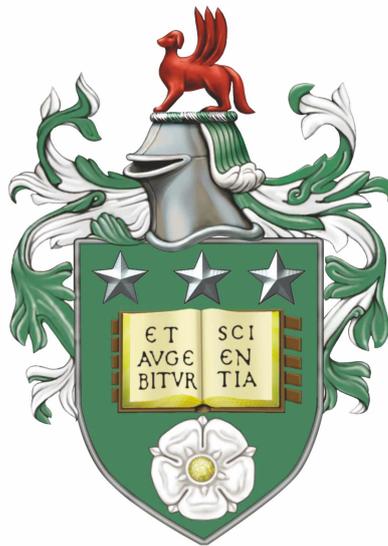


Measuring the Adapted Low Temperature
Resilience of Cold Shock Proteins - using Single
Molecule Force Spectroscopy and Nuclear
Magnetic Resonance



Katherine Ellen Kendrick

School of Physics and Astronomy
The University of Leeds

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

July 2018

Declaration

The candidate confirms that the work submitted is her own and that appropriate credit has been given 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 the thesis may be published without proper acknowledgement.

©2019 The University of Leeds and Katherine Ellen Kendrick.

The right of Katherine Ellen Kendrick to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

Dedication

I would like to dedicate this thesis to my grandparents: Kay and Steve, who supported me throughout my education and were always examples of the importance of hard work and kindness.

Acknowledgements

The work presented here was completed thanks to the help of a lot of other people. Firstly I must thank my supervisors Dr Lorna Dougan and Dr David Brockwell for being very supportive guides and for motivating through their visible excitement for science. I am also immensely grateful to the help I received from Dr Katarzyna Tych, Dr Lars Kuhn, Dr Arnout Kalverda, Dr Michael Wilson, Dr Danielle Walsh, Dr Megan Hughes, Dr Sam Lenton and everyone who was a part of the MNP and soft matter groups in Leeds. I also had a lot of support from members of the Radford-Brockwell Lab, especially Nasir Khan. They helped me with a lot of patience when I found myself wandering around in their lab as a confused physicist.

At Knowle Villa my friends Beth, Liam, Danielle, Godwin, Simon, Adam, Lucien and other temporary house members provided a good home. They made life great and full of games, films, cake, road trips, UV parties, dinosaurs, gin and general silliness. They also fed me. Fern and Julia put me up when needed and were excellent desk neighbours and yoga buddies. MNP and Astbury members have my gratitude for being such a great bunch who were a lot of fun to hang around with. The last few years have made me realise how lucky I am to have the family I was born into and I am so grateful for all of them. This is especially true for Mum, Dad and Rosie.

“The cold never bothered me anyway”

– Elsa, *Frozen*

Publications

Refereed Journals

K.M. Tych, T. Hoffmann, M. Batchelor, M.L. Hughes, **K.E. Kendrick**, D.L. Walsh, M. Wilson, D.J. Brockwell and L. Dougan. Life in extreme environments: Single molecule force spectroscopy as a tool to explore proteins from extremophilic organisms. *Biochemical Society Transactions*, 43:179–185, 2015.

List of Abbreviations

Experimental Techniques

AFM	Atomic Force Microscopy
CD	Circular Dichroism
CPMG	Carr-Purcell-Meiboom-Gill
FID	Free Induction Decay
FT	Fourier Transform
FX	Force Extension
HSQC	HeteroNuclear Single Quantum Coherence
MST	MicroScale Thermophoresis
NMR	Nuclear Magnetic Resonance
SDS-PAGE	Sodium Dodecyl Sulfate - - PolyAcrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEC-MALS	Size Exclusion Chromatography with - -Multi Angle Light Scattering
SMFS	Single-Molecule Force Spectroscopy

Variables

ΔG	Change in Gibbs free energy
------------	-----------------------------

ΔH	Change in enthalpy
ΔS	Change in entropy
k_D	Dissociation constant
F_U	Unfolding Force
P_2P	Inter-Peak Distance

Models

WLC	Worm-Like Chain
-----	-----------------

Biological Molecules

DNA	DeoxyRibonucleic Acid
ssDNA	single stranded DeoxyRibonucleic Acid
RNA	Ribonucleic Acid
ssRNA	single stranded Ribonucleic Acid
bp	basepairs
nt	nucleotides
CSP	Cold Shock Protein

Constants

k_B	Boltzmann's constant
h	Planck's constant
R	Gas constant

Units

mg	milligram
----	-----------

μg	microgram
g	gram
kg	kilogram
pN	pico Newton
ml	millilitre
μl	microLitre
nM	nanoMolar
μM	microMolar
mM	milliMolar
M	Molar
nm	nanometre
mm	millimetre
cm	centimetre
m	meter
s	seconds
Hz	Herz
IR	InfraRed
RF	Radio-Frequency
ppm	parts per million

Abstract

Cold Shock Proteins (CSPs) are produced when organisms experience cold-shock, a sudden drop to roughly 7-10 °C below the optimum temperature, and are thought to aid the melting of non-functional RNA secondary structures. For this function CSPs are required to stay flexible at temperatures determined by the optimum environment of the organism. While the thermodynamic properties of these proteins vary depending on the optimum temperature of the source organism, their structures are highly conserved, making this a good model system to explore non-structural adaptations to various environments.

Single molecule force spectroscopy was used to measure the mechanical properties of the CSP from *Bacillus Subtilis* while in the functional state, i.e. bound to nucleic acids. This binding was shown to stabilise the protein, without increasing the rigidity, even when the temperature of the complex was reduced.

The affinity of the CSP binding to ssDNA was measured and this was compared to the affinity to ssDNA of a CSP from an Arctic soil bacterium, *Psychrobacter sp6*. The CSP from the organism adapted to colder environments was found to have a lower binding affinity to ssDNA and the binding strength of ssDNA to both CSPs increased with decreasing temperature. When the binding of ssDNA was measured at the respective cold-shock temperatures of each organism the affinities were of the same order of magnitude, suggesting that the strength of this interaction is tuned to the native environment. To explore how functionality is maintained in organisms adapted to extremely cold environments, the dynamics of the CSP from *Psychrobacter sp6* was measured at low, physiologically relevant temperatures using a range of NMR techniques. Results showed evidence of slow exchange with a more disordered state at low temperatures together

with fast dynamics at the active site. These were not seen in measurements of the CSP from *Bacillus Subtilis* at low temperatures, indicating they are part of the protein adaptation to cold environments.

Contents

Publications	iv
List of abbreviations	v
Abstract	viii
Contents	x
List of tables	xv
List of figures	xviii
1 Introduction	1
1.1 Biomolecules	2
1.1.1 Nucleic Acids	2
1.1.2 Proteins	4
1.1.3 Protein Folding	5
1.1.4 Folding Thermodynamics	7
1.1.5 Dynamics and Function	12
1.1.6 Protein-Ligand binding	13
1.1.7 Dissociation Constant	13
1.1.8 Environmental Stress	14
1.2 Extremophiles	14
1.2.1 Psychrophiles	15
1.2.2 Psychrotrophic Organisms	16

1.2.3	Disorder in Cold Adaptation	18
1.3	The Cold Shock Protein (CSP)	18
1.3.1	Structure of the CSPs	20
1.3.2	Induction of the CSP	20
1.3.3	Binding of CSPs to Nucleic Acids	22
1.3.4	CSPs as RNA Chaperones	24
1.3.5	CSP Homologues Adapted to Extreme Environments as Model Systems	25
1.4	Single Molecule Force Spectroscopy (SMFS)	25
1.4.1	Models to Extract Energy Landscape Parameters from SMFS Ex- periments	26
1.4.2	CSPs in SMFS	28
1.4.3	Ligand Binding recorded with SMFS	30
1.5	Nuclear Magnetic Resonance Theory	37
1.5.1	The NMR signal	37
1.5.2	Multi-Dimensional NMR	40
1.5.3	Protein dynamics measured by NMR	45
1.5.4	Nuclear Spin Relaxation	46
1.5.5	Relaxation Dispersion Measured by Carl-Purcell-Meiboom-Gill (CPMG) Measurements	51
1.5.6	ZZ-Exchange	54
1.5.7	Relation of Protein Dynamics to Temperature Measured by Nu- clear Magnetic Resonance (NMR)	55
1.6	Discussion and Aims	55
2	Materials and Methods	58
2.1	Materials	58
2.1.1	General Chemicals	58
2.1.2	Molecular Biology Materials	59
2.1.3	AFM Materials	59
2.1.4	Buffers Used	59

2.2	Molecular Biology	60
2.2.1	Polymerase Chain Reaction (PCR)	62
2.2.2	Insertion of the TEV _p Restriction Site to the <i>Bs</i> -CSP Gene	63
2.2.3	Insertion of a Cysteine Residue into the <i>Bs</i> -CSP Gene	63
2.2.4	Agarose Gel Electrophoresis	64
2.2.5	Insertion of the <i>PB6</i> -CSP Gene into the pMal-c5x Plasmid	65
2.3	Production and Purification of the Cold Shock Proteins	68
2.3.1	Cell Culture Medium	68
2.3.2	Large Scale Growth	69
2.3.3	Large Scale Growth in ¹⁵ N and ¹³ C Labelled Minimal Medium	69
2.3.4	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE)	70
2.3.5	Lysing the Cells	71
2.3.6	Columns used for Protein Purification	71
2.3.7	Cleavage using Tobacco Etch Virus protease (TEV _p)	74
2.3.8	Purification of Each System	74
2.3.9	Dialysis and Lyophilisation	83
2.3.10	Circular Dichroism Spectroscopy	86
2.3.11	Monitoring Protein Aggregation with SEC-MALS	86
2.4	Measurements of Protein-DNA Interactions	87
2.4.1	Tryptophan Fluorescence Quenching	87
2.4.2	Isothermal Titration Calorimetry	89
2.4.3	MicroScale Thermophoresis (MST)	89
2.5	Single Molecule Force Spectroscopy (SMFS)	91
2.5.1	Polyprotein Unfolding Experimental Set-Up	91
2.5.2	Calibration of the AFM Cantilever	92
2.5.3	Polyprotein Pulling Experiments	94
2.5.4	The Worm Like Chain (WLC) Model	95
2.5.5	Protein Unfolding Analysis	95
2.5.6	Protein Unfolding Energy Landscape	97
2.5.7	Monte Carlo Simulations of Protein Unfolding	98

2.5.8	SMFS Experiments using Functionalised ssDNA Tips and <i>Bs</i> -CSP Functionalised Surfaces	99
2.6	Protein NMR Methods	102
2.6.1	Acquiring HSQC Spectra	102
2.6.2	Backbone assignment experiments	102
2.6.3	Relaxation measurements	103
3	Impact of Nucleic Acid Binding on Mechanical Stability of the Cold Shock Protein	105
3.0.1	Incorporating I27 in the Polyprotein Chain as a Marker Protein . .	106
3.1	Determining the Polyprotein-ssDNA Binding Association	106
3.2	Obtaining a Mechanical Fingerprint for (I27- <i>Bs</i> CSP) ₃ -I27 unfolding . . .	108
3.3	Triplicate Experiments show that a Robust Mechanical Fingerprint can be Obtained	110
3.4	Obtaining a Mechanical Fingerprint of the <i>Bs</i> -CSP Unfolding in the Pres- ence of ssDNA	113
3.4.1	Identification of Appropriate ssDNA Oligonucleotide Sequence for further SMFS Experiments	113
3.5	Determining the Pulling Speed Dependence of the <i>Bs</i> -CSP Unfolding Force	116
3.5.1	The Affect of Lower Temperature on Unfolding Force of the <i>Bs</i> - CSP	120
3.5.2	Determining the Effect of Low Temperature on the <i>Bs</i> -CSP Me- chanical unfolding in the Presence and Absence of ssDNA Binding	123
3.6	Determining Parameters of the Underlying Unfolding Energy Landscapes of the <i>Bs</i> -CSP using Monte Carlo (MC) Simulations	124
3.7	Discussion	127
4	Measuring the binding of ssDNA to the Cold Shock Protein	130
4.0.1	Limits of the Equilibrium Titrations	131
4.1	Tryptophan Fluorescence Quenching	133
4.1.1	<i>Bs</i> -CSP	134
4.1.2	<i>PB6</i> -CSP	136

4.2	Isothermal Titration Calorimetry	137
4.2.1	<i>Bs</i> -CSP	138
4.2.2	<i>PB6</i> -CSP	140
4.3	MicroScale Thermophoresis (MST) to explore Stoichiometry with labelled ssDNA	143
4.4	Single Molecule Force Spectroscopy of the ssDNA- <i>Bs</i> -CSP Complex Unbinding	145
4.4.1	Cysteine Mutated <i>Bs</i> -CSP binding capability	146
4.4.2	Traces Measured of Unbinding Force at Controllable Length	147
4.4.3	Statistical Analysis of Unbinding Forces	148
4.5	Discussion	151
4.5.1	Affinities Adapted to Organism Native Temperature	151
4.5.2	Stoichiometry of ssDNA Oligonucleotides binding to the CSPs	152
4.5.3	SMFS Measurements of ssDNA Unbinding from the <i>Bs</i> -CSP	154
5	Protein Dynamics measured by NMR	155
5.1	Sample Characterisation	158
5.1.1	Monitoring Polydispersity of <i>PB6</i> -CSP by SEC-MALS	158
5.1.2	<i>PB6</i> -CSP Thermo-Stability measurements using CD	158
5.2	NMR Spectra	159
5.2.1	HSQC	159
5.2.2	Temperature Dependent HSQC spectra	161
5.3	Protein Backbone Assignment using 3D Triple-Resonance Experiments	164
5.3.1	Assignment of <i>PB6</i> -CSP	165
5.3.2	Investigating the Additional Peaks	168
5.4	Relaxation Measurements	171
5.4.1	<i>ZZ</i> -Exchange	171
5.4.2	Relaxation Measurements and Time-scales	173
5.4.3	¹⁵ N- ¹ H-NOE Relaxation	174
5.4.4	Transverse Relaxation	176
5.4.5	Relaxation Dispersion measured with CPMG	180

5.4.6	Region Separation in Fit of Relaxation Dispersion	185
5.5	Discussion	185
6	Discussion and Future Work	189
6.1	Summary	189
6.2	Implications on the molecular adaptation of CSPs to low temperature function	196
6.3	SMFS to Confirm Localised Unfolding Identified in NMR Measurements .	197
6.4	Application to Protein Based Materials	198
A	Appendix Methods	199
B	Appendix Results	202

List of Tables

2.1	Composition of 50 times Tris-acetate-EDTA (TAE) buffer	59
2.2	Composition of 6× gel loading buffer	60
2.3	Composition of lysis buffer	60
2.4	Composition of wash buffer	60
2.5	Composition of elution buffer	60
2.6	Composition of high and low salt buffers for ion exchange chromatography	60
2.7	Table of the plasmids used throughout this project to produce the protein constructs.	61
2.8	Reagents used in PCR reactions for insertion and substitution mutations .	62
2.9	Thermocycling conditions used for PCR	63
2.10	Primers used for the mutation of the <i>Bs</i> -CSP plasmid to include the DNA encoding a TEVp recognition site	63
2.11	Primers used for the replacement of a serine residue with a cysteine in the <i>Bs</i> -CSP	64
2.12	Primers used for the amplification of the <i>PB6</i> -CSP gene, with EcoRI and BamHI recognition sites	65
2.13	Reagents used in PCR reactions for amplification of <i>PB6</i> -CSP gene	66
2.14	Composition of Lysogeny Broth (LB) used as growth medium for cell culturing	68
2.15	Composition of 5× M9 minimal salts.	69
2.16	Composition of M9 Minimal Medium	69
2.17	Compositions of the Sodium Dodecyl Sulfate PolyAcrylamide Gel	71

2.18	Table of the main constructs used throughout this project with the sections describing the purification of each.	75
2.19	Parameters used in Monte Carlo simulations of (I27- <i>Bs</i> CSP) ₃ -I27 unfolding events	99
2.20	Sequences of ssDNA strands used for SMFS unbinding experiments	100
3.1	Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27- <i>Bs</i> CSP) ₃ -I27 at 600 nms ⁻¹ and at 23 °C	112
3.2	ssDNA oligonucleotides used in SMFS unfolding experiments	114
3.3	Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27- <i>Bs</i> CSP) ₃ -I27 in the presence of different ssDNA oligonucleotides at 23 °C and with a pulling speed of 600 nms ⁻¹	115
3.4	Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27- <i>Bs</i> CSP) ₃ -I27 at 23 °C and at speeds of 100, 200, 600 and 2000 nms ⁻¹	119
3.5	Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27- <i>Bs</i> CSP) ₃ -I27 in the presence of CT2 ssDNA at 23 °C and at speeds of 100, 200, 600 and 2000 nms ⁻¹	119
3.6	Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27- <i>Bs</i> CSP) ₃ -I27 at 5 °C and at speeds of 100, 200, 600 and 2000 nms ⁻¹	122
3.7	Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27- <i>Bs</i> CSP) ₃ -I27 in the presence of CT2 ssDNA at 5 °C and at speeds of 100, 200, 600 and 2000 nms ⁻¹	122
4.1	K_D values for the <i>Bs</i> -CSP binding to ssDNA oligonucleotides determined by fits to the tight binding equation	135
4.2	K_D values for the <i>PB6</i> -CSP binding to ssDNA oligonucleotides determined by fits to the tight binding equation	136
4.3	Parameters obtained from the 1 to 1 binding model fit to ITC measurements of <i>PB6</i> -CSP and <i>Bs</i> -CSP binding to ssDNA oligonucleotides	141
A.1	Primers used for the sequencing of the final pMal c5x plasmid containing the <i>PB6</i> -CSP gene	199

