., Wolf-Watz M., Bosco D. A., Skalicky J. J., Kay L. E. and Kern D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature.* **438**(7064): 117-121
- 42) Hernandez G., Jenney F. E. Jr., Adams M. W. and LeMaster D. M. (2000) Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature *Proc. Natl. Acad. Sci. U.S.A.* **97**(7): 3166-3170
- 43) Amico S., Marx J. C., Gerday C., and Feller G. (2003) Activity-stability relationships in extremophilic enzymes. *J Biol Chem.* **278**(10): 7891-7896

- 44) Radestock S. and Gohlke H. (2011) Protein rigidity and thermophilic adaptation. *Proteins*. **79**(4): 1089-1108
- 45) Xu B.-L., Dai M., Chen Y., Meng D., Wang Y., Fang N., Tang X. and Tanga B. (2015) Improving the thermostability and activity of a thermophilic subtilase by Incorporating Structural elements of its psychrophilic counterpart. *Appl. Environ. Microbiol.* **81**(18): 6302-6313
- 46) Tian J., Wu N., Chu X. and Fan Y. (2010) Predicting changes in protein thermostability brought about by single- or multi-site mutations. *BMC Bioinf.* **11**: 370
- 47) De Baets G., Van Durme J., van der Kant R., Schymkowitz J. and Rousseau F. (2015) Solubis: Optimize your protein. *Bioinformatics*. **31**(15): 2580-2582
- 48) Yanase M., Takata H., Fujii K., Takaha T. and Kuriki T. (2005) Cumulative effect of amino acid replacements results in enhanced thermostability of potato type L alpha-glucan phosphorylase. *Appl. Environ. Microbiol.* **71**(9): 5433-5439
- 49) Mount D. M. (2004) Bioinformatics: Sequence and genome analysis (2nd ed.). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY
- 50) Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I. N. and Bourne P. E. (2000) The protein data bank. *Nucleic Acids Res.* **28**: 235-242
- 51) Geierhaas C. D., Nickson A. A., Lindorff-Larsen K., Clarke J., Vendruscolo M. (2007) BPPred: A Web-based computational tool for predicting biophysical parameters of proteins. *Protein Sci.* **16**: 125-134
- 52) Zhang Y. (2008). "Progress and challenges in protein structure prediction". *Curr. Opin. Struct. Biol.* **18**(3): 342-348
- 53) Howell S. C., Inampudi K. K., Bean D. P. and Wilson C. J. (2014) Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. *Structure* **22**(2): 218-229

- 54) Kundu S. and Roy D. (2009) Comparative structural studies of psychrophilic and mesophilic protein homologues by molecular dynamics simulation. *J. Mol. Graph Model* **27**(8): 871-880
- 55) Bartlett G. J., Porter C. T., Borkakoti N. and Thornton J. M. (2002) Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **324**: 105-121
- 56) Bae E. and Phillips G. N. Jr (2004) Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J Biol Chem.* **279**(27): 28202-28208
- 57) Marx J. C., Collins T., Amico S. D., Feller G. and Gerday C. (2007) Cold-adapted enzymes from marine Antarctic microorganisms. *Mar. Biotechnol.* **9**(3): 293-304
- 58) Perutz M. F. and Raidt H. (1975) Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin. *Nature.* **255**(5505): 256-259
- 59) Sadeghi M., Naderi-Manesh H., Zarrabi M. and Ranjbar B. (2006) Effective factors in thermostability of thermophilic proteins. *Biophys. Chem.* **119**: 256-270
- 60) Miroshnichenko M. L. and Bonch-Osmolovskaya E. A. (2006) Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. *Extremophiles* **10**: 85-96
- 61) Robinson-Rechavi M. and Godzik A. (2005) Structural genomics of *Thermotoga maritima* proteins shows that contact order is a major determinant of protein thermostability. *Structure.* **13**(6): 857-860
- 62) Siddiqui K. S., Feller G. , D'Amico S., Gerday C., Giaquinto L. and Cavicchioli R. (2005) The active site is the least stable structure in the unfolding pathway of a multidomain cold-adapted alpha-amylase. *J Bacteriol.* **187**: 6197-6205
- 63) Rathi P. C., Höffken H. W. and Gohlke H. (2014) Quality matters: extension of clusters of residues with good hydrophobic contacts stabilize (hyper)thermophilic proteins. *J. Chem. Inf. Model.* **54**(2): 355-361

- 64) Pack S. P. and Yoo Y. J. (2004) Protein thermostability: structure-based difference of amino acid between thermophilic and mesophilic proteins. *J. Biotechnol.* **111**(3): 269-277
- 65) Amico S. D., Collins T., Marx J. and Feller G. (2006) Psychrophilic microorganisms: challenges for life. *EMBO reports* **7**: 385-389
- 66) Rodrigues D. F. and Tiedje J. M. (2008) Coping with our cold planet. *Appl. Environ. Microbiol.* **74**(6): 1677-1686
- 67) Graumann P. and Marahiel M. A. (1996) Some like it cold: response of microorganisms to cold shock. *Arch Microbiol.* **166**(5): 293-300
- 68) Cary S. C., McDonald I. R., Barrett J. E. and Cowan D. A. (2010) On the rocks: The microbiology of Antarctic dry valley soils. *Nat. Rev. Microbiol.* **8**: 129-138.
- 69) García-Arribas O., Mateo R., Tomczak M. M., Davies P. L. and Mateu M. G. (2007) Thermodynamic stability of a cold-adapted protein, type III antifreeze protein, and energetic contribution of salt bridges. *prot. sci.* **16**(2): 227-238
- 70) Deming J. W. (2002) Psychrophiles and polar regions. *Curr. Op. Microbio.* **5**(3): 301-309
- 71) Feller G. and Gerday C. (2003) Psychrophilic enzymes: Hot topics in cold adaptation. *Nat. Rev. Microbiol.* **1**(3): 200-208
- 72) Phadtare S. (2004) Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.* **6**(2): 125-136
- 73) Bakermans C. and Nealson K. H. (2004). Relationship of critical temperature to macromolecular synthesis and growth yield in *Psychrobacter cryopegella*. *J. Bacteriol.* **186**(8): 2340-2345
- 74) Piette F., Struvay C. and Feller G. (2011) The protein folding challenge in psychrophiles: facts and current issues. *Environ. Microbiol.* **13**(8): 1924-1933
- 75) Chattopadhyay M. K. (2006) Mechanism of bacterial adaptation to low temperature. *J. Biosci.* **31**(1): 157-165