A.2	Amino Acid Sequences of <i>Bs</i> -CSP Variants used	199
A.3	Amino Acid Sequence of the MBP- <i>PB6</i> -CSP construct	200
B.1	Parameters of the underlying energy landscapes calculated by Monte Carlo Simulations from experiments at 23 °C and 5 °C both with and without ssDNA present	202

List of Figures

1.1	Representations of the Central Dogma of Molecular Biology	3
1.2	Condensation Reaction Forming Peptide Bonds between Amino Acids . .	4
1.3	Structures and Labels of the 21 common Amino Acid Side Chains	5
1.4	The Forces involved in Protein Folding	6
1.5	Protein Secondary Structures	7
1.6	Example Tertiary Protein Structures	7
1.7	Temperature dependence of Enthalpic, Entropic and Free Energies of fold- ing of a typical protein	9
1.8	Energy Landscape of Protein Folding and Fluctuations	10
1.9	Example of an RNA stem-loop or hairpin structure	11
1.10	The Energy Diagram of a Reaction Process	12
1.11	Predicted Energy landscapes for proteins from thermophiles and psy- chrophiles	17
1.12	The Cold Shock Response	19
1.13	OB-fold Structure of the CSP	20
1.14	Induction of the CSPs by Temperature Dependence of mRNA Structure .	21
1.15	Structures of the <i>Bs</i> -CSP complexed with hexathymidine ssDNA oligonu- cleotide. PDB code 2ES2	23
1.16	Schematic of the binding of ssDNA oligos with <i>Bs</i> -CSP in both (a) the crystal structure and (b) in solution	23
1.17	RNA Chaperone Activity of the CSP measured by the Unfolding of RNA hairpins and Formation of a Hetero-Duplex	24
1.18	The Bell Model for Protein Unfolding	27

1.19	Schematic depiction of the increase in distance to the transition state along the mechanical unfolding co-ordinate of the <i>Tm</i> CSP with increasing temperature	29
1.20	Structure of the SUMO1 in the ligand-free state and bound with the S12 ligand	32
1.21	Unfolding forces from SMFS experiments plotted against pulling speed for the SUMO1 protein when ligand-free and in complex with S12 and S10	33
1.22	SMFS Measurements of the Unfolding Force of the RRM1 domain of TDP-41 in the Presence and Absence of ssDNA	36
1.23	Diagrams showing the production of an idealised NMR signal	38
1.24	Representation of the Spin Echo Pulse Sequence Focusing Incoherent Spins	41
1.25	INEPT Pulse Sequence	42
1.26	Simple HSQC diagram	43
1.27	Approximate Delays from J Coupling Constants for Magnetisation Transfer in Backbone Assignment Experiments	44
1.28	Diagrams of the experiments used for protein backbone assignment in NMR	45
1.29	Diagram of the Dynamics Time-Scales Accessible by Different NMR Techniques	46
1.30	Diagrams showing the Longitudinal Relaxation process and the effect of relaxation on NMR signal	49
1.31	Representation of T_2 relaxation in the rotating reference frame	50
1.32	Plots demonstrating the signals generated by two states A and B, of populations $P_A=75\%$ and $P_B=25\%$, undergoing various rates of exchange, k_{ex}	52
1.33	Example pulses from the CPMG sequence to record $R_{2,Eff}$	53
1.34	Example Plot of the measured $R_{2,Eff}$ value with increasing numbers of CPMG pulses	53
1.35	Example of the spectra for a probe undergoing slow exchange as experimental exchange time is increased during the ZZ-Exchange experiment . .	54
2.1	Diagrams of the constructs used to express <i>Bs</i> -CSP and <i>PB6</i> -CSP monomers	62

2.2	Crystal structure of <i>Bs</i> -CSP-dT7 complex PDB code 2ES2	64
2.3	Agarose gel of the products of the PCR reaction amplifying the <i>PB6</i> -CSP gene	66
2.4	Agarose gel of the products of the digestion with enzymes EcoRI and BamHI	67
2.5	Agarose gel of the products of the digestion with enzymes EcoRI and BamHI of plasmid with the <i>PB6</i> -CSP inserted into the pMal-c5x vector .	68
2.6	SDS-PAGE of the lysate used in growing double labelled MBP- <i>PB6</i> -CSP	70
2.7	Diagram of the chemistry in the Ni-affinity column	72
2.8	Diagram of the Anion Exchange Column	73
2.9	Diagram of the Size Exclusion Column	73
2.10	Chromatogram and SDS-PAGE of the Ni-Affinity Chromatography of the (I27- <i>Bs</i> CSP) ₃ -I27	76
2.11	Chromatogram and SDS-PAGE of the Anion Exchange Chromatography of the (I27- <i>Bs</i> CSP) ₃ -I27	77
2.12	Chromatogram and SDS-PAGE of the Size Exclusion Chromatography of (I27- <i>Bs</i> CSP) ₃ -I27	78
2.13	Chromatogram and SDS-PAGE of the Ni Affinity Chromatography of the <i>Bs</i> -CSP Monomer	79
2.14	Chromatogram and SDS-PAGE of the Anion Exchange Chromatography of the <i>Bs</i> -CSP Monomer	80
2.15	Chromatogram and SDS-PAGE of the Size Exclusion Chromatography of the <i>Bs</i> -CSP Monomer	81
2.16	Chromatogram and SDS-PAGE of the Ni Affinity Chromatography of the MBP- <i>PB6</i> -CSP construct	82
2.17	Chromatogram and SDS-PAGE of the Anion Exchange Chromatography of the MBP- <i>PB6</i> -CSP construct	84
2.18	Chromatogram and SDS-PAGE of the Size Exclusion Chromatography of the <i>PB6</i> -CSP Monomer	85
2.19	Diagram of the MicroScale Thermophoresis Experimental Set-up	90
2.20	Power Spectrum used in Calibration of AFM Cantilevers	92

2.21	Measurement of the Relation between Voltage and Extension in AFM Cantilever Calibration	93
2.22	Diagram of the Atomic Force Microscope Set-Up for the Polyprotein Un- folding Experiment	94
2.23	Force Extension Trace for a Protein Unfolding	96
2.24	Force Extension Trace for the Unbinding of a Protein and a Ligand by SMFS	101
2.25	Peak Integration of Two Overlapped Peaks	103
3.1	(I27- <i>BsCSP</i>) ₃ -I27 construct and binding isotherm from the Tryptophan Fluorescence Quenching Titration of the polyprotein with CT2 ssDNA . .	107
3.2	Force Extension trace of the (I27- <i>BsCSP</i>) ₃ -I27 with fits to the WLC . . .	109
3.3	Scatter plots of the Unfolding force against the inter-peak distance for three FX experiments performed on the (I27- <i>BsCSP</i>) ₃ -I27 at 600 nms ⁻¹ and at 23 °C	110
3.4	Histograms of the unfolding forces and inter-peak distances of (I27- <i>BsCSP</i>) ₃ -I27 unfolding events from three experiments, measured at 23 °C and with a pulling speed of 600 nms ⁻¹	111
3.5	Histogram showing all of the unfolding forces measured in <i>Bs</i> -CSP un- folding events at 23 °C with a pulling speed of 600 nms ⁻¹	112
3.6	Force Extension trace of the (I27- <i>BsCSP</i>) ₃ -I27 in the presence of CT2 ssDNA with WLC fits	114
3.7	Comparison of the median <i>Bs</i> -CSP and I27 unfolding forces obtained from three experiments in the absence of ssDNA and in the presence of dA7 , CT2 and dT7 ssDNA oligonucleotides	116
3.8	Histogram showing all of the unfolding forces measured in <i>Bs</i> -CSP un- folding events at 23 °C with a pulling speed of 600 nms ⁻¹ in the presence and absence of CT2 ssDNA	116
3.9	Example FX traces recorded at 100, 200, 600 and 2000 nms ⁻¹ at 23 °C both in the absence and presence of the CT2 ssDNA oligonucleotide . . .	117

3.10	Pulling speed dependence of the unfolding force recorded for the (I27- <i>BsCSP</i>) ₃ -I27 at 23 °C	118
3.11	Example FX traces recorded at 5 °C at 100, 200, 600 and 2000 nms ⁻¹ both in the absence and presence of CT2 ssDNA	120
3.12	Pulling speed dependence of the unfolding force recorded for the (I27- <i>BsCSP</i>) ₃ -I27 at 5 °C	121
3.13	Histograms of the pooled unfolding forces of the <i>Bs</i> -CSP in all conditions, comparing F_U in the absence and presence of CT2 ssDNA	124
3.14	The median unfolding forces of the <i>Bs</i> -CSP in all conditions plotted against the natural log of the pulling speed	125
3.15	Underlying energy landscapes calculated from experiments at 23 °C and 5 °C	126
3.16	Underlying energy landscapes calculated for the free <i>Bs</i> -CSP and the <i>Bs</i> - CSP complexed with CT2 ssDNA	126
4.1	Fluorescence Emission Spectra of the <i>Bs</i> -CSP Before and After Quenching Titration with dT7	133
4.2	Fluorescence Emission Spectra of the <i>PB6</i> -CSP Before and After Quench- ing Titration with dT7	134
4.3	Tryptophan Fluorescence Quenching Titration of the <i>Bs</i> -CSP with dT7 at 10 °C, 20 °C and 30 °C. With fits to the tight binding equation	135
4.4	Tryptophan Fluorescence Quenching Titration of the <i>PB6</i> -CSP with dT7 at 10 °C, 20 °C and 30 °C. With fits to the tight binding equation	136
4.5	Temperature dependence of the dissociation constants, K_D , for the pro- teins <i>Bs</i> -CSP and <i>PB6</i> -CSP with the oligonucleotide dT7 as measured by the Tryptophan Fluorescence Quenching Titrations	137
4.6	ITC of <i>Bs</i> -CSP with CT2 ssDNA at 25 °C	139
4.7	ITC of <i>PB6</i> -CSP with dT7 ssDNA at 25 °C	140
4.8	ITC of <i>PB6</i> -CSP with dT7 ssDNA at 10 °C	141

4.9	Temperature Dependence of K_{DS} of <i>Bs</i> -CSP and <i>PB6</i> -CSP with ssDNA oligonucleotides, calculated from the thermodynamic parameters measured by ITC	142
4.10	Fractional change in fluorescence measured in MST titration of Fluorescently labelled dT7 with <i>Bs</i> -CSP	144
4.11	Tryptophan Fluorescence Quenching Titration of the Cys- <i>Bs</i> -CSP with dT7 at 15 °C and 25 °C	146
4.12	Example traces from SMFS experiments of the unbinding of the <i>Bs</i> -CSP-dT7 complex carried out at 600nms ⁻¹	148
4.13	Forces and Contour Lengths of Peaks measured in SMFS experiments of the Unbinding of the <i>Bs</i> CSP-dT7 Complex carried out at 600nms ⁻¹	149
4.14	Histograms of Forces and Contour Lengths of Peaks measured in SMFS experiments of the Unbinding of the <i>Bs</i> CSP-dT7 Complex carried out at 600nms ⁻¹	150
4.15	Binding Models	153
5.1	Crystal structure and Sequence of the <i>Bs</i> -CSP , pdb code 1CSP, with Side Chains of Residues that Interact Directly with dT6 ssDNA	156
5.2	Predicted Structure and Sequence of the <i>PB6</i> -CSP with conserved Side Chains of Residues that Interact Directly with dT6 ssDNA	156
5.3	Monitoring of <i>PB6</i> -CSP Polydispersity by SECMAALS	159
5.4	Temperature ramp of the <i>PB6</i> -CSP measured by far UV-CD spectroscopy	160
5.5	HSQC spectra of the <i>PB6</i> -CSP at 5 °C, 10 °C, 20 °C and 37 °C	162
5.6	Zoomed in Regions of the HSQC Spectra of the <i>PB6</i> -CSP	162
5.7	HSQC spectra of the <i>Bs</i> -CSP at 10 °C, 20 °C and 37 °C	163
5.8	Zoomed in Regions of the HSQC Spectra of the <i>Bs</i> -CSP	164
5.9	HSQC of the <i>Bs</i> -CSP with assignments, measured at 20 °C	165
5.10	Backbone Identification Experiments	166
5.11	HSQC of the <i>PB6</i> -CSP with assignments, measured at 20 °C	167
5.12	Backbone Identification Spectra for Unassigned Peaks	169

5.13 Comparison of HSQC spectra of the <i>PB6</i> -CSP at 5 °C and after heat denaturation	170
5.14 Overlaid HSQC spectra of the Tryptophan side chain of <i>PB6</i> -CSP in the Folded, Heat Denatured, and ssDNA Bound State	170
5.15 ZZ-exchange spectra of the identified cross-peaks in the <i>PB6</i> -CSP	171
5.16 ZZ-exchange spectra of the Tryptophan Side Chain Signal	172
5.17 Location on the predicted <i>PB6</i> -CSP structure of the residues identified as undergoing slow exchange by zz-exchange experiments	173
5.18 The ^{15}N - ^1H -NOE ratios for the <i>PB6</i> -CSP measured at 10 and 20 °C. The <i>Bs</i> -CSP measured at 10, 20 and 29 °C	175
5.19 R_2 rates for the <i>PB6</i> -CSP measured at 10 and 20 °C. The <i>Bs</i> -CSP measured at 10, 20 and 29 °C	177
5.20 R_2 rates for the <i>PB6</i> -CSP measured at 20 °C plotted against R_2 rates measured at 10 °C	178
5.21 R_2 rates for the <i>Bs</i> -CSP measured at 20 °C plotted against R_2 rates measured at 10 °C	179
5.22 R_2 rates for the <i>Bs</i> -CSP measured at 29 °C plotted against R_2 rates measured at 20 °C	180
5.23 The calculated R_2 values for the <i>PB6</i> -CSP from the fits to CPMG dispersion curves	181
5.24 Example relaxation dispersion curves with fits at two B_0 fields	182
5.25 Values measured at 10 °C of k_{ex} for the CPMG measurements on the <i>PB6</i> -CSP	183
5.26 Values of State populations from the CPMG Measurements on the <i>PB6</i> -CSP	184
5.27 Model of <i>Bs</i> -CSP dynamics at (a) 10, (b) 20 and (c) 29 °C represented by energy landscape diagrams	186
5.28 Model of <i>PB6</i> -CSP dynamics at (a) 10 and (b) 20 °C represented by energy landscape diagrams	186

6.1	Schematic of the effect of ssDNA binding on the transition state of unfolding when the <i>Bs</i> -CSP is under mechanical force	191
6.2	Schematic of the effect of reduced temperature on the dynamics of the <i>PB6</i> -CSP	195
6.3	Representations of how the low temperature exchange with a second state might influence the <i>PB6</i> -CSP binding to nucleic acids	196
A.1	CD scan of the <i>Cys-Bs</i> -CSP at room temperature to confirm folding of the variant.	200
A.2	CD scans of the <i>PB6</i> -CSP during the temperature ramp	201
B.1	Histograms of all the unfolding forces recorded at room temperature on the (I27- <i>Bs</i> CSP) ₃ -I27 polyprotein.	203
B.2	Histograms of all the unfolding forces recorded at 5 °C on the (I27- <i>Bs</i> CSP) ₃ -I27 polyprotein.	204
B.3	Raw data of Fluorescence signal measured in MST titration of Fluorescently labelled dT7 with <i>Bs</i> -CSP	205
B.4	Fractional change in fluorescence measured in MST titration of Fluorescently labelled dT7 with <i>Bs</i> -CSP at various IR laser strengths	205
B.5	First 15 relaxation curves	206
B.5	Second group of 15 relaxation curves	207
B.5	Third group of 15 relaxation curves	208
B.5	Relaxation curves and R_{ex} Fits (black) from the CPMG experiment on the <i>PB6</i> -CSP at 10 °C measured at 600 (blue) and 750 (green) MHz field strengths.	209

1

Introduction

Since life began on earth, the planet has undergone drastic changes in environment, and surviving organisms have had to adapt^[1]. Currently the most common environments are at low temperature, with 90 % of oceans on earth below 5 °C, and 80 % of the overall biosphere permanently cold^[2]. Organisms have successfully colonised the majority of these regions, with measured metabolic activity at -20 °C measured in organisms from the Siberian permafrost^[3]. Microbial communities have been found in subglacial outflows^[4] and sampled directly from subglacial Antarctic lakes^[5], where whole ecosystems exist in dark, cold and salty environments. Because of this, psychrophilic (cold-loving) organisms are one of the best models we have to explore potential life on other planets in our solar system^[6].

The work described in this thesis explores the molecular mechanisms involved in the adaptation of organisms to the cold, with a focus on the preservation of flexibility in biomolecules at low temperatures. The Cold Shock Protein (CSP) and its interactions with nucleic acids are employed as a model system for this objective, building on previous work with these proteins^[7].

This chapter begins with a brief description of some of the important properties of biomolecules, how they are affected by temperature and an introduction to the organisms, termed extremophiles, that have adapted to live in environments that were considered too hostile for life before the discovery of thermophiles at Yellowstone in the 1960s^[8]. This is followed by an overview of the literature surrounding extremophilic proteins and

some of the techniques employed to investigate adaptations of proteins to extremely cold environments.

1.1 Biomolecules

Living organisms known on earth are made up of cells, which have been called “The Building Blocks of Life”. Cells in turn are built up from a mixture of biomolecules each with its own role in helping the cell survive, grow and reproduce. These include lipids self assembling to form protective membrane boundaries, cytoskeletons acting as a scaffold to give cells shape and volume, nucleic acids in the form of DNA and RNA storing the genetic information and an array of bionanomachines, the proteins.

1.1.1 Nucleic Acids

Genetic information is stored and processed by nucleic acids. These flexible molecules carry information that informs all of life’s processes from one generation to the next.

1.1.1.1 DNA

DeoxyriboNucleic Acid (DNA) is a chain made up of nucleotides with a sugar-phosphate backbone. The four nucleotides in DNA, Adenine (A), Thymine (T), Cytosine (C) and Guanine (G), make up a four letter code in which all the information required for an organism to grow and live is written. This includes the code for the amino acid sequence of every protein in the organism. The section of DNA that encodes a protein is called a gene. DNA forms a double stranded, helical structure, as each nucleotide bonds to a partner through hydrogen bonding, forming a pair (figure 1.1a). A and T form base pairs containing two hydrogen bonds, while C and G form pairs with three, leading to a higher melting temperature. The double strand stabilises the molecule, protecting the information from damage which might cause a change in the genetic code, called a mutation.

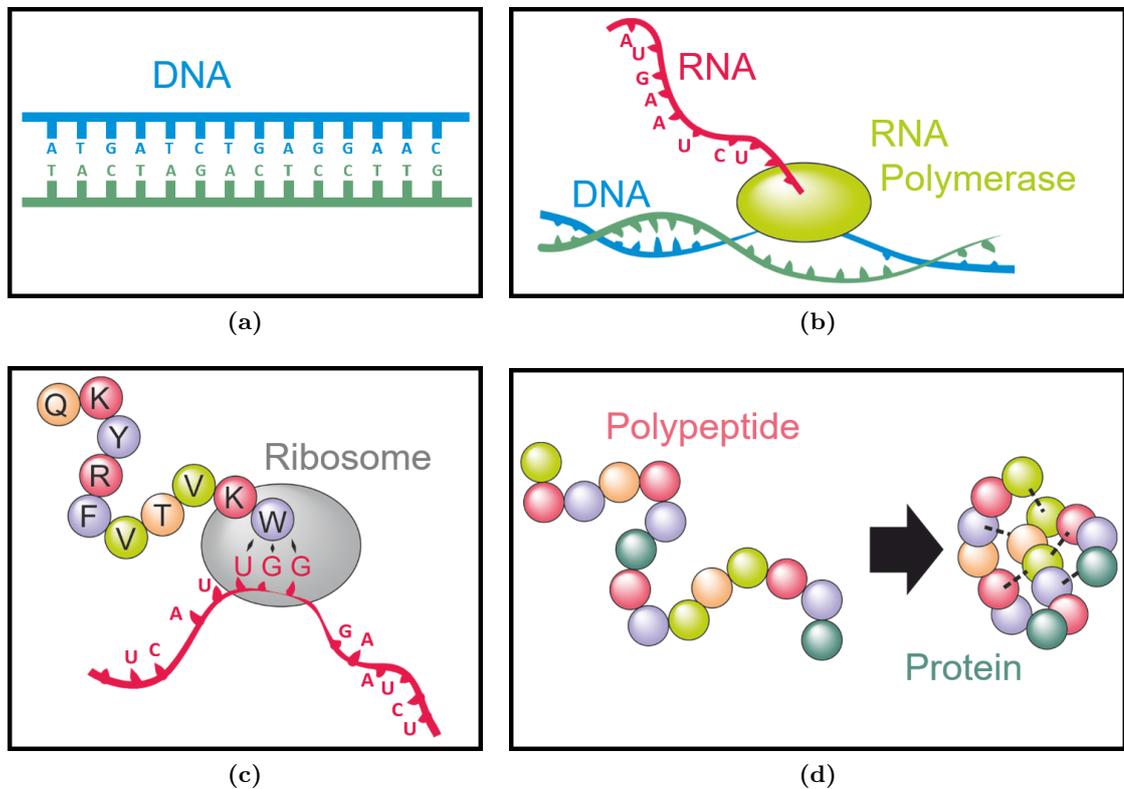


Figure 1.1: Representation of the production of a protein from the genetic code written in the DNA, or the central dogma of molecular biology. (a) The order of nucleotides A, T, C and G form the code in the double stranded DNA structure. (b) When this code is being read for transcription, the double strand is split and one of the strands is used to form RNA. (c) This messenger RNA is used by the ribosome for translation, using the RNA code to sequentially join amino acids into a polypeptide chain, (d) which then folds into the protein 3D structure.

1.1.1.2 RNA

RiboNucleic Acid (RNA) carries a 4 letter code just like DNA but the different sugar base makes it more flexible, and the thymine base is replaced by Uracil (U). It is usually found in the single stranded form and can fold into structures that can perform some catalyst reactions. This is part of the reason RNA is considered to have possibly formed the basis of the original self replication in early life^[9]. One role of RNA is to carry the genetic information from the DNA to the ribosome (figure 1.1 b and c), where the

protein polypeptide chain is produced. The versatility of RNA allows it to play many other roles in the cell, not all of which are fully understood.

1.1.2 Proteins

Proteins come in a huge variety of shapes^[10], as they carry out an equally huge range of tasks to maintain life. These 3D structures are formed from polypeptide chains made up of the 21 naturally occurring amino acids^[11], arranged in order according to code written in the DNA (figure 1.1 d).

All amino acids have the same base of a carboxyl group and an amino group, which bond to each other in the ribosome to form a chain (figure 1.2). The polypeptide chain thus formed is called the primary structure of the protein. The shape and chemistry of

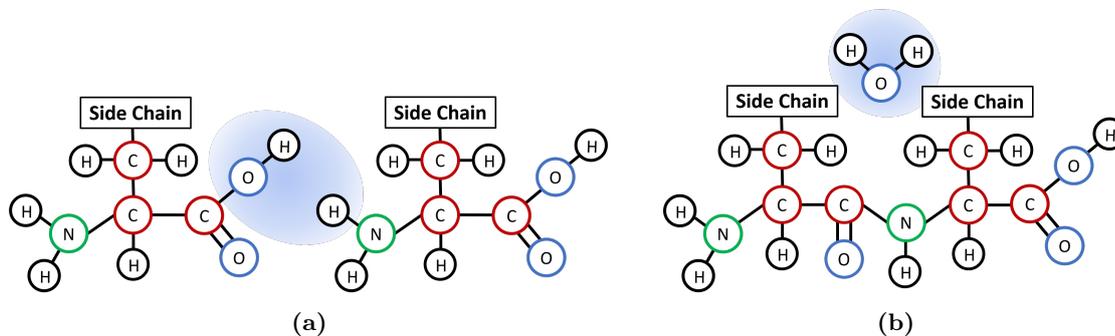


Figure 1.2: Formation of the peptide bond that makes up the backbone of proteins. (a) Two amino acids are brought together in the ribosome and (b) a peptide bond is formed through a condensation reaction, releasing a molecule of water.

each amino acid side chain determines the characteristics that influence the shape of the protein. By attracting or repelling each other, or the solvent surroundings, the amino acid make up of the protein determines the protein shape and behaviour. The 21 most common amino acids are shown in figure 1.3.

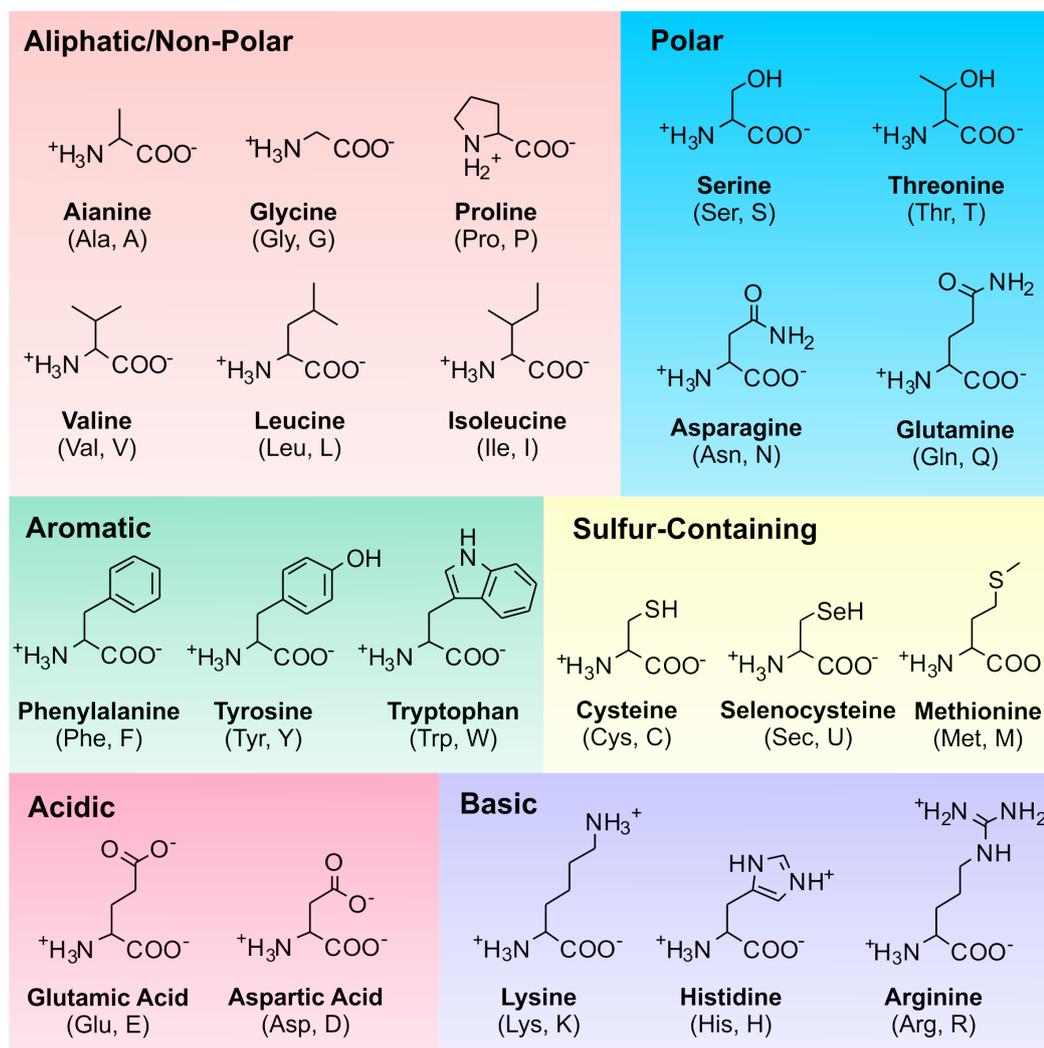


Figure 1.3: Amino acid side chain structures, labels and properties. Non-polar (orange) and aromatic (green) side chains are hydrophobic, but polar (blue) are hydrophilic. Sulfur containing side chains (yellow) have the ability to form cysteine bonds, which can be used to stabilise protein structures as it is a strong covalent bond. Acidic (red) and basic (purple) residues interact via electrostatics.

1.1.3 Protein Folding

The electric charge, size, shape, polarity or hydrophobicity of the amino acid side chain leads to forces of attraction or repulsion, demonstrated in figure 1.4. Charged residues are attracted to oppositely charged residues but repulsed by like charge. The shape of some side chains reduces the flexibility of the peptide backbone while glycine, containing only a hydrogen in the side chain position, results in a much more flexible region. Polar

regions will be attracted to each other and to the polar water molecules in the solvent. Hydrophobic residues are pushed toward the centre of the protein structure, away from the solvent and this leads to a hydrophobic core. π stacking is another interaction between aromatic rings^[12], which are also hydrophobic.

The balance of these interactions will push and pull the linear chain into a structure

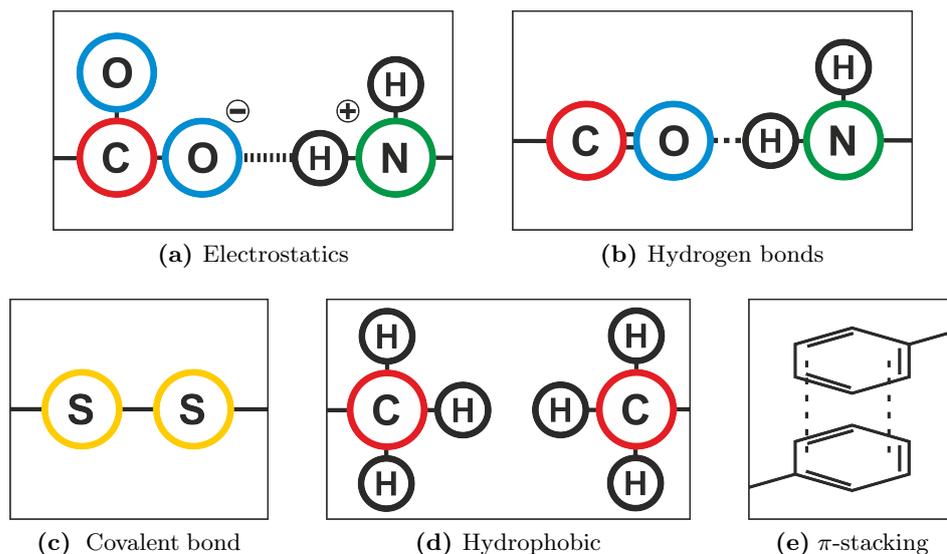


Figure 1.4: Interactions between protein residues (a) Electrostatic Interactions between acidic and basic amino acid side chains (b) Hydrogen bonds between an oxygen and hydrogen (c) Covalent bond between two sulfur atoms (d) Hydrophobic interaction as residues avoid exposure to the solvent (e) π -stacking interaction between aromatic rings.

that minimises the overall free energy of the system. Features called α -helices and β -sheets, make up the secondary structure (figure 1.5) which usually forms spontaneously. α -helices are generally quite flexible, while β -sheets form stable structures made up of β -strands arranged either parallel to one another, or anti parallel.

These features of the secondary structure are then folded into the tertiary structure of the protein, which is usually represented as a combination of α -helices and β -sheets with loop regions connecting them. Example 3D protein structures are shown in this representation in figure 1.6. The titin domain I27 is made up only of β -sheets (figure 1.6a) while the Maltose Binding Protein (MBP) structure is mostly α -helical.

Quaternary structure describes proteins with structures that include multiple domains. These can be formed from separate polypeptide chains coming together or can fold from

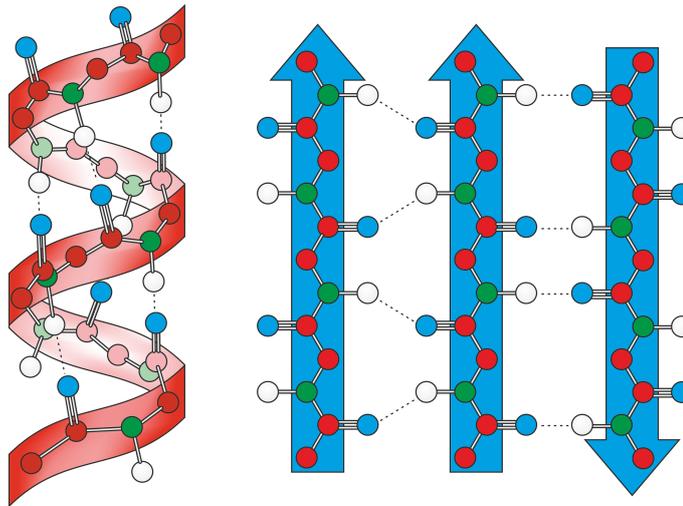


Figure 1.5: Secondary structures formed by folding of the polypeptide chain. An α -helix (red) is one feature that forms and a β -sheet (blue) is the second, which is formed in either parallel (left) or anti-parallel (right) orientation. Hydrogen bonds between the oxygen and hydrogen atoms of the peptide backbone are shown (dashed lines)

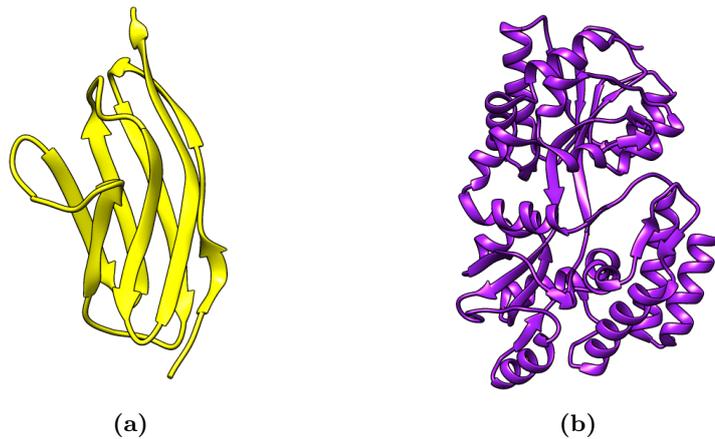


Figure 1.6: Example protein tertiary structures. (a) The 3D structure of I27 (PDB code: 1TIT) and (b) the 3D structure of MBP (PDB code: 1N3X)^[13,14]

a single long chain with different folded regions connected by linker regions.

1.1.4 Folding Thermodynamics

The process of protein folding from an unfolded (U), disordered 2D chain to a folded (F), defined 3D structure can be considered a two state system. This process, if reversible,

results in an equilibrium as shown in equation 1.1.



The available free energy of the system is the Gibbs free energy (G). The difference in this energy between the folded and unfolded state ($\Delta G = G_U - G_F$) determines the stability of the protein folded state, with a larger negative ΔG describing a folded state that is more energetically favourable. This available energy is a combination of the enthalpic (ΔH) and the entropic (ΔS) energy difference as given by equation 1.2.

$$\Delta G = \Delta H - T\Delta S \quad (1.2)$$

Thermodynamically, entropy is defined by $\Delta S = \frac{q_{REV}}{T}$, where q_{REV} is the heat transfer between the two states and T is the temperature. This change describes the energy dispersed during the process of folding or unfolding. The entropic component accounts for the ordering needed for the protein to fold. In terms of entropy the lowest energy state is an unfolded polypeptide chain, and without any other forces the protein would not fold. As temperature is increased, the entropic component becomes dominant, as shown in figure 1.7, and the unfolded state becomes more energetically favourable, leading to heat denaturing or melting of the protein.