- 76) Feller G. (2003) Molecular adaptations to cold in psychrophilic enzymes. *Cell. Mol. Life Sci.* **60**: 648-662
- 77) Goldstein R. A. (2007) Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: Insights from the quasi-chemical approximation. *Protein Sci.* **16**(9): 1887-1895
- 78) D'Amico S., Claverie P., Collins T., Georlette D., Gratia E., Hoyoux A., Meuwis M. A., Feller G. and Gerday C. (2002) Molecular basis of cold adaptation. *Phil. Trans. R. Soc. Lond. B.* **357**(1423): 917-925
- 79) Josepha B., Ramtekea P. W. and Thomas G. (2008) Cold active microbial lipases: Some hot issues and recent developments. *Biotechnology Adv.* **26**(5): 457-470
- 80) Chiuri R., Maiorano G., Rizzello A., del Mercato L. L., Cingolani R., Rinaldi R., Maffia M. and Pompa P. P. (2009) Exploring local flexibility/rigidity in psychrophilic and mesophilic carbonic anhydrases. *Biophys. J.* **96**(4): 1586-1596
- 81) Becktel W. J. and Schellman J. A. (1987) Protein stability curves. *Biopolymers* **26**: 1859-1877
- 82) Feller G. and Gerday C. (1997) Psychrophilic enzymes: molecular basis of cold adaptation. *Cell Mol Life Sci.* **53**(10) 830-841
- 83) Kumar S. and Nussinov R. (2004) Different roles of electrostatics in heat and in cold: adaptation by citrate synthase. *Chem. Bio. Chem.* **5**(3): 280-290
- 84) Saelsminde G., Halskau O. and Jonassen I. (2009) Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role. *Extremophiles.* **13**: 11-20
- 85) Feller G. (2007) Life at low temperatures: is disorder the driving force? *Extremophiles.* **11**(2): 211-216
- 86) Kim S., Hwang K. Y., Kim S., Sung H., Han Y. S., and Cho Y. (1999) Structural basis for cold adaptation sequence, biochemical properties, and crystal structure of

malate dehydrogenase from a psychrophile *aquaspirillum articum*. *J. Biol. Chem.* **274**: 11761-11767

87) Yan B. X., Sun Y. Q. (1997) Glycine residues provide flexibility for enzyme active sites. *J Biol. Chem.* **272**: 3190-3194

88) Lopez C. F., Darst R. K. and Rossky P. J. (2008) Mechanistic elements of protein cold denaturation. *J. Phys. Chem. B.* **121**(19): 5961-5967

89) Davidovic M., Mattea C., Qvist J. and Halle B. (2009) Protein cold denaturation as seen from the solvent. *J. Am. Chem. Soc.* **131**(3): 1025-1036

90) Dias C. L., Ala-Nissila T., Wong-ekkabut J., Vattulainen I., Grant, M. and Karttunen M. (2010) The hydrophobic effect and its role in cold denaturation. *Cryobiology* **60**(1): 91-99

91) Feller G. (2010) Protein stability and enzyme activity at extreme biological temperatures. *J. Phys. Condens. Matter* **22**(32): 323101

92) Siddiqui K. S. and Cavicchioli R. (2006) Cold-Adapted enzymes. *Annu. Rev. Biochem.* **75**: 403-433

93) Ueda M., Goto T., Nakazawa M., Miyatake K., Sakaguchi M. and Inouye K. (2010) A novel cold-adapted cellulase complex from *Eisenia foetida*: Characterization of a multienzyme complex with carboxymethylcellulase, beta-glucosidase, beta-1,3 glucanase, and beta-xylosidase. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **157**: 26-32

94) Tronelli D., Maugini E., Bossa F. and Pascarella S. (2007) Structural adaptation to low temperatures – analysis of the subunit interface of oligomeric psychrophilic enzymes. *FEBS* **274**(17): 4595-4608

95) Zhou. H. Z. (2002) Toward the physical basis of thermophilic proteins: Linking of enriched polar interactions and reduced heat capacity of unfolding. *J. Biophys.* **83**: 3126-3133

- 96) Shivaji S. and Prakash S. S. (2010) How do bacteria sense and respond to low temperature? *Arch. Microbiol.* **192**(2): 85-95
- 97) Siddiqui K.S. and Thomas T. (2008) Protein adaptation in extremophiles, Nova Biomecial Books, New York
- 98) Schmid B., Klumpp J., Raimann E., Loessner M. J., Stephan R. and Tasara T. (2009) Role of cold shock proteins in growth of *Listeria monocytogenes* under cold and osmotic stress conditions. *Appl. Environ. Microbiol.* **75**(6): 1621-1627
- 99) Prévost D., Drouin P., Laberge S., Bertrand A., Cloutier J. and Lévesque G. (2013) Cold-adapted rhizobia for nitrogen fixation in temperate regions. *Can. J. Botany* **81**: 1153-1161
- 100) Cavicchioli R., Charlton T., Ertan H., Omar S. M., Siddiqui K. S. and Williams T. J. (2011) Biotechnological uses of enzymes from psychrophiles. *Microb. Biotechnol.* **4**(4): 449-460
- 101) van den Burg B. B. (2003) Extremophiles as a source for novel enzymes. *Curr. Op. Microbio.* **6**: 213-218
- 102) Margesin R. and Feller G. (2010) Biotechnological applications of psychrophiles. *Environ. Technol.* **31**(8): 835-844
- 103) Kasana R. C. and Gulati A. (2011) Cellulases from psychrophilic microorganisms: a review. *J. Basic Microbiol.* **51**(6): 572-579
- 104) Bajaj S., and Singh D. K. (2015). Biodegradation of persistent organic pollutants in soil, water and pristine sites by cold-adapted microorganisms. *Int. Biodeter. Biodegr.* **100**: 98-105
- 105) Dahiya N., Tewari R. and Hoondal G. S. (2006) Biotechnological aspects of chitinolytic enzymes: a review. *Appl. Microbiol. Biotechnol.* **71**(6): 773-782
- 106) Collins T., Gerday C. and Feller G. (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev.* **29**(1): 3-23