The enthalpy is defined thermodynamically as $\Delta H = \Delta U - \Delta(PV)$ where U is the internal energy, P is the internal pressure and V is the internal volume of the system. In biological systems this is usually equated with the internal energy as the pressure and volume are negligible. The interactions between the amino acid side chains provide the enthalpic, chemical component of the free energy.

Unlike most ordered molecular systems, proteins usually have a temperature of maximum stability. The plot of free energy shown in figure 1.7 shows an optimal temperature at which the free energy difference is largest, where the protein is most stable.

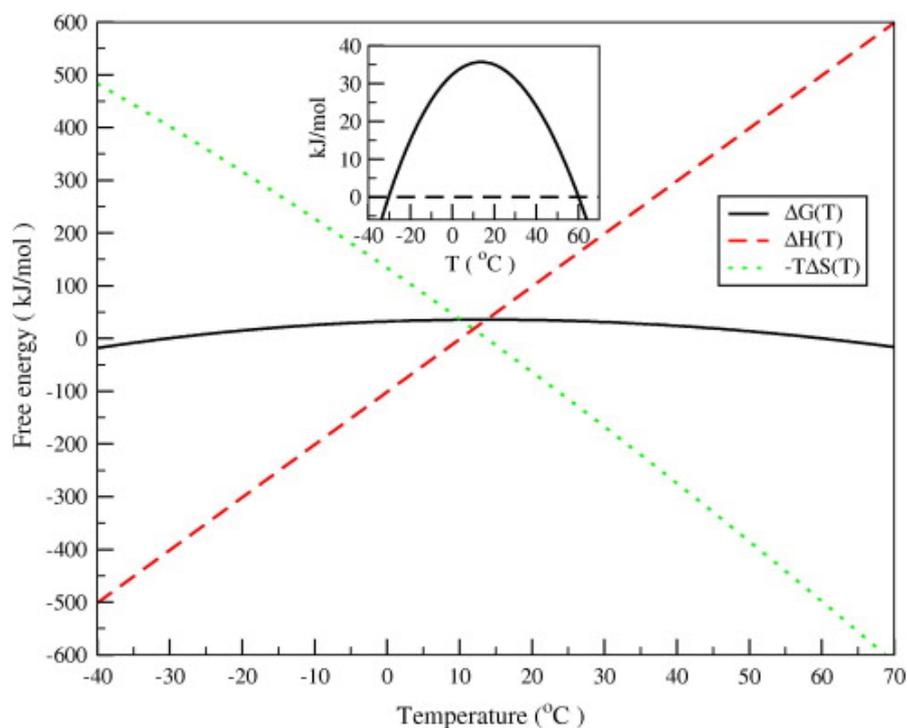


Figure 1.7: Temperature Dependence of the thermodynamic parameters free energy $\Delta G(T)$, enthalpic energy $\Delta H(T)$ and entropic energy $-T\Delta S(T)$ for the folding of a typical protein. Inset shows an enlarged Gibbs free energy plot. The cold and hot denaturation temperatures are where the free energy is reduced to zero, and energies of enthalpy and entropy are equal. Taken from Dias 2010^[15]

1.1.4.1 Cold Denaturation

As there is a temperature of maximum stability, decreasing the temperature below this point causes the protein stability to decrease, leading eventually to cold denaturation^[16]. For most proteins the temperature at which cold denaturation is predicted to occur is below the freezing point of water^[15]. Here the protein stability breaks down as all the amino acids are compacted and the balance of forces between them is disrupted. In particular, the hydrophobic force (figure 1.4d) is disrupted^[15]. This force is a result of the high entropic cost of water molecules becoming more ordered at the hydrophobic surface. This makes the hydrophobic core a consistent feature of proteins, but as temperatures decrease, the reordering of water molecules surrounding the protein reduces this effect. The hydrophobic force is generally strongest around room temperature and is weakened by both increasing and decreasing temperatures, making it an important

feature of protein thermal stability^[17]. Experimentally, cold induced unfolding is not always observed. Some proteins have been measured becoming kinetically trapped in the state of their native fold while other proteins were measured to be more compacted at low temperatures^[18].

1.1.4.2 Protein Folding Energy Landscapes

This minimisation of energy is described by the protein energy landscape, shown in figure 1.8. In a particular set of conditions, the top of the energy landscape is the free energy of

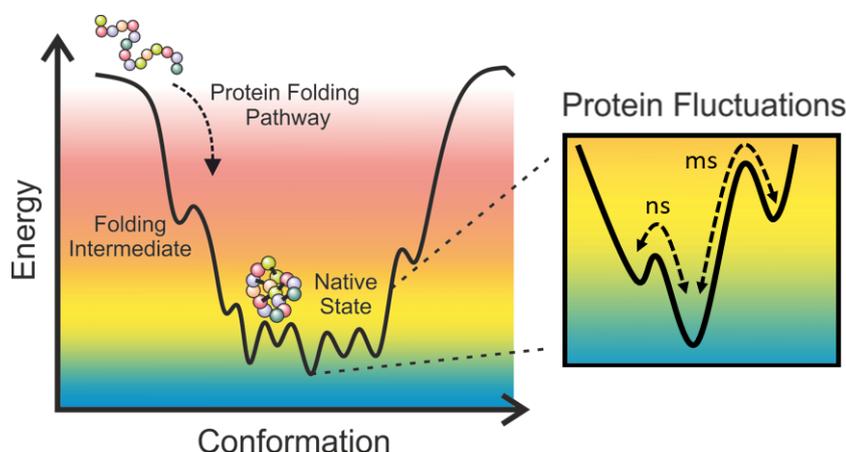


Figure 1.8: The process of protein folding described by an energy landscape. An unfolded protein is in a high energy state as the forces between amino acid side chains are not optimised. This pushes the polypeptide chain down the energy funnel, towards a lower energy state where the forces are balanced. Protein fluctuations can also be described as movement between states as shown in the expanded picture. The height of the energy barrier between the energy minima will determine the rate of exchange (represented here by ns or ms time-scales) between the two states.

the unfolded protein and the bottom of the funnel is the fully folded protein. The height of the funnel is the ΔG of folding as represented in figure 1.7, so the funnel is deepest at the temperature that the protein is most stable. The process of a protein structure being formed from the polypeptide chain can be described as a protein journey down the slope of the energy funnel. Each possible conformation of the protein has an energy associated with it, and the final folded structure is usually where this is at a minimum. This schematic representation provides a convenient description encompassing all of the forces behind protein folding and structure. The structure of each protein has been honed

during the evolution of the organism, and over time each protein will find a structure in an energy minimum that performs the desired function. The landscape will appear different for each protein, and can be affected by the surrounding physical and chemical environment. Some proteins do not have a smooth journey down the energy funnel, and can get stuck in smaller energy minima in a conformation that is not functional. Protein chaperones act as catalysts, helping the proteins reach the desired structure by reducing energy barriers on the slope.

1.1.4.3 RNA folding

As well as proteins, RNA molecules can fold into structures that help RNA in function and regulation. Because RNA is more flexible than DNA, the formation of base pairs between complementary nucleotides can result in a large variety of shapes. These can include loops, where the nucleotides are exposed, and hairpin regions (figure 1.9), where the nucleotides are paired up, and inaccessible to potential binding partners. In the

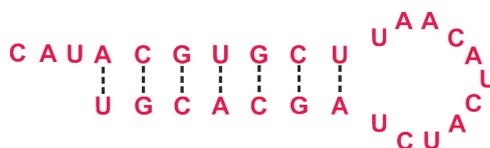


Figure 1.9: Example of an RNA stem-loop or hairpin structure

same way that proteins can misfold, RNA can form structures that are unfavourable to the RNA function^[19]. This can be caused by external stresses such as solvent conditions and temperature, or be due to long strands of RNA finding an energy minimum as it explores the conformational landscape of long polypeptide chains. Misfolded RNA will not function in the cell and could cause major problems, so RNA chaperones aid folding by binding RNA and making the misfolded states energetically unfavourable^[20].

1.1.4.4 Rates

Binding and folding processes are governed by the relative energies of the states involved, whether that is folded and unfolded states, or free and bound states. The rate at which

a system will change from a high energy state to a lower energy state depends on the energy barrier in between them. This corresponds to the energy of the transition state (TS) corresponding to the state the system would have to pass through in order to get to the lower energy state. This is represented in figure 1.10. The rate at which molecules

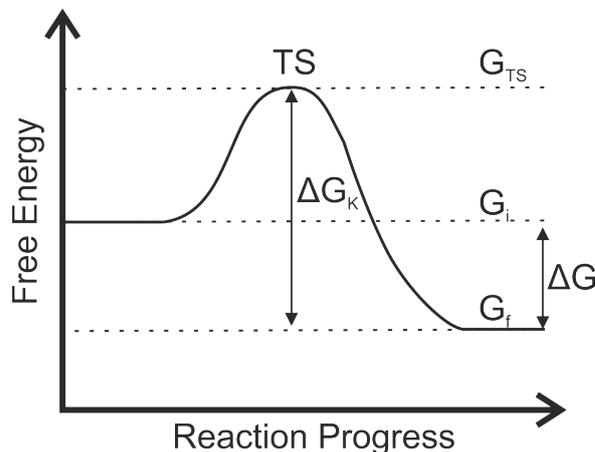


Figure 1.10: Energy diagram of the process of changing states. The energy barrier between the high energy starting state, i, and the low energy final state, f is the difference between the energy of the transition state (TS) and the final state, $G_{TS} - G_f$.

change from a high energy state to a low energy state is described by the Arrhenius equation.

$$k_R = A e^{-\frac{\Delta G_K}{RT}} \quad (1.3)$$

This describes the transition rate, k_R , between two states where the activation energy is ΔG_K , A is the attempt frequency, R is the gas constant and T is the temperature.

1.1.5 Dynamics and Function

Biomolecules need to be flexible in order to function. DNA needs to be flexible so that it can be stored compactly in the cell and read by proteins. Too flexible and DNA is susceptible to damage and tangling, but too rigid and it will be useless to the organism. At the scale of proteins and nucleic acids, the nanoscale, thermal fluctuations dominate. Thousands of proteins in each cell have their own very specific task to carry out which usually involves binding to some other biomolecule. An inflexible, entirely rigid protein would need to be thrown by random fluctuations into the perfect position and rotation

with its binding partner before it could carry out this role. Flexibility allows proteins to use thermal energy to sample a range of conformations (right hand inset of figure 1.8) which helps them search for binding sites. When two binding partners come together a subset of the protein conformations are stabilised in the bound state. This adjustment of protein conformations can trigger a change in protein activity. Activity that is dependent on ligand binding is a mechanism for protein control in the cell. The protein dynamics are key to this process, as the free protein samples all of the active conformations^[21].

1.1.6 Protein-Ligand binding

The role of many proteins involves the binding of some other molecule termed a ligand, which can be DNA, RNA, small molecules or other proteins. This can be part of a signalling process, a regulation process or a chaperoning process where one of the binding partners assists the other in reaching the correctly folded state. The process of binding shares commonalities with the process of folding, involving many of the same forces seen in figure 1.4. Molecules are constantly diffusing around the cell, and when they come into contact with one another, if there are attractive forces between them they will bind. High affinity binding partners will have binding rates limited only by this diffusion, then the stability of the complex will determine how long they remain bound.

1.1.7 Dissociation Constant

The overall affinity between the two binding partners can be described by the dissociation constant, K_D . It can also be described as the inverse of the association constant K_A .

$$K_D = \frac{1}{K_A} = \frac{k_{off}}{k_{on}} = \frac{[P][L]}{[PL]} \quad (1.4)$$

Where k_{on} and k_{off} are the rates of binding and unbinding respectively, $[P]$ and $[L]$ are the concentrations of protein P and ligand L, and $[PL]$ is the concentration of the complex. The K_D has units of molarity which is useful as it describes the concentration of ligand, $[L]$, at which 50 % of the protein will be bound. It is also useful as this

number is usually of the same magnitude as the natural concentration of the ligand *in vivo*. Affinity that generates a K_D that is stronger than the natural ligand concentration will not be selected for during evolution, as it will not lead to any significant increase in the fraction of bound states. Similarly, if the K_D is much smaller than the natural ligand concentration then the chances of binding become so much smaller that the protein is essentially useless.

1.1.8 Environmental Stress

The energetics of biomolecules and the interactions between them are finely balanced and have been selected for during evolution in the organism's native environment. If the energy difference between two states is too large, then processes will be blocked, too small and there will be no stability and rates will be affected. These energies can be affected by a number of factors. These include ionic concentration, acidity, pressure, radiation and temperature which will all affect the folding and binding reactions that keep the organism alive. Changes in any of these conditions will affect all of the processes described here, as well as potentially damaging the genetic information stored in the DNA. Survival in fluctuating conditions requires a stress response mechanism that allows the system to acclimate to new environments.

RNA is more flexible than DNA and hence more susceptible to environmental stress. This property has been exploited by organisms, using RNA as sensor mechanisms that can adjust protein production in a responsive manner^[22-24].

1.2 Extremophiles

Life has been found to exist across earth in a wide range of seemingly extremely unfavourable conditions. From the high temperatures and pressures of deep sea thermal vents, to freezing temperatures deep below the ice sheets, to highly saline waters. Organisms that live in environments we would consider as hostile are termed extremophiles.

A major component of this adaptation is protein stability^[25]. Warm blooded organisms

maintain an internal environment at the temperature and salinity at which their proteins function best. Without this protective sheltering, the proteins must adapt at the molecular level to function in whatever environment the organism lives^[26].

Extremophiles are categorised by the optimum growth conditions of the organisms. Thermophiles are defined as organisms with optimum growth temperatures between 45 and 80 °C and hyperthermophiles organisms with optimal growth temperatures over 80 °C^[25]. These organisms would be unable to survive if their proteins unfolded in these high temperatures^[27]. Other biomolecules such as DNA and RNA are also more vulnerable to damage in these environments and must therefore be protected for species survival^[28,29]. At the other end of the temperature-scale psychrophiles are adapted to thrive at temperatures lower than 10 °C^[30], where there is reduced source of thermal energy for molecular processes.

1.2.1 Psychrophiles

Living at low temperatures between 10 °C and -20 °C, psychrophilic organisms are adapted to function with lower levels of thermal energy. This low thermal energy can lead to cold denaturation of proteins^[16], as described in section 1.1.4.1. The temperature also leads to a reduction in reaction rates, especially in enzyme catalysis which is very temperature dependent. This can be seen in equation 1.5, an expanded version of equation 1.3 for enzyme catalysis^[2,31].

$$k_{cat} = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G_K}{RT}} = \kappa \frac{k_B T}{h} e^{\frac{\Delta S_K}{R}} e^{-\frac{\Delta H_K}{RT}} \quad (1.5)$$

Here κ is the transmission coefficient generally close to 1, k_B is the Boltzmann constant (JK^{-1}), h the Planck constant (Js), R the universal gas constant ($8.31\text{JK}^{-1}\text{mol}^{-1}$), and $-\Delta G_K$ the free energy of activation or the variation of the Gibbs energy between the activated enzyme-substrate complex and the ground state ES. This temperature dependence of reaction rates is one of the reasons refrigerators are so useful, as they slow down bacterial functions. This also makes it one of the major issues that psychrophiles must counter if they are to thrive at lower temperatures (below 10 °C).

1.2.2 Psychrotrophic Organisms

While the optimal growth temperature of organisms defined as psychrophiles is below 15 °C, psychrotrophs are defined as organisms that can survive in cold environments but will grow better above 15 °C^[6,32]. This is a feature that is true for many cold-adapted microorganisms, as survival in cold environments is helped by the ability to withstand a wide range of temperatures. In the rest of this work the term psychrophiles will be used to cover both psychrophiles and psychrotrophic organisms.

For example, *Psychrobacter* sp. B6 is a Gram-negative coccobacillus bacterium found in Antarctic soils^[33]. It grows at temperatures ranging from 0-30 °C with an optimal growth temperature of 20 °C, meaning it is defined as a psychrotrophic organism. An aminotransferase enzyme from this organism has an optimum thermal activity that is 20 °C lower than the mesophilic aminotransferase^[34]. The crystal structure showed that this was not due to an increase in loop length but that an increase in the number of glycines in key hinge regions allows more flexibility.

1.2.2.1 Flexibility in Cold Adaptation

Much work has gone into understanding how cold adapted enzymes counter this exponential decrease in enzymatic rates with lower temperatures^[31,35]. Simulations have shown that low temperature catalytic rates are directly promoted by a higher surface mobility in the enzyme trypsin^[36]. In these simulations, the amino acids of the cold adapted trypsin were not changed but the motion of the protein surface was constrained. Changing this around the active site was enough to control the low temperature catalytic rate. Molecular Dynamics (MD) simulations have also been used to show that an increase in surface mobility, not necessarily near the active site, will increase low temperature catalytic activity^[37]. In a cold environment, this increase in protein flexibility will be selected for over protein thermal stability, making psychrophilic proteins an interesting route of investigation when trying to understand the balance between protein function and stability^[38]. To maintain resilience at low temperatures, cold adapted proteins need to counter the increase in viscosity. One tactic for this, would be to shift to a more

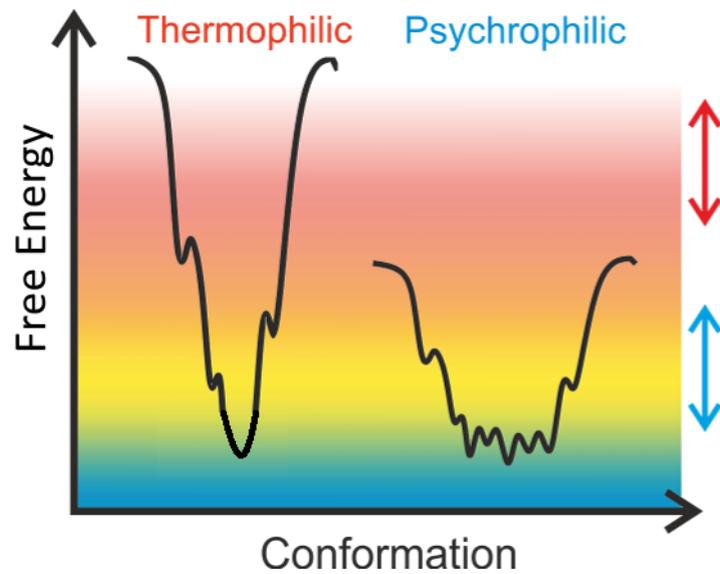


Figure 1.11: Predicted Energy landscapes for proteins from thermophiles and psychrophiles. Arrows indicate the range of energies that the majority of proteins will have at the optimum temperatures for each organism, with thermophilic proteins (red) sampling high energy states and psychrophilic proteins (blue) sampling the lower energy states.

“frustrated” state, meaning the non-covalent interactions are tuned to cause regions of the protein to repel other spatially close regions. This can be described as a “rough” energy landscape (figure 1.11), where different protein states are separated by low energy barriers and each energy well is shallow as the repulsive forces counter stabilisation. As less energy is required for exchange between these states, the protein will sample a range of conformations even if there is less thermal energy available.

The schematic energy landscapes shown in figure 1.11 represent the theoretically favourable funnels for psychrophiles and thermophiles. The thermophile has a higher energy barrier between the unfolded and folded state. This accounts for the necessary increase in melting temperature of the protein. The psychrophilic protein conformation funnel has a lower overall energy well, with more native conformational states at the low energy levels. At the optimal temperatures for each organism, the proteins are in an energy conformation at a wide part of the funnel, representing the flexibility of the protein. At the lower temperatures, the thermophilic protein is trapped in a narrow energy well, where it would likely be too rigid to function optimally. At higher temperatures the psychrophilic protein has too much thermal energy to maintain a folded structure and would denature. The evolution of these proteins has selected for adaptations that

suit the native environment of the organism.

1.2.3 Disorder in Cold Adaptation

Intrinsically Disordered Proteins (IDPs) are proteins that have a primary structure that does not lead to a defined 3D structure. Instead the energetically favourable state for these proteins is that of an unfolded or disordered state, making them less susceptible to cold denaturation^[18]. This can describe a whole protein, or just a part of the protein structure that is linked to other more ordered domains. Stress conditions can induce partial disorder in certain conditionally disordered chaperones, which is thought to aid in recognition of multiple aggregation-prone proteins^[39]. When it comes to cold adaptation, an IDP region of a cold induced protein can increase the freeze tolerance of the whole organism^[40]. This is thought to aid in maintaining membrane flexibility through lipid binding.^[41]

For low temperature catalysis of enzymes, disordered regions far from the active site have been shown to increase low temperature activity through dynamic allostery^[42]. By making mutants of the adenylate kinase that favour the unfolding of domains far from the active site, low temperature catalytic activity was obtained. It was also demonstrated that affinity and activity could be independently controlled in this manner.

1.3 The Cold Shock Protein (CSP)

To survive changing conditions, bacteria need a strategy of stress response that will protect fragile components of the cell from damage and adapt functionality to the new environment^[43]. A common source of stress is a change in temperature. Without an adequate response mechanism, any significant change in temperature would disrupt the delicate energetic balance of the cellular processes which maintain life. When there is a rapid decrease in temperature, post-transcriptional events induce an increase in a group of proteins that help adapt the molecular processes in the cell^[45]. This is known as the “Cold-Shock” response. These proteins are induced during an “acclimation” phase,

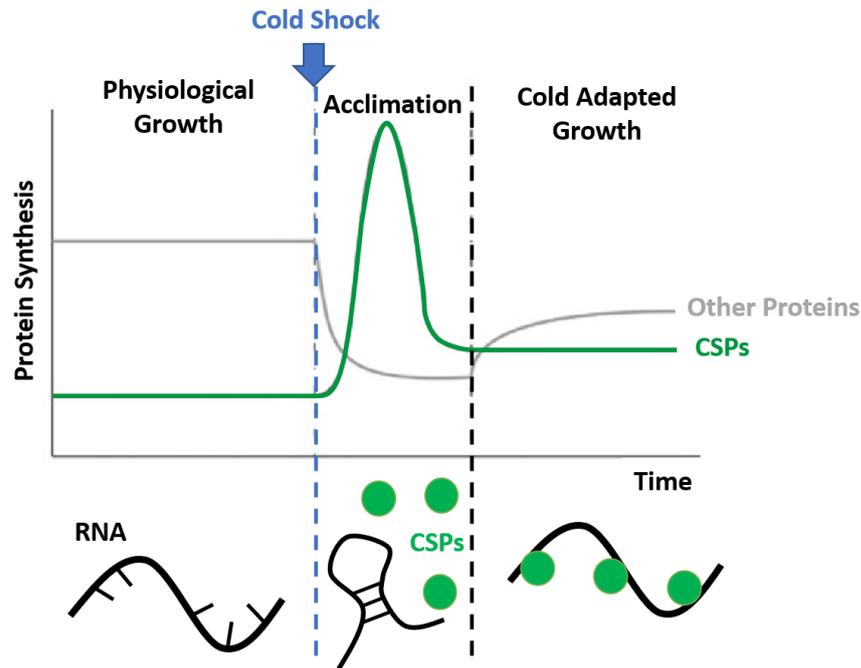


Figure 1.12: The Cold Shock response. Cold Induced Proteins, which includes the CSPs (green), are induced when organisms experience a rapid drop in temperature. These proteins help acclimate the organism to the new conditions. One role of the CSPs is to destabilise RNA secondary structures, thus allowing the translation of RNA in the new Cold-Adapted Growth phase. Taken from Horn 2007^[44]

as shown in figure 1.12, while the production of other proteins decreases rapidly, partially due to translation impairment caused by the energetic favouring of non-functioning mRNA secondary structures. After this initial response to the cold shock, global protein synthesis is restored at a lower rate, adapted to the colder temperature and this is the cold shock response^[46]. In total nine CSPs have been identified in the genome of *E. Coli*, labelled (CspA-CspI). The major CSPs induced at cold temperature are termed CspA, CspB and CspG^[47]. In *Bacillus Subtilis* three CSPs were identified as CspB, CspC and CspD. There are many homologues of CSPs, and each play roles of varying importance under different stress conditions^[48], but the mechanism of this stress response remains similar enough that they can stand in for one another and phenotype alterations are only seen after two or more CSP genes are deleted^[49].

1.3.1 Structure of the CSPs

Cold shock proteins are ubiquitous throughout prokaryotic life with highly conserved nucleic acid binding regions^[44]. Versions are found in most other life forms as well^[50], as Y-box proteins in eukaryotes^[51]. The highly conserved structure^[52] consists of 65-70 amino acids that form 5 β -strands folded into a closed barrel, called an OB-fold and shown in figure 1.13a. The crystal structure of the *Bs*-CSP is shown in figure 1.13b^[53].

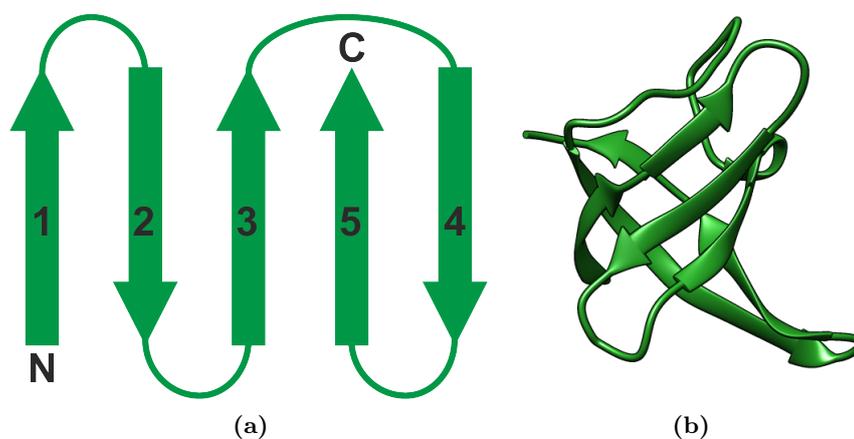


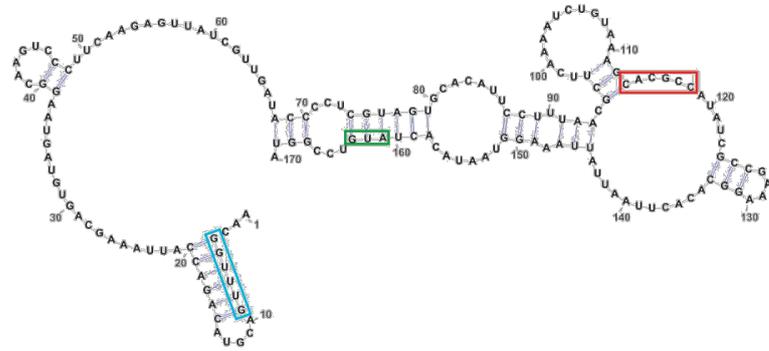
Figure 1.13: Greek Key motif of the OB-fold that is typical for CSPs. (a) The structure is made up of five β -strands, numbered from the N to the C-terminus (b) Folded into a closed barrel structure of the CSP (PDB code: 1CSP^[53]).

This domain structure is found frequently as a general nucleic acid recognition site^[54], and is present in the human Y-box binding protein^[51]. Mutations of the CspA from *E. Coli* to mimic the loop regions of the Y-box can even confer the ability to bind double stranded DNA to the bacterial CSP^[55].

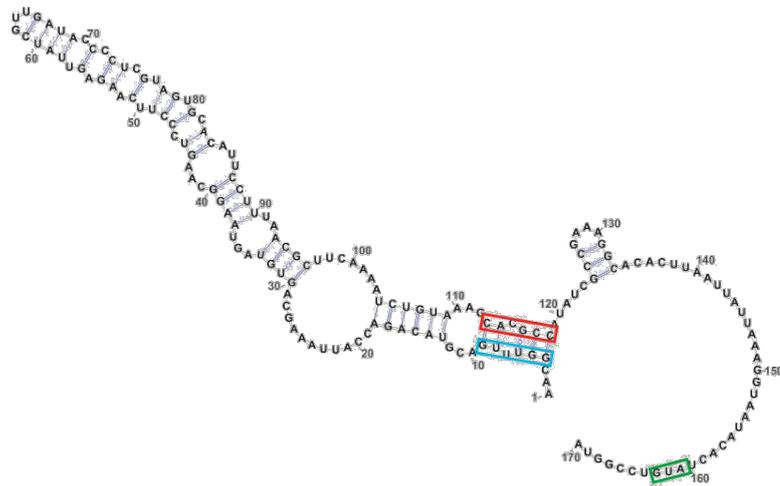
1.3.2 Induction of the CSP

In *E. Coli*, the two proteins identified as the most important to the acclimation phase are RNase R and CspA^[56]. As the decrease in temperature leads to increased misfolding of RNA into non-functional secondary structures, CspA works to unfold these structures while RNase R promotes mRNA degradation to achieve optimal Csp:mRNA ratios. The induction of CspA is caused by a temperature dependent mRNA structure^[23], as shown

in figure 1.14. The start codon, RNA code AUG that initiates the ribosome protein



(a) 37 °C



(b) 10 °C

Figure 1.14: Induction of the cold shock proteins. The RNA structures that translate to cspA at (a) 37 °C and (b) cold shock temperature, 10 °C. The boxes indicate the start codon of cspA (green), the conserved “cold box element (blue), and its long-range interaction region at 10 °C (red). Taken from Zhang 2018^[56].

production, for the CSP is more accessible to the ribosome in the mRNA structure predicted at 10 °C (figure 1.14b). Through this up-regulation of translation of CSPs and RNase R, after 3 hours of cold shock they constitute 40 % of the protein synthesis in *E. coli*. The scale of this RNA regulation component of the cold shock response is a sign that the effect of temperature on RNA is an important aspect of cold adaptation.

1.3.3 Binding of CSPs to Nucleic Acids

The structures of the *Bs*-CSP bound to ssDNA^[57] and ssRNA^[58] have been obtained from both crystallisation and in solution by NMR^[59]. This provides a very detailed picture of the interactions between the binding site of *Bs*-CSP with nucleic acid bases. These structures identified a binding site consistent for both DNA and RNA, across the hydrophobic region of the protein surface^[60]. This site was found to preferentially bind thymine nucleotides and can hold 6-7 bases^[61]. The site contains two conserved sequence motifs known as RNP-1 and RNP-2, which are sites Lys13 – Val20 and Val26 – Phe30 respectively in *Bs*-CSP^[62]. It was found in the crystal structure that this site could bind two short oligos over the two RNA recognition sites and that these oligos can link multiple proteins to create a crystal lattice^[62].

This complex structure is shown in figure 1.15 with the hydrophobic side chains shown in figure 1.15a and the side chain charge shown in figure 1.15b. The binding region has a large number of surface hydrophobic residues which bind the nucleotides through hydrophobic and pi-stacking interactions (figure 1.4). While the overall protein surface appears to have a mostly negative charge density, the binding region is positive (figure 1.15b). This will help in the binding of negatively charged nucleic acids. The amino acid side chains involved in the binding of dT6, and shown in figure 1.15a were identified as Lys 7, Trp 8, Ser 11, Glu 12, Phe 17, Asp 25, Phe 27, His 29, Phe 30, Phe 38, Lys 39, Thr 40, Arg 56, Gly 57, Pro 58, and Gln 59. Most of these residues are aromatic or Binding studies of the *Bs*-CSP interacting with nucleic acids was used to design an oligo with a very high binding affinity. By replacing the 3rd thymine of dT7 with a cytosine (so the oligo sequence is TTCTTTT), the affinity of dT7 with *Bs*-CSP at 15 °C increased by ninefold, with the K_D reducing from 1.8 ± 0.4 nM to 0.2 ± 0.07 nM, as measured by tryptophan fluorescence quenching^[62]. Figure 1.16 shows schematics representing binding both in the crystal structure and in solution based on this study.

The hydrophobic residues Phe 15, Phe 17 and Phe 27 are surface exposed hydrophobic residues that help in binding nucleic acids, but they have also been shown to stabilise the *Bs*-CSP^[60]. This makes the CSP an unusual protein in that it contains a hydrophobic surface region that stabilises rather than destabilises the protein conformation.

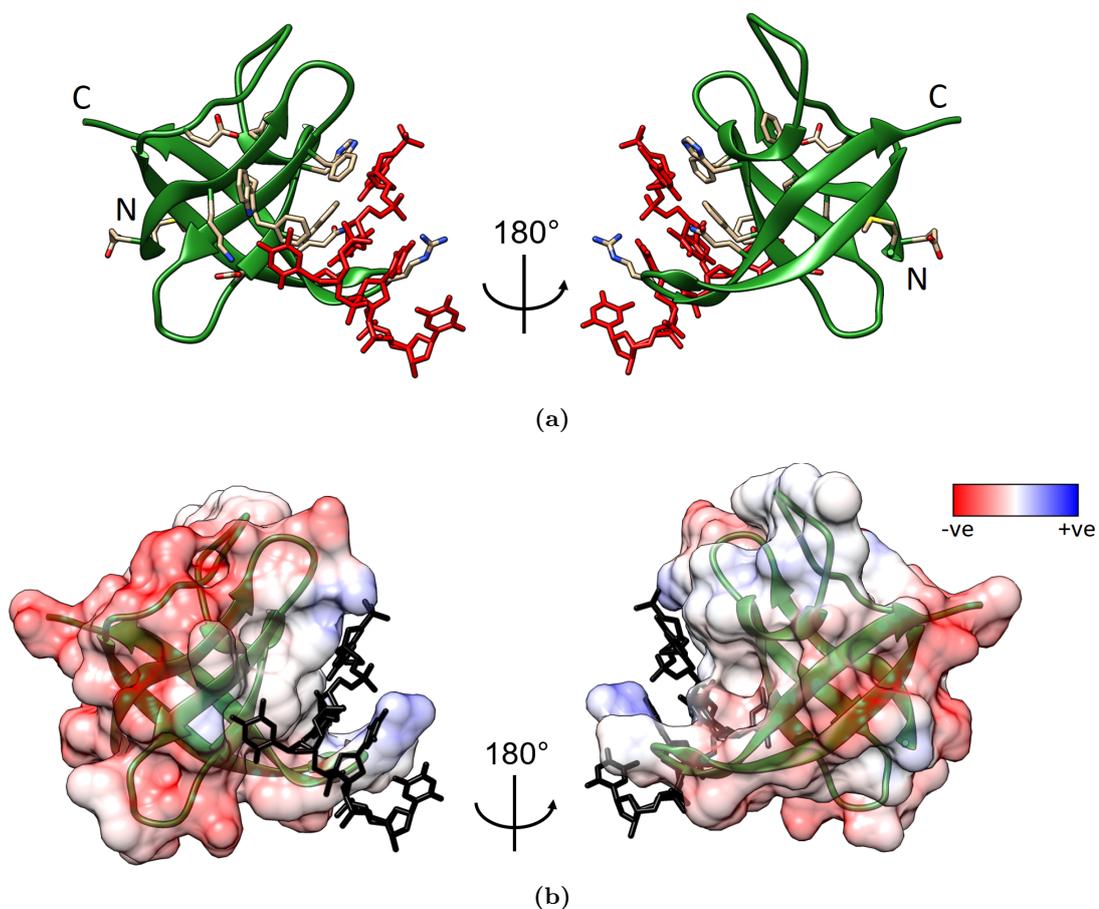


Figure 1.15: Crystal structure of the *Bs*-CSP bound to hexathymidine ssDNA oligo (dT6)^[57] (a) β -barrel structure with the side chains identified as part of the binding surface shown with the ssDNA bound (red). (b) Surface charge showing an overall more negative (red) surface charge but a positive (blue) area around the ssDNA (black) binding site.