- 107) Jeon J., Kim J. T., Kang S., Lee J. H., Kim S. J. (2009) Characterization and its potential application of two esterases derived from the Arctic sediment metagenome. *Mar. Biotechnol.* **11**: 307-316
- 108) Kobori H., Sullivan C. W. and Shizuya H. (1984) Heat-labile alkaline phosphatase from Antarctic bacteria: rapid 5' end labelling of nucleic acids. *Proc Natl. Acad. Sci. U.S.A.* **81**: 6691-6695
- 109) Wang E., Koutsioulis D., Leiros H. K. S., Andersen O. A., Bouriotis V., Hough E. and Heikinheimo P. (2007) Crystal structure of alkaline phosphatase from the Antarctic bacterium TAB5. *J. Mol. Biol.* **366**(4): 1318-1331
- 110) Mihailovich M., Militti C., Gabaldón T. and Gebauer F. (2010) Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays.* **32**(2): 109-118
- 111) Jiang W., Hou Y. and Inouye M. (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.* **272**(1): 196-202
- 112) Schindelin H., Jiang W., Inouye M. and Heinemann U. (1994) Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **91**(11): 5119-5123
- 113) Tang Y., Schneider W. M., Shen Y., Raman, S., Inouye M., Baker D., Roth, M. J. and Montelione G. T. (2010) Fully automated high-quality NMR structure determination of small (2)H-enriched proteins. *J. Struct. Funct. Genomics* **11**(4): 223-232
- 114) Yamanaka K. and Inouye M. (1997) Growth-phase-dependent expression of cspD, encoding a member of the CspA family in *Escherichia coli*. *J. Bacteriol.* **179**(16): 5126-5130
- 115) Horn G., Hofweber R., Kremer W. and Kalbitzer H. R. (2007) Structure and function of bacterial cold shock proteins. *Cell. Mol. Life Sci.* **64**(12): 1457-1470
- 116) Ermolenko D. N. and Makhatadze G. I. (2002) Bacterial cold-shock proteins. *Cell. Mol. Life Sci.* **59**(11): 1902-1913

- 117) Lopez M. M., Yutani K. and Makhatadze G. I. (1999) Interactions of the major cold shock protein CspB with single-stranded DNA templates of different base composition. *J. Biol. Chem.* **274**(47): 33601-33608
- 118) Max K. E. A., Zeeb M., Bienert R., Balbach J. and Heinemann U. (2006) T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein *Bs-CspB*. *J. Mol. Biol.* **360**(3): 702-714
- 119) Giuliadori A. M., Di Pietro F., Marzi S., Masquida B., Wagner R., Romby P., Gualerzi C. O. and Pon C. L. (2010) The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol. Cell.* **37**(1): 21-33
- 120) Mitta M., Fang L. and Inouye M. (1997) Deletion analysis of *cspA* of *Escherichia coli*: requirement of the AT-rich UP element for *cspA* transcription and the downstream box in the coding region for its cold shock induction. *Mol. Microbiol.* **26**(2): 321-335
- 121) Jiang W., Fang L., and Inouye M. (1996) The role of the 5'-end untranslated region of the mRNA for CspA, the major cold-shock protein of *Escherichia coli*, in cold-shock adaptation. *J. Bacteriol.* **178**(16): 4919-4925
- 122) Bae W., Jones P. G. and Inouye M. (1997). CspA, the major cold shock protein of *Escherichia coli*, negatively regulates its own gene expression. *J. Bacteriol.* **179**(22): 7081-7088
- 123) Phadtare S., Inouye M. and Severinov K. (2002) The nucleic acid melting activity of *Escherichia coli* CspE is critical for transcription antitermination and cold acclimation of cells. *J. Biol. Chem.* **277**(9): 7239-7245
- 124) Gualerzi C. O., Giuliadori A. M. and Pon C. L. (2003) Transcriptional and post-transcriptional control of cold-shock genes. *J. Mol. Biol.* **331**(3): 527-539
- 125) Hofweber R., Horn G., Langmann T., Balbach J., Kremer W., Schmitz G. and Kalbitzer H. R. (2005) The influence of cold shock proteins on transcription and translation studied in cell-free model systems. *FEBS J.* **272**(18): 4691-4702

- 126) Dersch P., Kneip S. and Bremer E. (1994) The nucleoid-associated DNA-binding protein H-NS is required for the efficient adaptation of *Escherichia coli* K-12 to a cold environment. *Mol. Gen. Genet.* **245**(2): 255-259
- 127) Mykytczuk N. C., Foote S. J., Omelon C. R., Southam G., Greer C. W. and Whyte L. G. (2013) Bacterial growth at 15 °C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1 *ISME J.* **7**: 1211-1226
- 128) Jones P. G. and Inouye M. (1996) RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response. *Mol. Microbiol.* **21**(6): 1207-1218
- 129) D. K. Schiffenbauer J., Woulfe S. L., Zacheis M. and Schwartz B. D. (1988) Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl. Acad. Sci. U.S.A.* **85**(19): 7322-7326
- 130) Lindquist J. A., Brandt S., Bernhardt A., Zhu C. and Mertens P. R. (2014) The role of cold shock domain proteins in inflammatory diseases. *J. Mol. Med. (Berl).* **92**(3): 207-216
- 131) Matsumoto G., Yajima N., Saito H., Nakagami H., Omi Y., Lee U. and Kaneda Y. (2010) Cold shock domain protein A (CSDA) overexpression inhibits tumor growth and lymph node metastasis in a mouse model of squamous cell carcinoma. *Clin. Exp. Metastasis.* **27**(7): 539-547
- 132) Zeng Y., Kulkarni P., Inoue T. and Getzenberg R. H. (2009) Down-regulating cold shock protein genes impairs cancer cell survival and enhances chemosensitivity. *J. Cell. Biochem.* **107**(1): 179-188
- 133) Sutherland B. W., Kucab J., Wu J., Lee C., Cheang M.C.U., Yorida E., Turbin D., Dedhar S., Nelson C., Pollak M., Grimes H. L., Miller K., Badve S., Huntsman D., Blake-Gilks C., Chen M., Pallen C. J. and Dunn S. E. (2005) Akt phosphorylates the Y-box binding protein 1 at Ser102 located in the cold shock domain and affects the anchorage-independent growth of breast cancer cells. *Oncogene* **24**(26): 4281-4292

- 134) Knott G. (2015) Neurodegeneration: Cold shock protects the brain. *Nature* **518**(7538): 177-178
- 135) Peretti D., Bastide A., Radford H., Verity N., Molloy C., Martin M. G., Moreno J. A., Steinert J. R., Smith T., Dinsdale D., Willis A. E. and Mallucci G. R. (2015) RBM3 mediates structural plasticity and protective effects of cooling in neurodegeneration. *Nature* **518**(7538): 236-239
- 136) Hoffmann T., Tych K. M., Brockwell D. J and Dougan L. (2013) Single molecule force spectroscopy identifies a small cold shock protein as being mechanically robust. *J. Phys. Chem. B.* **117**(6): 1819-1826
- 137) Tych K. M., Hoffmann T., Brockwell D. J. and Dougan L. (2013) Single molecule force spectroscopy reveals the temperature-dependent robustness and malleability of a hyperthermophilic protein. *Soft Matter* **9**: 9016-9025
- 138) Hansen L., Knudsen S. and Sørensen S. (1998). The effect of the lacY gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*. *S. Curr. Microbiol.* **36**: 341
- 139) Malke H., Davis R. W., Botstein D. and Roth J. R. (1981) A manual for genetic engineering, advanced bacterial genetics. *Allg. Mikrobiol.* **21**: 767
- 140) Waterhouse A. M., Procter J. B., Martin D. M. A., Clamp M. and Barton G. J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**(9): 1189-1191
- 141) Greenfield N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. protoc.* **1**(6): 2527-2535
- 142) Moon C. P. and Fleming K. G. (2011) Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes. *Methods Enzymol.* **492**: 189-211.
- 143) Rees D. C. and Robertson A. D. (2001) Some thermodynamic implications for the thermostability of proteins, *Protein Sci.* **10**: 1187-1194

- 144) Pace C. N. and Shaw K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **4**: 1-7
- 145) Scholtz J. M., Grimsley G. R. and Pace C. N. (2009) Solvent denaturation of proteins and interpretations of the m value. *Method Enzymol.* **466**: 549-565
- 146) Pace C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**: 266-280
- 147) Tanford C. (1968) Protein denaturation. *Adv. Protein Chem.* **23**: 121-282
- 148) Jackson S. E. and Fersht A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state. *Transition Biochemistry* **30**: 10428-10435
- 149) Hutter J. L. and Bechhoefer J. (1993) Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64**: 1868-1873
- 150) Florin E. L., Rief M., Lehmann H., Ludwig M., Dornmair C., Moy V. T. and Gaub H. E. (1995) Sensing specific interactions with the atomic force microscope. *Biosens. Bioelectron.* **10**: 895-901
- 151) Levy R. and Maaloum M. (2002) Measuring the spring constant of atomic force microscope cantilevers : thermal fluctuations and other methods. *Nanotech.* **13**: 33-37
- 152) Marko J. F., Siggia E. D. and Smith S. (1994) Entropic elasticity of X-phage DNA explicit and implicit learning and maps of cortical motor output. *Science* **265**(5178): 1599-1600
- 153) Rief M., Fernandez J. M. and Gaub H. E. (1998) Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* **81**: 4764
- 154) Leea D., Honga Y., Choea E., Leea S., Kima S., Leeb H., Oha J., Shinc H. and Pyuna Y. (2005) A thermodynamic study of mesophilic, thermophilic, and hyperthermophilic l-arabinose isomerases: The effects of divalent metal ions on protein stability at elevated temperatures. *FEBS Letters* **579**(5): 1261-1266

- 155) Razvi A. and Scholtz J. M. (2006) Lessons in stability from thermophilic proteins *Protein Sci.* **15**(7): 1569-1578
- 156) Sawle L. and Ghosh K. (2011) How do thermophilic proteins and proteomes withstand high temperature? *Biophys. J.* **101**(1): 217-227
- 157) Hollien J. and Marqusee S. (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38**: 3831-3836
- 158) Ohmura T., Ueda T., Ootsuka K., Saito M. and Imoto T. (2001) Stabilization of hen egg white lysozyme by a cavity-filling mutation. *Protein Sci.* **10**(2): 313-320
- 159) Robic S., Guzman-Casado M., Sanchez-Ruiz J. M. and Marqusee S. (2003) Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. U.S.A.* **100**(20): 11345-11349
- 160) Fu H., Grimsley G. and Pace C. N. (2010) Increasing protein stability: importance of $\Delta C(p)$ and the denatured state. *Protein Sci.* **19**: 1044-1052
- 161) Beadle B. M., Baase W. A., Wilson D. B., Gilkes N. R. and Shoichet B. K. (1999) Comparing the thermodynamic stabilities of a related thermophilic and mesophilic enzyme. *Biochemistry* **38**: 2570-2576
- 162) Graziano G. (2008) Is there a relationship between protein thermal stability and the denaturation heat capacity change? *J. Therm. Anal. Calorim.* **93**: 429-438
- 163) Cipolla A., Delbrassine F., Da Lage J. L. and Feller G. (2012) Temperature adaptations in psychrophilic, mesophilic and thermophilic chloride-dependent alpha-amylases. *Biochimie.* **94**(9): 1943-1950
- 164) Xu Y., Feller G., Gerday C., Glansdorff N. (2003) Moritella cold-active dihydrofolate reductase: are there natural limits to optimization of catalytic efficiency at low temperature? *J. Bacteriol.* **185**(18): 5519-5526
- 165) Peng Q. and Li H. (2008) Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: Evidence for a kinetic partitioning mechanism *Proc. Natl. Acad. Sci. U.S.A.* **105**(6): 1885-1890

- 166) Makhatadze G. I., Loladze V. V., Gribenko A. V. and Lopez M. M. (2004) Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions *J. Mol. Biol.* **336**(4): 929-942
- 167) Heilmayer O., Brey T. and Portner H. O. (2004) Growth efficiency and temperature in scallops: a comparative analysis of species adapted to different temperatures. *Funct. Ecol.* **18**: 641-647
- 168) Waller C. L., Overall A., Fitzcharles E. M. and Griffiths H. (2016) First report of *Laternula elliptica* in the Antarctic intertidal zone. *Polar Biol.* **3**: 1-4
- 169) Gerasimova A. V., Martynov F. M., Filippova N. A. and Maximovich N. V. (2016) Growth of *Mya arenaria* L. at the northern edge of the range: heterogeneity of soft-shell clam growth characteristics in the White Sea. *Helgol Mar Res* **457**: 1-14
- 170) Wassenberg D., Welker C. and Jaenicke R. (1999) Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol.* **289**: 187-193
- 171) Motono C., Gromiha M. M. and Kumar S. (2008) Thermodynamic and kinetic determinants of *Thermotoga maritima* cold shock protein stability: a structural and dynamic analysis. *Proteins.* **71**(2): 655-669
- 172) Kaufman-Szymczyk A., Wojtasik A., Parniewski P., Białkowska A., Tkaczuk K. and Turkiewicz M. (2009) Identification of the csp gene and molecular modelling of the CspA-like protein from Antarctic soil-dwelling psychrotrophic bacterium *Psychrobacter* sp. B6. *Acta Biochim. Pol.* **56**(1): 63-69
- 173) Auman A. J., Breezee J. L., Gosink J. J., Kämpfer P. and Staley J. T. (2006) *Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. *Int. J. Syst. Evol. Microbiol.* **56**(5): 1001-1007
- 174) Riley M., Staley J. T., Danchin A., Wang T. Z., Brettin T. S., Hauser L. J., Land M. L. and Thompson L. S. (2008) Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. *BMC Genomics* **6**(9): 210