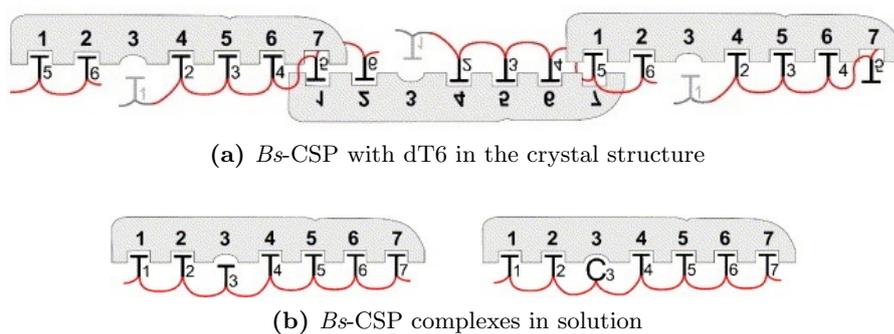


Figure 1.16: Schematic of the binding of ssDNA oligos with *Bs*-CSP in both (a) the crystal structure and (b) in solution. Taken from Max 2006^[62]

to respond to stresses on the molecular level is essential for bacteria, and it has been shown that by taking the genes of these CSPs and transferring them into wheat, this tolerance for stress can be appropriated^[64]. The modified wheat exhibits a much greater drought tolerance when the bacterial gene for CspA was incorporated, although the effect was not observed using the gene for CspB. This ability to maintain translation efficiency is useful even when it is not part of a stress response, and CSPs have been shown to be present in high concentrations, above 50 μM , during normal exponential growth conditions^[65]. The location of the proteins in the cell have been shown by fluorescent labelling to be localised to nucleoids in growing cells^[66]. Double CSP deletion mutants, with two of cspB, cspD and cspB, cspC deleted, showing abnormal nucleoid shapes, indicating CSPs as important to connecting transcription to translation initiation.

1.3.5 CSP Homologues Adapted to Extreme Environments as Model Systems

The conserved structure of CSPs make them good models to study mechanisms of adaptation to different environments. Structures have been obtained by both Nuclear Magnetic Resonance (NMR) and crystallography of CSPs from organisms adapted to a variety temperature environments^[67], including those adapted to colder environments, *Listeria monocytogenes* and *Colwellia psychrerythraea* 34H^[68,69]. The structure of a CSP from the hyperthermophile *Thermotoga Maritima* (*Tm*-CSP) has been obtained at the physiological temperature for this organism at 70 °C (343 K)^[70]. The *Tm*-CSP at this high temperature maintains the OB-fold structure.

1.4 Single Molecule Force Spectroscopy (SMFS)

The melting temperature (T_M) of a protein is the temperature at which 50% of the proteins in a sample have denatured. This gives an idea of the average stability of a protein population which is very useful for understanding differences between types of proteins. Differences between individual protein molecules however, are lost in this averaging.

Single-molecule techniques provide a new level of understanding of protein properties at the individual level, making details of the kinetic unfolding pathway accessible (figure 1.10).

A range of techniques can be applied to this goal, including optical and magnetic tweezers, FRET and Atomic Force Microscopy (AFM)^[71,72]. Each method has advantages and limitations, with optical tweezers able to precisely monitor small forces between molecules attached to two beads^[73], and magnetic tweezers that can apply both stretching and twisting forces onto molecules^[74]. The application of AFMs to single molecule studies is very versatile. It can be in the form of imaging molecules that have been deposited onto a surface with recent advances in scanning speed making it possible to watch processes unfold at this scale^[75], using tips functionalised with molecules to create force maps of a surface^[76], or using the AFM cantilever to pick up and pull on a molecule to measure intermolecular forces directly, called Single Molecule Force Spectroscopy (SMFS).

1.4.1 Models to Extract Energy Landscape Parameters from SMFS Experiments

The unfolding of proteins by applied force can be described as a tilting of the energy landscape, according to the Bell model^[77]. The effect of an applied force on the protein folding energy landscape according to the Bell model is shown in figure 1.18. The force dependent unfolding rate ($k_U(F)$) and the distance to the unfolding transition state along the reaction coordinate Δx_U are described by equation 1.6^[77].

$$K_U(F) = A \exp\left(\frac{-(\Delta G_U^* - F\Delta x_U)}{k_B T}\right) \quad (1.6)$$

F is the applied force, A is the attempt frequency, k_B is Boltzmann's constant, T is the temperature and ΔG_U^* is the transition state free energy when the force is applied as shown in figure 1.18. This incorporates the force of pulling to the reaction energies shown in figure 1.10 by applying a tilt to the energy landscape, so that the unfolded state becomes more favourable.

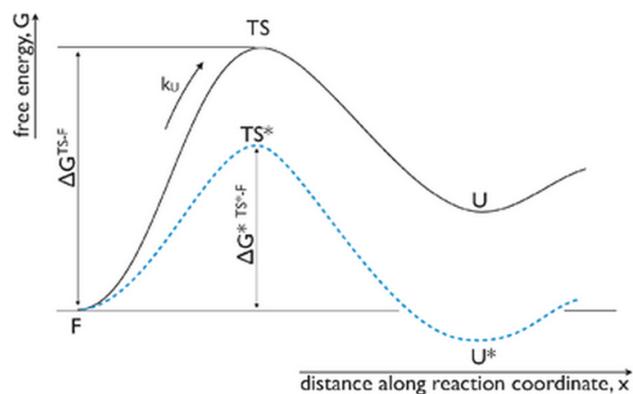


Figure 1.18: Schematic free-energy profile of a protein folding pathway leading from the native, folded state (F) to its denatured, unfolded state (U) through a transition state (TS) at an unfolding rate k_U . Applying a force tilts the energy landscape (blue dotted line), lowering the energy barrier so that the unfolded state (U^*) is more populated than the folded. Taken from Hoffmann 2012^[78]

Thermal fluctuations cause the proteins to unfold in a stochastic manner, the probability of which is described by the energy landscape. As pulling speed is increased, there is less time for this stochastic process to cause the proteins to unfold, so this leads to higher unfolding forces being recorded. The relation between unfolding force and pulling speed depends on the distance between the folded and unfolded state in the conformational dimension. This can also be described as the “malleability” of the protein, or how far from it’s native structure it can be stretched before it unfolds. By performing the FX experiment at a variety of speeds, this underlying property of the protein can be investigated. One model for the underlying energy landscape of protein unfolding is the Bell model^[77].

Monte Carlo simulations are used to obtain these parameters from the experimental speed dependence of the unfolding force measured by SMFS. Using this model, the speed dependence of an underlying energy landscape is simulated by the probability of each domain unfolding.

A model that provides more detailed information about the roughness of the unbinding energy landscape can be applied by using a range of temperatures and forces^[79]. This model can be fit with single molecule results in which the unfolding rate and spread of forces obtained depends on an auxiliary variable representing an entropy value arising from fluctuations in internal dynamics^[80]. This model incorporates the intrinsic disorder

of biomolecules^[81], as a rough energy landscape (as shown in the right hand insert in figure 1.8) and uses the effect of temperature on this system to fit SMFS data. Experimentally this has been applied to measure the “roughness” of the unbinding energy landscape of the small GTPase Ran and the nuclear transport receptor importin-beta^[82]. This measured an energy landscape roughness with a scale of over $5 k_B T$, by applying dynamic force spectroscopy across the complex.

1.4.2 CSPs in SMFS

The cold shock protein has been established as a good model for SMFS studies^[7,83]. The small size (7.4 kDa), single domain, high β -sheet content and two state folding pathway^[52] make it mechanically robust with a recorded unfolding force of over 20 pN, which can be readily measured in Force Extension (FX) measurements.

1.4.2.1 Multiple Unfolding Trajectories

As well as unfolding proteins at a constant speed to explore the mechanics of unfolding, Force Clamp (FC) measurements apply a constant force and measure the response of biomolecules^[78]. This has been applied to the *Tm*-CspB and accompanied by Molecular Dynamics (MD) simulations to explore the unfolding trajectory of this protein^[84]. It was found that underneath the simple two state unfolding process seen in FX, and single molecule fluorescence^[85], was a wide array of possible unfolding trajectories. Which unfolding trajectory occurred depended on the applied force. These trajectories include mechanical intermediate states held under the applied force, some of which were recorded to be held by the protein for a few seconds. The widest range of trajectories was at an applied force of 50 pN, and the results of these experiments matched with theoretical predictions. This study highlights the importance of single-molecule studies, as the individual unfolding of a protein molecule can vary between identical molecules, and this behaviour would be lost in bulk measurements that record only the population average.

1.4.2.2 Temperature Dependence of CSP Unfolding

The temperature dependent malleability of the CSP from the hyperthermophile *Thermatoga Maritima* (*Tm*CSP) has been measured using SMFS^[7]. This protein was found to be mechanically stable across the temperatures used (5-40 °C). The main effect of tem-

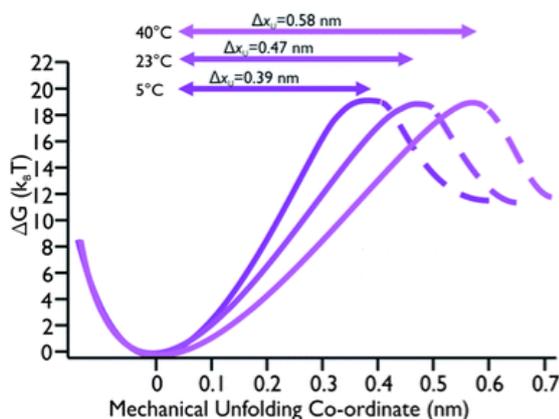


Figure 1.19: Schematic depiction of the increase in distance to the transition state along the mechanical unfolding co-ordinate of the *Tm*CSP with increasing temperature. Taken from Tych 2013^[7]

perature was to change the distance along the mechanical unfolding dimension (Δx_U) of the TS, as shown in figure 1.19. This effectively softens the protein at higher temperatures, meaning it can withstand larger perturbations in this direction without unfolding. For a protein from a hyperthermophile this characteristic may aid in maintaining stability at the high temperatures in which the organism lives. The key residues for this property were identified as electrostatically charged residues around the region of the N and C termini, known as the mechanical clamp. This region is the portion of the protein bearing the majority of the tension when the protein is pulled by the N and C termini, and so is therefore the section most likely to trigger unfolding in these measurements. By mutating the mesophilic *Bs*-CSP to incorporate three of these charged residues from the *Tm*-CSP, the possibility of mimicking this property was investigated^[17]. This study found that the ionic residues around the mechanical clamp did increase the malleability of the *Bs*-CSP, while the unfolding forces were not increased.

1.4.3 Ligand Binding recorded with SMFS

There are two primary ways that SMFS can be used to measure protein-ligand binding. The first is to measure directly the interaction force by pulling apart the complex and measuring the force of unbinding. The other method is to unfold one of the binding partners while the ligand is bound, measuring any change in the unfolding process caused by the ligand presence. These methods have been established in measuring protein stability^[78,82] and complex binding strength^[86] with dynamic force spectroscopy being applied to explore the energy landscape of either unfolding^[7] or unbinding^[87].

1.4.3.1 DNA Stretching Forces

The application of single-molecule force spectroscopy (SMFS) to gain a deeper insight into the mechanics of protein-DNA interactions can be utilised in a number of ways. These depend on where the force is applied across the protein-DNA complex. Previous SMFS work has been implemented by stretching the DNA strand and observing the effects caused by the presence of DNA binding proteins in the solution^[88]. This provides information on how the protein affects the DNA's mechanical properties^[89] which is an important function of many DNA binding proteins. It can also be used as a method for measuring the stoichiometry and binding free energy of ligands on a single molecule level^[90]. One issue with this method is that it usually requires the DNA to be in the stretched form for force measurements. This may affect the protein-DNA binding mechanics as it will not allow the highly flexible ssDNA to bind in the way it normally would.

1.4.3.2 Unbinding Forces

A further method applies the force across the complex to directly measure the unbinding force. Unbinding SMFS measurements have been used to accurately measure disassociation rates, unbinding forces and energy barriers^[91,92].

Using ssDNA as the binding ligand in such experiments, the rupture force between the

Single Stranded DNA- binding protein and a 60 nucleotide long strand of ssDNA has been measured in this way^[93]. This study used the change in buffer salt conditions and the influence of this on the flexibility of the ssDNA to observe the change in binding modes. More recently this technique was used to investigate the binding of APOBEC3A, a ssDNA cytidine deaminase, with ssDNA^[94]. This found a sequence dependent rupturing force, with deaminase-specific sequences forming more stable structures. Investigation into the unbinding energy landscape of aptamers with the proteins IgE and α -thrombin^[95] found that 20-40 nt long strands of ssDNA went through at least one intermediate state with two energy barriers to overcome.

1.4.3.3 Protein Unfolding Forces

Another SMFS method involves applying the force across the protein rather than the DNA to measure any changes in unfolding force corresponding to changes in the protein stability^[96]. As a key biological function for many proteins, the binding of ligands induces conformational changes in the protein structure and an altered thermodynamic stability^[97] as the ligand binding induces a population shift in the protein conformation^[98]. While an increase in thermodynamic stability does not always lead directly to an increase in mechanical stability^[99] the AFM has been applied to investigate whether there is any accompanying change in mechanical stability in many systems^[100]. One of the first found an increase in unfolding force of 200% in dihydrofolate reductase from the Chinese hamster ovary when bound to its ligands^[101]. Another found an increase of 40% at 400nm s^{-1} in the small protein GB1 when it was bound to fragments of the IgG antibody^[102]. Macromolecular crowding was also found to influence the unfolding pathway in some cases, leading to an increase in observed unfolding force without a binding interaction^[103]. While some proteins are found to be stabilised by bound ligands, others show no enhancement at all. This is thought to be due to the mechanical stability being dependent on the transition state of the protein structure under an applied force^[104] while the thermodynamic stability depends only on the difference between bound and unbound states. If ligand binding affects the transition state to the same degree as the bound state then the unfolding force will not increase on binding as the overall change

in stability is not detectable. This transition state depends on the unfolding pathway of the protein in the pulling experiments.

1.4.3.4 Ligand Binding Affect on Protein Unfolding

A SMFS investigation into the binding of small ubiquitin-like modifiers (SUMOs) found a drastic increase in SUMO unfolding force upon the non-covalent binding of small peptides with SUMO binding motifs (SBMs)^[105]. Ligands bind to these proteins and extend the β -sheet of the SUMO protein, as seen in figure 1.20. In this complex the binding site is not directly between the N and C termini where the force is applied in unfolding experiments. Two such SBM ligands, S10 and S12, with differing dissociation constants were derived from target proteins RanBp2 and PiasX, respectively. RanBp2 catalyzes SUMO E3 ligase activity and PiasX plays a role in the interaction with SUMOylated transcription factor Elk-1. The effect of ligand binding on the mechanical stability was compared in unfolding experiments.

The SUMO unfolding force was found to increase from around 130 pN to 170 pN,

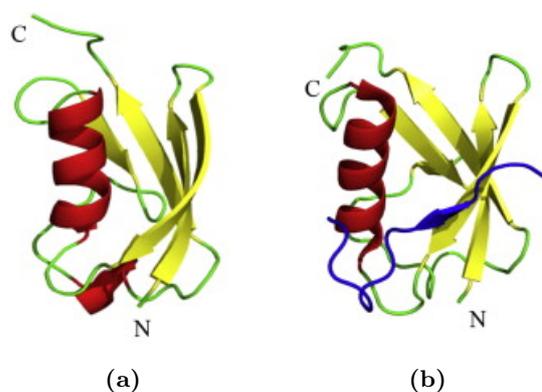


Figure 1.20: Structure of the SUMO1 (a) in the ligand-free state and (b) bound with the S12 ligand shown in blue, PDB entries 1A5R and 2ASQ, respectively. Taken from^[105]

or 30%, at a pulling speed of 400nm s^{-1} upon binding and at higher speeds this difference was increased, as can be seen in figure 1.21. Monte Carlo simulations were used to calculate the kinetic parameters of unfolding from these pulling-speed dependencies and demonstrated that the distance to the unfolding transition state was decreased on

binding. A surprising finding was that when comparing two ligands of varying binding strength, the weaker bound SBM S12, with a dissociation constant of $6\mu\text{M}$, led to a greater increase in mechanical strength than the stronger bound S10, with a dissociation constant of $70\mu\text{M}$.