- 175) Breezee J., Cady N. and Staley J. T. (2004) Subfreezing growth of the sea ice bacterium "*Psychromonas ingrahamii*". *Microbial Ecol.* **47**(3): 300-304
- 176) Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402
- 177) Bozal N., Jesus Montes M., Tudela E., Jimenez F. and Guinea J. (2002) *Shewanella frigidimarina* and *Shewanella livingstonensis* sp. nov. isolated from Antarctic coastal areas. *Int. J. Syst. Evol. Micr.* **52**: 195-205
- 178) Schindler T., Perl D., Graumann P., Sieber V., Marahiel M. A. and Schmid F. X. (1998), Surface-exposed phenylalanines in the RNP1/RNP2 motif stabilize the cold-shock protein CspB from *Bacillus subtilis*. *Proteins* **30**: 401-406
- 179) Perl D. and Schmid F. X. (2001) Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **313**: 343-357
- 180) Schindler T., Graumann P. L., Perl D., Ma S., Schmid F.X. and Marahiel M.A. (1999) The family of cold shock proteins of *Bacillus subtilis* : Stability and dynamics in vitro and in vivo. *J. Biol. Chem.* **274**(6): 3407-3413
- 181) Welker C., Bohm G., Schurig H. and Jaenicke, R. (1999) Cloning, overexpression, purification, and physicochemical characterization of a cold shock protein homologue from the hyperthermophilic bacterium *Thermotoga maritima*. *Protein Sci.* **8**: 394-403
- 182) Martin S. R. and Schilstra M. J. (2008) Circular dichroism and its application to the study of biomolecules. *Methods Cell. Biol.* **84**: 263-293
- 183) Makhatadze G. I. and Marahiel M. A. (1994) Effect of pH and phosphate ions on self-association properties of the major cold-shock protein from *Bacillus subtilis*. *Protein Sci.* **3**: 2144-2147
- 184) Wunderlich M., Martin A. and Schmid F. X. (2005) Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of coulombic interactions. *J. Mol. Biol.* **347**(5): 1063-1076

- 185) Motono C., Gromiha M. and Kumar S. (2008) Thermodynamic and kinetic determinants of *Thermotoga maritima* cold shock protein stability: A structural and dynamic analysis. *Proteins* **71**(2): 655-669
- 186) Myers J. K., Pace C. N., and Scholtz J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein. Sci.* **4**(10): 2138-2148
- 187) Zhou H-X. (2002) Toward the physical basis of thermophilic proteins: Linking of Enriched polar interactions and reduced heat capacity of unfolding. *Biophys. J.* **83**: 3126-3133
- 188) Liu Y. and Sturtevant J. M. (1996) The observed change in heat capacity accompanying the thermal unfolding of proteins depends on the composition of the solution and on the method employed to change the temperature of unfolding. *Biochemistry* **35**(9): 3059-3062
- 189) Nishii I., Kataoka M., Tokunaga F. and Goto Y. (1994) Cold denaturation of the molten globule states of apomyoglobin and a profile for protein folding. *Biochemistry.* **33**(16): 4903-4909
- 190) Tych K. M., Batchelor M., Hoffmann T., Wilson M. C., Paci E., Brockwell D. J. and Dougan L. (2016) Tuning protein mechanics through an ionic cluster graft from an extremophilic protein. *Soft Matter* **12**: 2688-2699
- 191) Tych K. M., Batchelor M., Hoffmann T., Wilson M. C., Hughes M. L., Paci E., Brockwell D. J. and Dougan L. (2016) Differential effects of hydrophobic core packing residues for thermodynamic and mechanical stability of a hyperthermophilic protein. *Langmuir* **32**(29): 7392-7402
- 192) Závodszy P., Kardos J., Svingor S. and Petsko G. A. (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**(13): 7406-7411

- 193) Best R. B., Rutherford T. J., Freund S. M. and Clarke J. (2004) Hydrophobic core fluidity of homologous protein domains: relation of side-chain dynamics to core composition and packing. *Biochemistry* **43**(5): 1145-1155
- 194) Berezovsky I. N., Chen W. W., Choi P. J. and Shakhnovich E. I. (2005) Entropic stabilization of proteins and its proteomic consequences. *PLoS Comput. Biol.* **1**(4): 47
- 195) Lee B. I., Chang C., Cho S. J., Eom S. H., Kim K. K., Yu Y. G. and Suh S. W. (2001) Crystal structure of the MJ0490 gene product of the hyperthermophilic archaeobacterium *Methanococcus jannaschii*, a novel member of the lactate/malate family of dehydrogenases. *J. Mol. Biol.* **307**(5): 1351-1362
- 196) Hoyoux A., Blaise V., Collins T., D'Amico S., Gratia E., Huston A. L., Marx J. C., Sonan G., Zeng Y., Feller G. and Gerday C. (2004) Review: Extreme catalysts from low-temperature environments. *J. Biosci. Bioeng.* **98**(5): 317-330
- 197) Wrba A., Schweiger A., Schultes V., Jaenicke R. and Závodszy P. (1990) Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. *Biochemistry US.* **29**(33): 7584-7592
- 198) Daniel R. M., Cowan D. A., Morgan H. W. and Curran M. P. (1982) A correlation between protein thermostability and resistance to proteolysis. *Biochem. J.* **3**: 641-644
- 199) Danson M. J., Hough D. W., Russell R. J., Taylor G. L. and Pearl L. (1996) Enzyme thermostability and thermoactivity. *Protein Eng.* **9**(8): 629-630
- 200) Lam S. Y., Yeung R. C., Yu T. H., Sze K. H. and Wong K. B. (2011) A rigidifying salt-bridge favours the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. *PLOS* **9**(3): 1027.....
- 201) Varley P. G. and Paint R. H. (1991) Relation between stability, dynamics and enzyme activity in 3-phosphoglycerate kinases from yeast and *Thermus thermophilus*. *J. Mol. Biol.* **220**: 531-538