For the SUMOs to bind to their targets, the authors concluded that having a higher

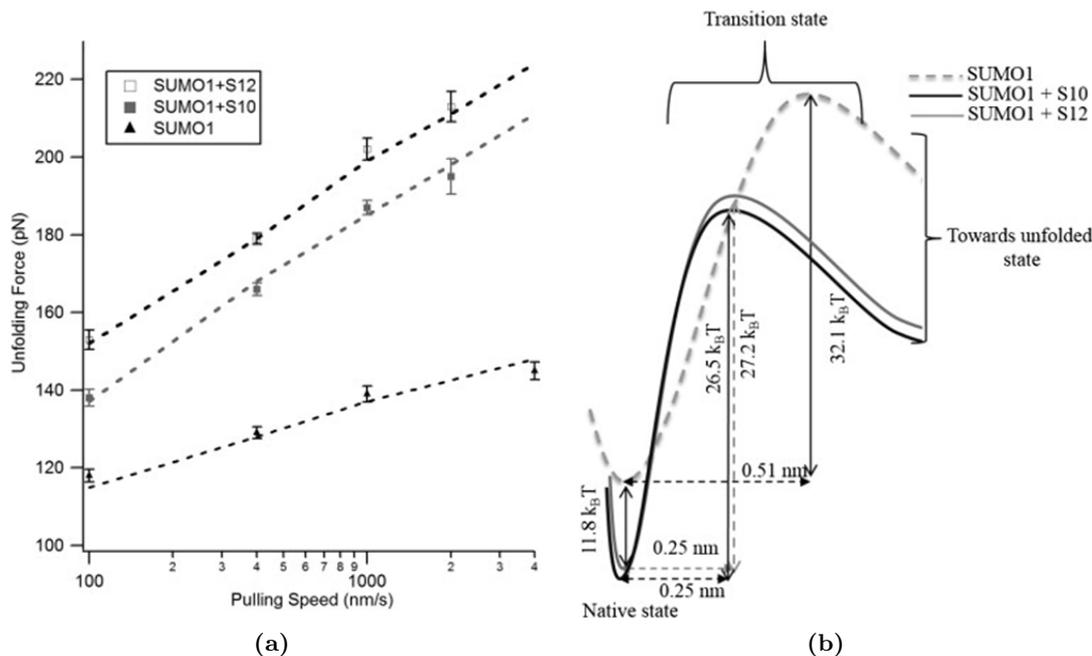


Figure 1.21: (a) Unfolding forces from SMFS experiments plotted against pulling speed for the SUMO1 protein when ligand-free and in complex with S12 and S10. (b) The effect of ligand-binding on the unfolding barrier as calculated from Monte Carlo simulations based on the fits shown in A. Taken from^[105]

flexibility while unbound, measured as a spring constant of around 1 Nm^{-1} , allowed the protein to access more structures. Upon binding, the stiffness of the protein increases to around 3.5 Nm^{-1} when the protein no longer needs to be adjustable. The results from this system demonstrate that residues can directly influence protein mechanical stability even when they are positioned far from the mechanical clamp. It also demonstrates the trend usually seen in SMFS studies on the effect of ligand binding on protein stability, which usually involves an increase in protein rigidity^[96,102,106]. This can be compared to results using other experimental techniques, where the effect of ligand binding has been shown to both increase and decrease flexibility^[107].

1.4.3.5 Ligand Concentration

The stabilisation effect on the mechanical unfolding force of ligand binding to SUMOs was not dependent on the concentration of ligand^[105]. When this concentration exceeded a saturation level that was dependent on the dissociation constant of the complex, increasing ligand concentration did not cause an increase in unfolding force. This is the case for most complexes that induce a mechanical stabilisation in the protein being pulled in the SMFS experiments^[100]. When the mechanical stabilisation effect of binding does depend in ligand concentration it is supposed that the binding kinetics are responsible for this effect^[108]. If the off-rate (k_{off} in equation 1.4) of binding is significant in relation to the time-scale of the SMFS experiment then an increase in ligand concentration will continue to increase, or shift, the measured mechanical stability based on a time average. Increasing the concentration leads to a higher percentage of unfolding forces measured when there is a ligand bound to the protein as it unfolds. Using this property, the mechanical difference in unfolding force of bound and unbound proteins can be used in a binding assay to measure a dissociation constant^[109]. This allows measurement of the relative number of unfolding forces on the bound and unbound states measured in the force extension traces. Measuring these at a range of ligand concentrations, and counting the number of peaks observed in the bound and unbound state provides a binding curve for the complex. This allows characterisation of the protein ligand interaction directly from induced mechanical stability, provided that a measurable change occurs.

For small force measurements (below 10 pN) the unfolding pathway can be influenced by low concentrations of ligand which affect the apparent unbinding transition state^[110]. This is due to the influence of the k_{off} rate becoming dominant in inducing protein unfolding at these low forces. The unfolding of a protein under force that is stabilised by a ligand will depend on the rate at which the ligand unbinds more than the rate of unfolding of the free protein. In this model, the ligand concentration will affect this apparent rate of unfolding, causing a shift in the apparent transition state, in the calculated energy landscape of unfolding.

1.4.3.6 DNA Binding Affect on Protein Unfolding

The technique of using SMFS to investigate protein-ligand binding has seldom been applied to DNA binding. One endeavour to observe the effect of DNA binding on an unfolding trajectory of a protein used the p53 DNA Binding Domain (DBD)^[111]. To explore the interactions of different sections of this protein and DNA, three complexes were used in SMFS experiments: The full DBD and no DNA present; The DBD with the N-terminal removed and no DNA; and the full DBD with DNA present. In the traces measured there was a correlation seen between the position of unfolding peaks in the case with no DNA present and the N-terminal removed, and the full DBD with DNA present.

When there is no DNA present and the whole DBD is unfolded there is only one unfolding peak seen, indicating that the N-terminal binds across the two domains causing the whole structure to unfold at once. From this change in the unfolding pathway it was concluded that the N-terminal and the DBD interact when no DNA is present and that DNA binding blocks this interaction. This method of detecting ligand binding by SMFS uses the change in direction of pulling force caused by ligand binding, which can only be applied to systems with this particular property.

The number of clear peaks measured in this study is fewer than 10 for each of the set-ups due to a low yield of successful extension. Quantitative measurements of the unfolding forces involved are therefore statistically unreliable. While this study provides some insight into this particular DBD and the effect of DNA binding on the unfolding pathway, due to its complexity, it does not provide any results that could be extracted to apply to other DNA binding proteins. For results that can be applied more generally a system with a simpler unfolding pathway is required.

Recently SMFS was applied to the stabilisation of RRM1 domain of TAR DNA binding Protein (TDP)-43 when bound with short strands of ssDNA^[112]. The TDP-41 is a protein involved in various aspects of RNA processing, and has the ability to bind both RNA and DNA. This study used an engineered structure incorporating one of the RNA Recognition Motifs (RRM1) into a chimeric polyprotein chain with 4 I27 domains. SMFS on this structure was used to measure the unfolding force of the RRM1 domain with

and without the presence of binding ssDNA oligos. Without the ssDNA bound, only a

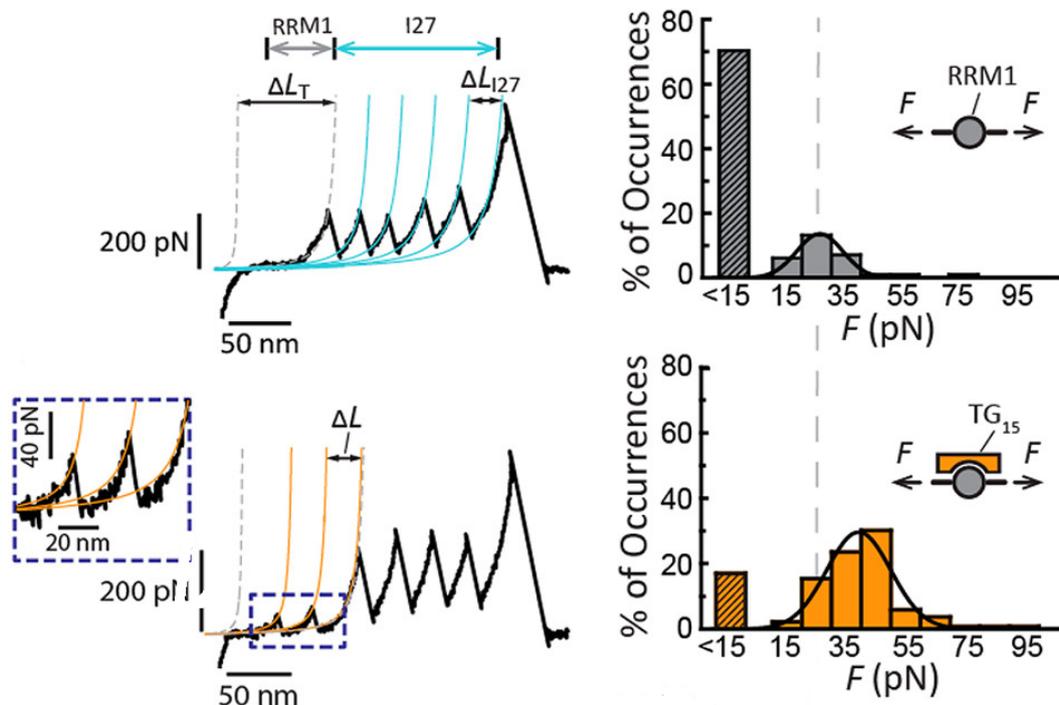


Figure 1.22: Force extension traces (left) of the chimeric polyprotein unfolding in the absence (above) and presence (below) of ssDNA oligos which bind to the RRM1 domains. Histograms of the measured unfolding forces from multiple unfolding events (right) show an increase in the average unfolding force of the RRM1 domain in the presence of ssDNA. There is also a decrease in unfolding events recorded with an unfolding force below 15 pN. Taken from Wang 2018^[112]

small (mean value less than 30 pN) mechanical unfolding force was recorded, as can be seen in figure 1.22. When the ssDNA was added this increased to around 40 pN to a clear, unfolding peak, indicating the binding of ssDNA caused the structure to become mechanically resistant. This effect was described as the ssDNA acting as a “mechanical staple” on the protein. This demonstrates that the binding of nucleic acids can impact the mechanical unfolding properties of proteins.

1.4.3.7 Pulling Direction

The influence of the pulling direction on the unfolding pathway and the effect of ligand binding was investigated for the maltose binding protein (MBP)^[113]. In a similar way

to the pulling geometry affecting the mechanical resistance measured in proteins^[114], it was found that in this system the pulling direction was key to measuring the difference caused by maltose binding to the MBP. When the polyprotein was first pulled at the N and C termini then no effect of maltose binding on the unfolding force could be seen^[115]. The locations of the links between proteins in the polyprotein chain was then engineered so that the applied force direction was across the binding site of the maltose. This led to a clear enhancement in the unfolding force of around 15 pN, or 12%, at a pulling speed of 1000 nms⁻¹ when maltose was present as opposed to the MBP alone. In this relatively large (42 kDa) protein structure the unfolding force must be applied across the site of ligand binding to detect the mechanical stability caused locally by ligand binding.

1.5 Nuclear Magnetic Resonance Theory

Nuclear Magnetic Resonance (NMR) spectroscopy is a versatile technique that can be used to obtain chemical, structural and dynamic information on small molecules in solution. As most proteins are flexible molecules that function in solution in the cell NMR is an invaluable measurement method for this class of biomolecules^[116]. Here is a brief overview of some of the theory behind the NMR methods used here. More thorough descriptions can be found elsewhere^[117-120].

1.5.1 The NMR signal

NMR active nuclei have magnetic moments which under normal conditions are randomly oriented. When a strong magnetic field (B_0) is applied, these magnetic moments are aligned either with B_0 in a low energy spin $+\frac{1}{2}$ state, or against B_0 in a high energy spin $-\frac{1}{2}$ state. The energy (E) of each state is given by equation 1.7 where μ_z is the magnetic moment of the nucleus in the direction of B_0 , usually taken as the z axis.

$$E = -\mu_z B_0 \tag{1.7}$$

μ_z is equal to $\gamma m \hbar$ where γ is the gyromagnetic ratio, m is the magnetic quantum number and \hbar is the reduced Planck constant. Putting this into equation 1.7 the energy of each state is given by equation 1.8

$$E = -\gamma m \hbar B_0 \quad (1.8)$$

The magnetic quantum number is $-\frac{1}{2}$ for the high energy state and $+\frac{1}{2}$ for the low energy state, so the difference in energy between the two states is given by equation 1.9

$$\Delta E = \gamma \hbar B_0 \quad (1.9)$$

This energy difference leads to a population difference as more magnetic moments align in the low energy state, as the energy increases with B_0 so does the population difference. All the nuclei spins precess roughly around the B_0 field axis at the Larmor frequency. This makes it convenient to use the rotating frame to describe the magnetisation vectors. This frame rotates at the same rate as the observation frequency, allowing the isochromats to be tracked in a more understandable way. At equilibrium, all these individual spins are randomised and incoherent, so they average each other out, producing no net magnetisation.

The NMR signal is produced by hitting these nuclei with an electromagnetic radiation

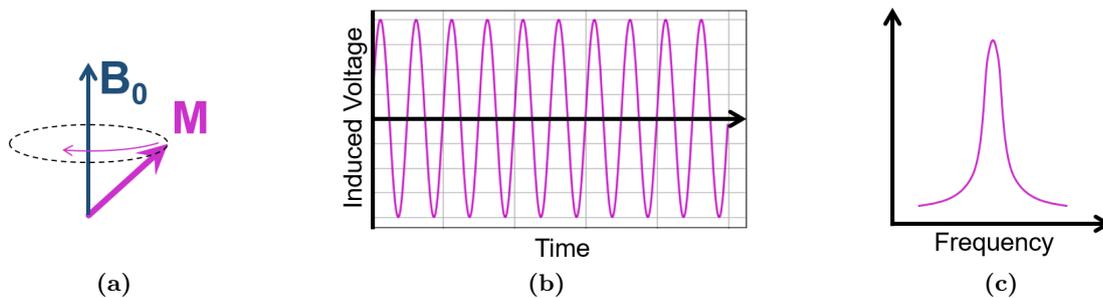


Figure 1.23: Diagrams showing the production of an idealised NMR signal. (a) After the RF pulse of resonant frequency, the excited spins form a coherent magnetisation out of alignment with the B_0 field, rotating around the B_0 field at roughly the Larmor frequency. (b) This rotation of a net magnetisation produces a signal in the receiver that (ignoring any loss of energy through relaxation effects described in section 1.5.4) follows a sinusoidal wave pattern. (c) Taking the Fourier Transform of this signal produces a frequency map with a peak at the resonant frequency for all active nuclei sharing that chemical environment.

pulse which is of the appropriate energy, near the Larmor frequency, to cause spins to

flip to the high energy state. This produces a set of spins acting coherently, which produce a net magnetisation, M , as shown in figure 1.23a. This net magnetisation M , is the sum of the magnetisation produced by all the coherent spins, and rotates around the B_0 field at the resonant frequency. If it is out of alignment with the B_0 field, this produces a signal in the receiver induction coil. Ignoring the loss of energy through relaxation processes, this rotation of M induces a sinusoidal signal as shown in figure 1.23b. This is an idealised form of the Free Induction Decay (FID), without the decay which will be described in section 1.5.4. Taking a Fourier transform of this signal will identify the resonant frequencies of the nuclei, as shown in figure 1.23c.

The frequency of this resonance will vary depending on the chemical environment surrounding the nucleus. In the simplest case, nuclei surrounded by a higher electron cloud density are more shielded, and so have a weaker local magnetic field (B_{local} in equation 1.10).

$$\Delta E = \gamma \hbar B_{local} \quad (1.10)$$

Here the local magnetic field $B_{local} = B_0 - B_{ind}$ where B_{ind} is the field induced by the electron shielding, opposing the external field. The lower ΔE between high and low energy spin states in a shielded nuclei means a lower resonant frequency, ω . This difference in ω is what allows NMR to differentiate between atoms in a protein molecule, as each one has a unique chemical environment and therefore a unique resonant frequency. To differentiate these unique frequency signatures a strong magnetic field is required. As the B_{local} of equation 1.10 scales with B_0 , stronger fields will have fewer overlapping signals due to the increase in energy difference between high and low energy spin states. This is why very powerful magnets are desirable in protein NMR, and also why frequency is not usually the unit used for comparing resonance values. The frequency can be recorded as angular frequency, ω , or as a rate per second, ν , where $\nu = \omega/2\pi$. This is generally converted to a chemical shift, with the unit of parts per million (ppm).

$$\delta = \frac{(\nu - \nu_0)}{\nu_0} \quad (1.11)$$

Here δ is the chemical shift, with units ppm, describing the change in frequency, ν , relative to the frequency of a universally used marker, ν_0 , which here is 4,4-dimethyl-4-

silapentane-1-sulfonic acid (DSS). This is a useful unit because the resonant frequency, ν , will change with the strength of B_0 fields due to the change in B_{local} , equation 1.10. The relative chemical shift will be consistent for the same sample measured in different B_0 fields of different instruments, making data easier to compare.

1.5.2 Multi-Dimensional NMR

Proteins contain a large number of hydrogen atoms, with two in each unit of the peptide bond and more in the amino acid side chains (figure 1.3). Measuring the NMR signal from these protons results in a 1D frequency spectrum filled with many peaks corresponding to the chemical shift for each proton in the protein, plus a large peak corresponding to the signal from the water in the solvent. Even with the strong magnetic fields used in NMR increasing the resonant frequency differences between protons, it can be very difficult to identify separate peaks in even a relatively small protein. This is where multi-dimensional NMR becomes very useful in studying proteins, where each peptide bond has a known structure containing hydrogen, nitrogen and carbon atoms (figure 1.2a). By exploiting the influence neighbouring nuclei have on each other, the net magnetisation can be passed from one type of nuclei to a neighbour. Each type of nuclei provides another dimension to the protein spectrum, allowing the signal from each amino acid residue to be separated and identified.

1.5.2.1 NMR Active Nuclei

^1H , ^{13}C and ^{15}N are the most useful NMR active nuclei for studies of proteins. While ^1H is already present in abundance and can be used for initial 1D studies on protein NMR properties, ^{13}C (natural abundance = 1.10%^[121]) and ^{15}N (natural abundance = 0.37%^[122]) are not. NMR measurements on these isotopes would require high sample concentrations and long experiment times if reliant on the natural abundance. To be used practically as probes, these isotopes can be enriched in the protein under study, which will increase signal-to-noise ratio in measurements so they require less instrument

time. This is done by growing the bacterium in ^{13}C and ^{15}N containing medium during protein production.