- 202) Lazaridis T., Lee I. and Karplus M. (1997) Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein. Sci.* **6**(12): 2589-2605
- 203) Mamonova T. B., Glyakina A. V., Kurnikova M. G. and Galzitskaya O. V. (2010) Flexibility and mobility in mesophilic and thermophilic homologous proteins from molecular dynamics and FoldUnfold method. *J. Bioinform. Comput. Biol.* **8**(3): 377-394
- 204) Svingor A., Kardos J., Hajdú I., Németh A. and Závodszky P. (2001) A better enzyme to cope with cold. Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic IPMDHs. *J. Biol. Chem.* **276**(30): 28121-28125
- 205) Paredes D. I., Watters K., Pitman D. J., Bystroff C. and Dordick J. S. (2011) Comparative void-volume analysis of psychrophilic and mesophilic enzymes: Structural bioinformatics of psychrophilic enzymes reveals sources of core flexibility *BMC Struct. Biol.* **20**(11): 42
- 206) Tehei M., Madern D., Franzetti B. and Zaccai G. (2005) Neutron scattering reveals the dynamic basis of protein adaptation to extreme temperature. *J Biol. Chem.* **280**: 40974–40979
- 207) Colombo G. and Merz K. M. (1999) Stability and activity of mesophilic subtilisin E and its thermophilic homolog: Insights from molecular dynamics simulations. *J. Am. Chem. Soc.* **121**(29): 6895-6903
- 208) Merkley E. D., Parson W. W. and Daggett V. (2010) Temperature dependence of the flexibility of thermophilic and mesophilic flavoenzymes of the nitroreductase fold. *Prot. Eng. Des. Sel.* **23**(5): 327-336.
- 209) Fitter J. and Heberle J. (2000) Structural equilibrium fluctuations in mesophilic and thermophilic α -Amylase. *Biophys. J.* **79**(3): 1629-1636.
- 210) Butterwick J. A., Loria J. P., Astrof N. S., Kroenke C. D., Cole R., Rance M. and Palmer A. G. 3rd. (2004) Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J. Mol. Biol.* (2004) **339**: 855-871

- 211) Hernandez G., Jenney F. E. Jr, Adams M. W. and LeMaster D. M. (2000) Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proc. Natl. Acad. Sci. U.S.A.* **97**(7): 3166-3170
- 212) Blöchl E., Rachel R., Burggraf S., Hafenbradl D., Jannasch H. W. and Stetter K. O. (1997) *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**(1): 14-21
- 213) Meinhold L., Clement D., Tehei M., Daniel R., Finney J. L. and Smith J. C. (2008) Protein Dynamics and Stability: The Distribution of Atomic Fluctuations in Thermophilic and Mesophilic Dihydrofolate Reductase Derived Using Elastic Incoherent Neutron Scattering. *Biophys J.* **94**(12): 4812-4818
- 214) Kotamarthi H. C., Yadav A., and Koti Ainavarapu S. R. (2015) Small peptide binding stiffens the ubiquitin-like protein SUMO1. *Biophys J.* **108**(2): 360-367
- 215) Papaleo E., Pasi M., Riccardi L., Sambì I., Fantucci P. and De Gioia L. (2008) Protein flexibility in psychrophilic and mesophilic trypsins. Evidence of evolutionary conservation of protein dynamics in trypsin-like serine-proteases *FEBS Lett.* **582**(6): 1008-1018
- 216) Olufsen M., Smalås A. O., Moe E. and Brandsdal B. O. (2005) Increased flexibility as a strategy for cold adaptation: a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase *J. Biol. Chem.* **280**: 18042-18048
- 217) Oyeyemi O. A., Sours K. M., Lee T., Kohen A., Resing K. A., Ahn N. G. and Klinman J. P. (2011) Comparative hydrogen–deuterium exchange for a mesophilic vs thermophilic dihydrofolate reductase at 25 °C: Identification of a single active site region with enhanced flexibility in the mesophilic protein. *Biochemistry.* **50**(38): 8251-8260.
- 218) Wolf-Watz M., Thai V., Henzler-Wildman K., Hadjipavlou G., Eisenmesser E. Z. and Kern D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* **11**: 945-949

- 219) Vemparala S., Mehrotra S. and Balaram H. (2011) Role of loop dynamics in thermal stability of mesophilic and thermophilic adenylosuccinate synthetase: a molecular dynamics and normal mode analysis study. *Biochim. Biophys. Acta.* **1814**(5): 630-637
- 220) Martinez R., Schwaneberg U. and Roccatano D. (2011) Temperature effects on structure and dynamics of the psychrophilic protease subtilisin S41 and its thermostable mutants in solution. *Protein Eng. Des. Sel.* **24**(7): 533-544
- 221) Kim H., Kim S., Jung Y., Han J., Yun J. H., Chang I. and Lee W. (2016) Probing the folding-unfolding transition of a thermophilic protein, MTH1880. *PLoS One.* **11**(1)
- 222) Borgia A., Kemplen K. R., Borgia M. B., Soranno A., Shammass S., Wunderlich B., Nettels D., Best R. B., Clarke J. and Schuler B. (2015) Transient misfolding dominates multidomain protein folding. *Nat. Commun.* **17**(6): 8861
- 223) Kelch B. A. and Agard D. A. (2007) Mesophile versus thermophile: insights into the structural mechanisms of kinetic stability. *J. Mol. Biol.* **370**: 784-795
- 224) Magg C., Kubelka J., Holtermann G., Haas E. and Schmid F. X. (2006) Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol.* **360**: 1067-1080
- 225) Perl D., Welker C., Schindler T., Schroder K., Marahiel M. A., Jaenicke R. and Schmid F. X. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat. Struct. Biol.* **5**: 229-235
- 226) Jackson S. E. (1998) How do small single-domain proteins fold? *Fold Des.* 81-91
- 227) Riddle D. S., Santiago J. V., Bray-Hall S. T., Doshi N., Grantcharova V. P., Yi Q. and Baker D. (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **4**(10): 805-809
- 228) Dams T. and Jaenicke R. (1999) Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemistry.* **38**(28): 9169-9178

- 229) Cavagnero S., Debe D. A., Zhou Z. H., Adams M. W. and Chan S. I. (1998) Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. *Biochemistry*. **37**(10):3369-3376
- 230) Baumann H., Knapp S., Lundback T. and Ladenstein R. (1994) Hard T: Solution structure and DNA-binding properties of a thermostable protein from the archeon *Sulfolobus solfataricus*. *Nat. Struct. Biol.* **1**: 808-819
- 231) Gunasekaran K., Eyles S. J., Hagler A. T. and Gierasch L. M. (2001) Keeping it in the family: folding studies of related proteins. *Curr. Opin. Struct. Biol.* **11**(1):83-93
- 232) Kishan K. V. and Agrawal V. (2005) SH3-like fold proteins are structurally conserved and functionally divergent. *Curr. Protein Pept. Sci.* **6**(2): 143-50
- 233) Kay L. E. (2005) NMR studies of protein structure and dynamics *Journal of Magn. Reson.* **173**: 193-207
- 234) Mueller U., Perl D., Schmid F. X. and Heinemann U. (2000) Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J. Mol. Biol.* **297**(4): 975-988
- 235) Dinner A. R. and Karplus M. (2001) The roles of stability and contact order in determining protein folding rates. *Nat. Struct. Biol.* **8**: 21-22
- 236) Otzen D. E., Kristensen O., Proctor M. and Oliveberg M. (1999) Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry* **38**: 6499-6511
- 237) Zeeb M. and Balbach J. (2003) Single-stranded DNA binding of the cold-shock protein CspB from *Bacillus subtilis*: NMR mapping and mutational characterization *Protein Sci.* **12**: 112-123
- 238) Sachs R., Max K. E., Heinemann U. and Balbach J. (2012) RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA* **18**: 65-76

- 239) Kremer W., Schuler B., Harrieder S., Geyer M., Gronwald W., Welker C., Jaenicke R. and Kalbitzer H. R. (2001) Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. *Eur. J. Biochem.* **268**: 2527-2539
- 240) Jung A., Bamann C., Kremer W., Kalbitzer H. R. and Brunner E. (2004) High-temperature solution NMR structure of *TmCsp*. *Protein Sci.* **13**(2): 342-350
- 241) Urry D. W. (1999) Elastic molecular machines in metabolism and soft-tissue restoration. *Trends Biotechnol.* **17**(6): 249-257
- 242) Johnson J. C. and Korley L. T. J. (2012) Enhanced mechanical pathways through nature's building blocks: amino acids. *Soft Matter* **8**: 11431-11442
- 243) Rico F., Rigato A., Picas L. and Scheuring S. (2013) Mechanics of proteins with a focus on atomic force microscopy. *J Nanobiotechnology.* **11**(1): 3
- 244) Hoffmann T. and Dougan L. (2012) Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.* **41**: 4781-4796
- 245) Li Y., Wray R., Eaton C. and Blount P. (2009) An open-pore structure of the mechanosensitive channel MscL derived by determining transmembrane domain interactions upon gating. *J. FASEB* **23**(7): 2197-204
- 246) Best R. B., Brockwell D. J., Toca-Herrera J. L., Blake A. W., Smith D.A., Radford S. E. and Clarke J. (2003) Force mode atomic force microscopy as a tool for protein folding studies. *Anal. Chim. Acta* **479**(1): 87-105
- 247) Linke W. A., Grützner A. (2008) Pulling single molecules of titin by AFM- recent advances and physiological implications *Pflugers Arch.* **456**(1): 101-115
- 248) Hughes M. L. and Dougan L. (2016) The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep. Prog. Phys.* **79**(7) 1-39

- 249) Willett R. L., Baldwin K. W., West K. W. and Pfeiffer L. N. (2005) Differential adhesion of amino acids to inorganic surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **102**(22): 7817-7822
- 250) Butt H. J., Cappella B. and Kappl M. (2005) Force measurements with the atomic force microscope: Technique, interpretation and applications. *Surf. Sci. Rep.* **59**: 1-152
- 251) Binnig G., Quate C. F. and Gerber C. (1986) Atomic Force Microscope. *Phys. Rev. Lett.* **56**: 930
- 252) Mitsui K., Hara M. and Ikai A. (1996) Mechanical unfolding of alpha 2-macroglobulin molecules with atomic force microscope. *FEBS Lett.* **385**: 29-33
- 253) Lee G. U., Chrisey L. A. and Colton R. J. (1994) Direct measurement of the forces between complimentary strands of DNA. *Science* **266**: 771-773
- 254) Muller D. J. and Drene Y. F. (2008). Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nat. Nano.* **3**(5): 261-269
- 255) Oberhauser A. F. and Carrión-Vázquez M. (2008) Mechanical biochemistry of proteins one molecule at a time. *J. Biol. Chem.* **283**: 6617-6621.
- 256) Best R. B., Li B., Steward A., Daggett V. and Clarke J. (2001) Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* **81**(4): 2344-2356
- 257) Dey A. and Szoszkiewicz R. (2012) Complete noise analysis of a simple force spectroscopy AFM setup and its applications to study nanomechanics of mammalian Notch 1 protein. *Nanotechnology* **23**(17): 175101
- 258) Valbuena A., Vera A. M., Oroz J., Menéndez M., Carrión-Vázquez M. (2012) Mechanical properties of β -catenin revealed by single-molecule experiments. *Biophys J.* **103**(8): 1744-1752
- 259) Best R. B. and Clarke J. (2002) What can atomic force microscopy tell us about protein folding? *Chem. Commun. (Camb).* **7**(3): 183-192

- 260) Brockwell D. J. (2007) Probing the mechanical stability of proteins using the atomic force microscope. *Biochemical Soc. T.* **35**(6): 1564-1568
- 261) Rounsevell R., Forman J. R. and Clarke J. (2004). Atomic force microscopy: mechanical unfolding of proteins. *Methods* **34**(1): 100-111
- 262) Brockwell D. J., Paci E., Zinober R. C., Beddard G. S., Olmsted P. D., Smith D. A., Perham R. N., Radford S. E. (2003) Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat. Struct. Biol.* **10**(9): 731-737
- 263) Fisher T. E., Oberhauser A. F., Carrion-Vazquez M., Marszalek P. E. and Fernandez J. M. (1999) Review: The study of protein mechanics with the atomic force microscope. *Trends in Biochem. Sci.* **24**(10): 379-384
- 264) Brockwell D. J., Beddard G. S., Paci E., West D. K., Olmsted P. D., Smith D. A., Radford S. E. (2005) Mechanically unfolding the small, topologically simple protein L. *Biophys J.* **89**(1): 506-519
- 265) Hoffmann T., Tych K. M., Hughes M. L., Brockwell D. J., Dougan L. (2013) Towards design principles for determining the mechanical stability of proteins. *Phys. Chem. Chem. Phys.* **15**: 15767-15780
- 266) Tych K. M., Hoffmann T., Batchelor M., Hughes M. L., Kendrick K. E., Walsh D. L., Wilson M. C., Brockwell D. J. and Dougan L. (2015) Life in extreme environments: Single molecule force spectroscopy as a tool to explore proteins from extremophilic organisms. *Biochem. Soc. T.* **43**: 179-185
- 267) Yang G., Cecconi C., Baase W. A., Vetter I. R., Breyer W. A., Haack J. A., Matthews B. W., Dahlquist F. W. and Bustamante C. (2000) Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 139-144
- 268) Steward A., Toca-Herrera J. L. and Clarke J. (2002) Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci.* **11**(9): 2179-2183

- 269) Zheng P., Cao Y. and Li H. (2011) Facile Method of Constructing Polyproteins for Single-Molecule Force Spectroscopy Studies. *Langmuir* **27**(10): 5713-5718
- 270) Zinober R. C., Brockwell D. J., Beddard G. S., Blake A. W., Olmsted P. D., Radford S. E. and Smith D. A. (2002) Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold. *Protein Sci.* **11**(12): 2759-2765
- 271) Tych K. M., Hughes M. L., Bourke J., Taniguchi Y., Kawakami M., Brockwell D. J., Dougan L. (2015) Optimizing the calculation of energy landscape parameters from single-molecule protein unfolding experiments. *Soft Matter Phys.* **91**(1): 12710-12719
- 272) Evans E. and Ritchie K. (1997) Dynamic Strength of Molecular Adhesion. *Biophys J.* **72**(4): 1541-1555
- 273) Rief M., Gautel M., Oesterhelt F., Fernandez J. M. and Gaub H. E. (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**: 1109-1111
- 274) Fowler S. B., Best R. B., Toca Herrera J. L., Rutherford T. J., Steward A., Paci E., Karplus M. and Clarke J. (2002) Mechanical unfolding of a Titin Ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.* **322**(4): 841-849
- 275) Rief M., Pascual J., Saraste M. and Gaub H. E. (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.*, **286**: 553-561
- 276) Carrion-Vazquez M., Marszalek P. E., Oberhauser A. F. and Fernandez J. M. (1999) Atomic force microscopy captures length phenotypes in single proteins. *Proc. Natl. Acad. Sci. USA.* **96**: 11288-11292
- 277) Scholl Z. N., Josephs E. A. and Marszalek P. E. (2016) Modular, nondegenerate polyprotein scaffolds for atomic force spectroscopy. *Biomacromolecules.* **17**(7): 2502-2505

- 278) Valbuena A., Vera A. M., Oroz J., Menéndez M. and Carrión-Vázquez M. (2012) Mechanical properties of β -Catenin revealed by single-molecule experiments. *Biophys J.* **103**(8): 1744-1752
- 279) Dietz H., Bertz M., Schlierf M., Berkemeier F., Bornschlöggl T., Junker J. P., Rief M. (2006) Cysteine engineering of polyproteins for single-molecule force spectroscopy. *Nat. Protoc.* **1**(1): 80-84
- 280) Carrion-Vazquez M., Oberhauser A. F., Fisher T. E., Marszalek P. E., Li H. and Fernandez J. M. (2000) Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Trends in Biochem. Sci.* **24**(10): 379-384
- 281) Hoffmann T., Tych K. M., Crosskey T., Schiffrin B., Brockwell D. J. and Dougan L. (2015) Rapid and robust polyprotein production facilitates single-molecule mechanical characterization of β -barrel assembly machinery polypeptide transport associated domains. *ACS. Nano.* **9**: 8811-8821
- 282) Gibson D. G., Benders G. A., Andrews-Pfannkoch C., Denisova E. A., Baden-Tillson H., Zaveri J., Stockwell T. B., Brownley A., Thomas D. W., Algire M. A., Merryman C., Young L., Noskov V. N., Glass J. I., Venter J. C., Hutchison C. A. and Smith H. O. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220
- 283) Gibson D. G., Young L., Chuang R. Y., Venter J. C., Hutchison C. A., Smith H. O. (2009) "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nat. Meth.* **6**: 343-345
- 284) Peng Q. and Li H. (2008) Biophysics atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: Evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **105**(6): 1885-1890
- 285) Brockwell D. J., Beddard G. S., Clarkson J., Zinober R. C., Blake A. W., Trinick J., Olmsted P. D., Smith D. A. and Radford S. E. (2002) The effect of core destabilization on the mechanical resistance of I27. *Biophys J.* **83**(1): 458-472

- 286) Carrion-Vazquez M., Oberhauser A. F., Fowler S. B., Marszalek P. E., Broedel S. E., Clarke J. and Fernandez J. M. (1999) Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl. Acad. Sci. U.S.A.* **30**(7): 3694-3697
- 287) Shortle D. and Ackerman M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**(5529): 487-489
- 288) Stirnemanna G., Kangb S., Zhou R. and Bernea B. J. (2014) How force unfolding differs from chemical denaturation. *Proc. Natl. Acad. Sci. U.S.A.* **111**(9): 3413-3418
- 289) Carrion-Vazquez M., Li H., Lu H., Marszalek P. E., Oberhauser A. F. and Fernandez J. M. (2003) The mechanical stability of ubiquitin is linkage dependent *Nat. Struct. Biol.* **10**: 738-743
- 290) Schönfelder J. and Muñoz V. (2016) A simple two-state protein unfolds mechanically via multiple heterogenous pathways at single-molecule resolution. *Nat. Comms.* **7**: 11777
- 291) Li H., Oberhauser A. F., Fowler S. B., Clarke J. and Fernandez J. M. (2000) Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. U.S.A.* **97**(12): 6527-6531
- 292) Cieplak M. and Sikora M. (2013) Topological features in stretching of proteins *Biochem. Soc. T.* (2013) **41**(2): 519-522
- 293) Brockwell D. J., Beddard G. S., Clarkson J., Zinober R. C., Blake A. W., Trinick J., Olmsted P. D., Smith D. A. and Radford S. E. (2002) *Biophys. J.* **83**: 458
- 294) Hlady V., Buijs J. and Malmsten M. (1998), *Biopolymers at Interfaces*, New York, 181.
- 295) Taniguchi Y., Brockwell D. J., Kawakami M. (2008) The effect of temperature on mechanical resistance of the native and intermediate states of I27. *Biophys J.* **95**(11): 5296-5305

- 296) Schlierf M. and Rief M. (2005) Temperature softening of a protein in single-molecule experiments. *J. Mol. Biol.* **354**: 497-503
- 297) Bell G. I. (1978) Models for the specific adhesion of cells to cells a theoretical framework for adhesion mediated by reversible bonds between cell surface molecules. *Science* **200**: 618
- 298) Sadler D. P., Petrik., Taniguchi Y., Pullen J. R., Kawakami M., Radford S. E. and Brockwell D. J. (2009) Identification of a mechanical rheostat in the hydrophobic core of protein L. *J. Mol. Biol.* **393**(1): 237-348
- 299) Randles L. G., Rounsevell R. W. S. and Clarke J. (2007) Spectrin domains lose cooperativity in forced unfolding. *Biophys. J.* **92**: 571-577
- 300) Lia M. S., Klimov D.K. and Thirumalai D. (2004) Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer* **45**(2): 573-579
- 301) Guzman D. L., Randall A., Baldi P. and Guan Z. B. (2010) Computational and single-molecule force studies of a macro domain protein reveal a key molecular determinant for mechanical stability. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 1989-1994