1.5.2.2 Spin Coherence Transfer

Even with enriched protein samples the low signal-to-noise ratio for ^{15}N and ^{13}C nuclei is a limiting factor. By passing the spin coherence from the protons to these nuclei this signal can be increased. The Inensitive Nuclei Enhanced by Polarisation Transfer (INEPT) method allows the spin coherence to be passed through covalently attached atoms via the J-coupling between them. J-coupling is the interaction via the local electrons between two covalently bonded nuclei.

The first step in the spin coherence transfer involves the spin echo pulse sequence, as

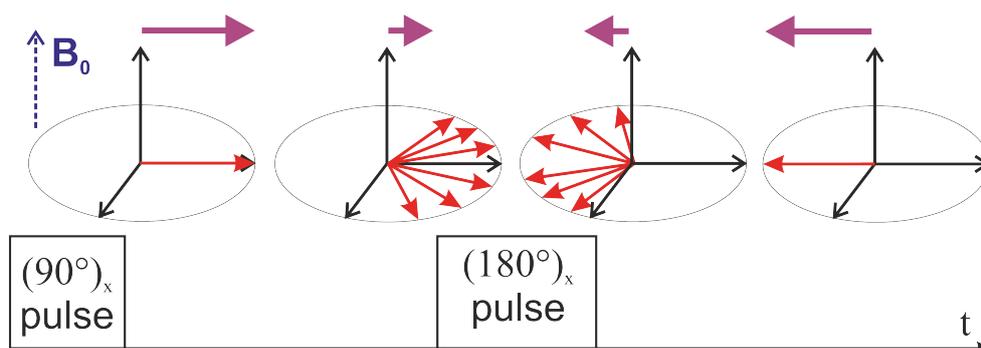


Figure 1.24: Representation of the spin echo pulse sequence focusing incoherent spins. Initially the system follows the same relaxation seen in figure 1.31, then after a time period, τ applying a 180°_x pulse. This will cause a refocusing of the spins after the same time τ , as the faster drifting isochromats catch up to the slower ones.

shown in figure 1.24. This uses a radio-frequency pulse to first align the isochromats along the x-y plane, with a 90°_x pulse. In the rotating reference frame, the nuclei spins will be aligned immediately after this pulse but will quickly drift out of alignment according with the T_2 relaxation of the system, described in section 1.5.4.3. By then applying a 180°_x pulse after a time, τ , the isochromats are rotated around the y axis. As the direction of drift is unchanged, the faster moving isochromats now catch up to the slower ones, similar to a race, if at some point the finishing line is changed to where the starting line is. This results in all of the magnetisation being focused after the same

amount of time between the two pulses, τ .

Some of this drift from coherence is due to the J-coupling with neighbouring nuclei. As the rate of this drift is directly related to the J-coupling constant, this can be used to transfer the spin coherence to the coupled nuclei. If the time period, τ , is set to the reciprocal of the J-coupling frequency (figure 1.27), this will allow the spin coherence to be transferred to the less sensitive nuclei as shown in figure 1.25.

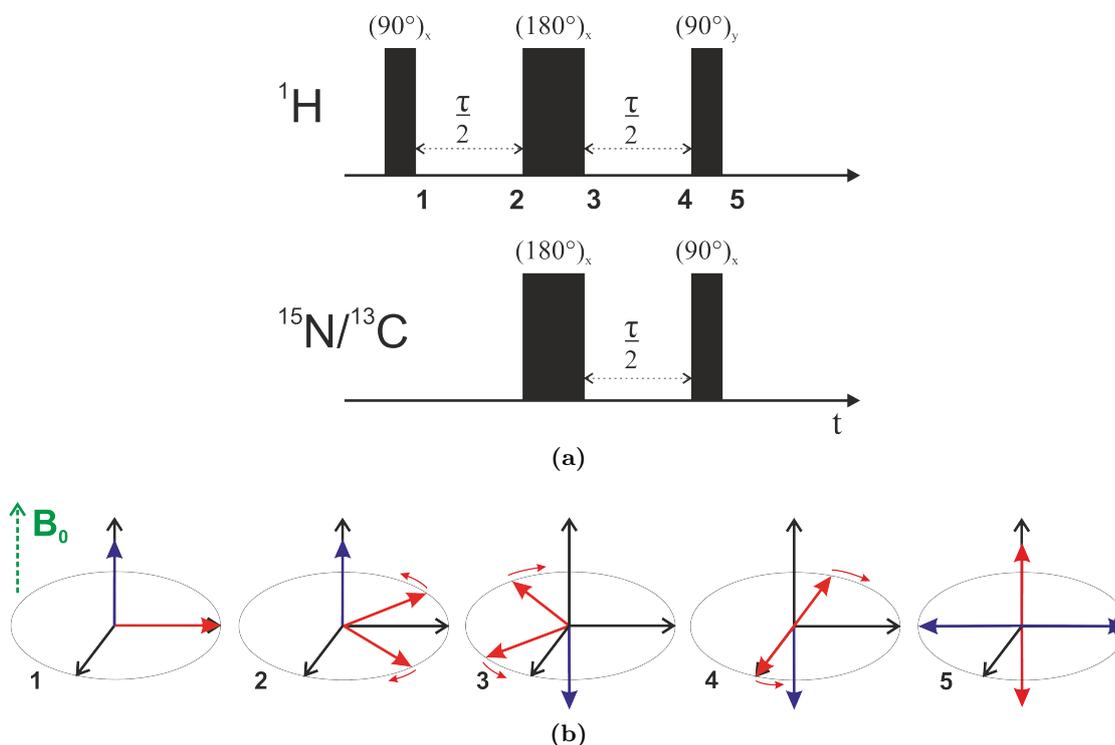


Figure 1.25: (a) INEPT pulse sequence passing the spin coherence from the ^1H protons to the ^{15}N or ^{13}C nuclei. The ^1H nucleus undergoes a spin echo focusing pulse with delay time $\tau/2$, where τ is the reciprocal of the J-coupling frequency between the proton and the second nucleus. (b) Vector representation of the magnetisation of the protons (red) and second nitrogen or carbon nuclei (blue) for each stage of the sequence. Initially (steps 1-2) this follows the same sequence as the spin echo to focus the proton spins. Then with the 180° pulse applied to both nuclei (step 3) those proton spins that are J-coupled to the second nucleus with a frequency of τ^{-1} will be pulled in the reverse direction. After the same delay (step 4) they will be oriented in the y direction, so the application of 90° pulses to both nuclei will result in a spin coherence with the second nuclei on the x-axis and the protons oriented in the z-direction (step 5).

1.5.2.3 2D

This INEPT method is used to produce the ^1H - ^{15}N HeteroNuclear Single Quantum Coherence (HSQC) spectra. This is done by transferring the spin coherence from the protons to the ^{15}N nuclei, allowing it to evolve for a time t_1 , then transferring it back to the ^1H with the inverse of the INEPT pulse sequence. This pathway for a single residue in the protein backbone is shown in figure 1.26a. The FID of the ^1H is recorded while the ^{15}N signal is decoupled to ensure only one peak is produced per ^1H - ^{15}N bond. Varying the length of t_1 will select ^{15}N signals by their resonant frequency, producing the second dimension of the spectrum as shown in figure 1.26b.

These HSQC spectra are useful for protein NMR as each amino acid residue (other

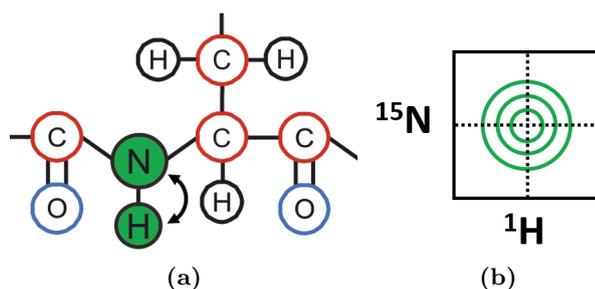


Figure 1.26: Production of a simple HSQC spectrum (a) Pathway of the spin coherence in the peptide bond from the proton to the ^{15}N and back to the proton for detection (b) The resulting 2-Dimensional spectrum with the ^{15}N chemical shift as the Y-axis against the ^1H chemical shift on the X-axis. A single peak is represented with contour lines (green) representing the intensity of the signal from the ^1H - ^{15}N bond. Dotted lines identify the chemical shift at the peak maximum for each dimension.

than proline residues) contain this ^1H - ^{15}N bond and should produce a single peak. A few side chains also contain the ^1H - ^{15}N bond and so also appear in the spectrum but are usually easily identifiable. To identify which peak corresponds to which residue in the protein sequence, more information is required. Some peaks can be identified if the amino acid has a very characteristic chemical shift in the proton or nitrogen dimension, but to identify the source of each signal the carbon atoms connecting each residue need to be included in these measurements.

1.5.2.4 3D

When a sample labelled with both ^{15}N and ^{13}C is used, this method can be applied to extend the spectra into a third dimension. By selecting a delay time, τ examples shown in Figure 1.27, to transfer the magnetisation to various carbon nuclei the residues can be connected sequentially.

Each experiment is named by the pathway this spin coherence transfer follows, all

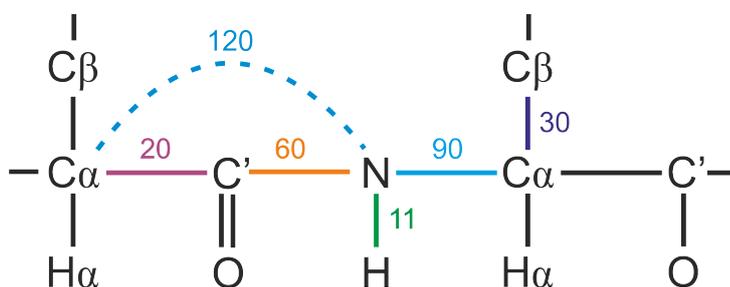


Figure 1.27: Approximate delays, τ in ms, for magnetisation transfer between nuclei. Delays quoted are $1/J$ where J is the coupling constant between the nuclei, in Hz^[123]

starting with HN for the ^1H - ^{15}N bonds in the protein backbone. HNCA experiments transfer the magnetisation from ^1H nuclei to ^{15}N nuclei to the $^{13}\text{C}\alpha$ nuclei, and in HN(CO)CA experiments it transfers from H to N to C' to the neighbouring $\text{C}\alpha$ nuclei, as shown in figure 1.28a. This also describes the HNC(O), HN(CA)CO and HN(CA)CB experiments used in backbone and side chain assignment.

Each ^1H - ^{15}N bond should have two associated HNCA peaks, one for the residue i and one for the previous residue in the sequence $i-1$, as shown in figure 1.28. The chemical shift of the $\text{C}\alpha$ peak in the ^{13}C dimension can be used to identify matching $\text{C}\alpha$ signals from other parts of the ^1H - ^{15}N spectrum. By matching HNCA signals in the ^{13}C dimension, the two ^1H - ^{15}N bonds that are bridged via this particular $\text{C}\alpha$ can be identified. The HN(CO)CA peak will then overlap with the HNCA peak, identifying which peak belongs to the preceding, $i-1$, residue. This helps build up an order for groups of connected residues. By then identifying the type of amino acids using the chemical shift values of the different carbon signals, the peaks in the protein ^1H - ^{15}N HSQC spectrum can be assigned to the corresponding protein residues.

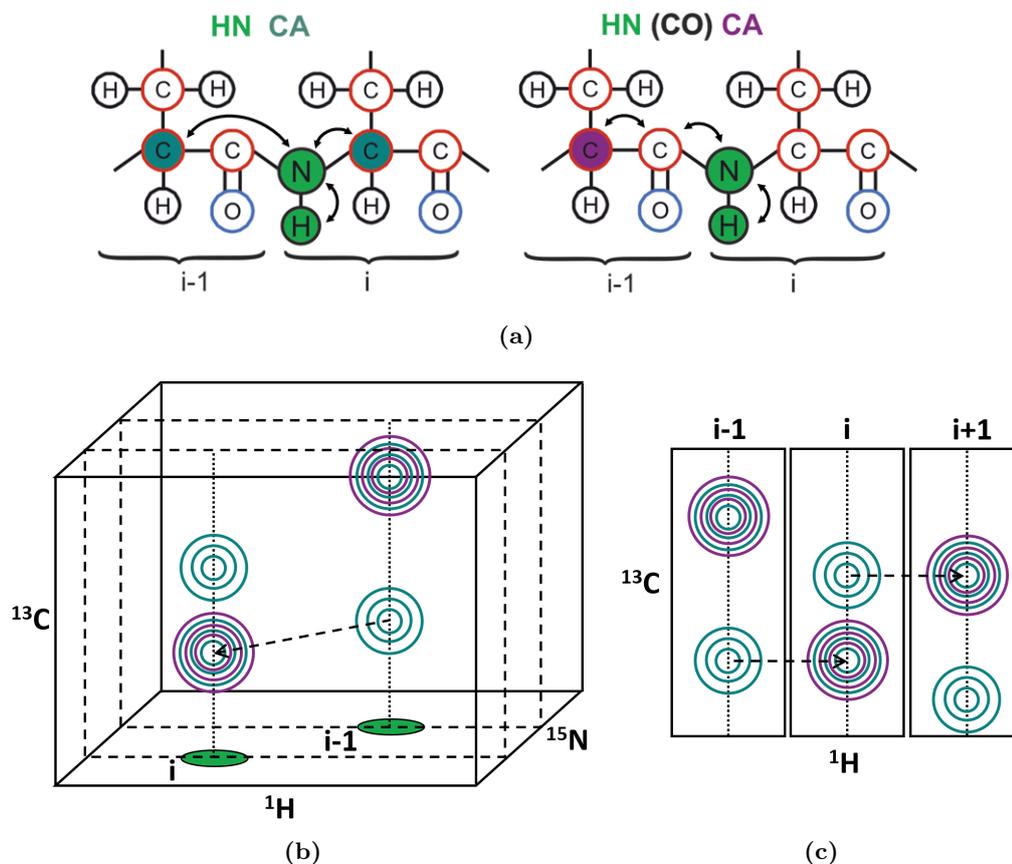


Figure 1.28: (a) Pathways for the spin coherence transfer in HNCA (teal) and HN(CO)CA (purple) measurements. Each backbone ^1H - ^{15}N bond has two corresponding HNCA signals in the ^{13}C dimension, corresponding to the $\text{C}\alpha$ nuclei of the residue (i) and the preceding residue ($i-1$). There is only one HN(CO)CA signal for the preceding $\text{C}\alpha$ ($i-1$) (b) Diagram of the spectra extracted from the 2D HSQC (green) into the 3rd ^{13}C dimension in HNCA and HN(CO)CA experiments. Two 2D slices of in the ^{15}N dimension show two HNCA signals (teal) and a single HN(CO)CA (purple) signal for each peak in the HSQC corresponding to a backbone residue. (c) 2D slices showing how these signals are used to sequentially connect the residues ($i-1, i$ and $i+1$) corresponding to each peak in the HSQC spectrum.

1.5.3 Protein dynamics measured by NMR

Due to the dynamic nature of proteins, and the importance this has for protein function^[21,124], having only the static structure does not give the full picture. NMR is a useful technique to probe these dynamics as it can provide information on the time-scale of each structural change^[125], as well as providing structural information about each sub-state. As shown in figure 1.29, a range of NMR techniques can be applied to understanding dynamics at all the different time-scales of protein motion. At the

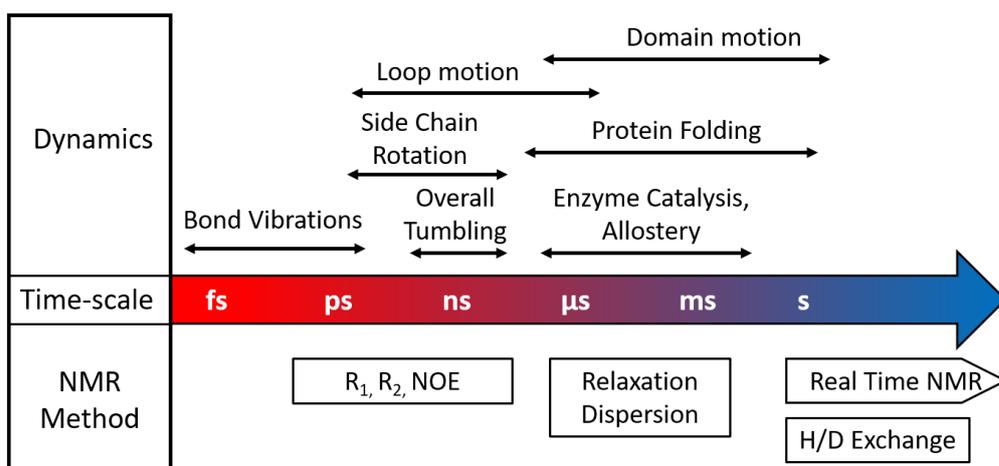


Figure 1.29: Diagram of the time-scales accessible by different NMR techniques. Adapted from Kleckner(2011)^[120].

fastest time-scales accessed by NMR experiments, picoseconds to nanoseconds, there is the physical processes of bond vibration, side chain rotamer interconversion, random coil and loop motions and backbone torsion angle rotation. At slower scales of 0.3–10 ms processes include side chain reorientation, loop motion, secondary structure changes and hinged domain movements.

1.5.4 Nuclear Spin Relaxation

Fluctuations of the molecule in the magnetic field cause fluctuations in the local fields experienced by each nuclei. These variations cause a change in the rate at which spins return to the low energy state during the measurement. This is the basic principle of how NMR relaxation measurements can be used to investigate dynamics of the molecule. These relaxation rates are related to fast ps-ns molecular motion by 3 principal mechanisms. These are Chemical Shift Anisotropy (CSA), Dipolar Coupling (DC) and Quadrupolar Interactions. CSA is the effect of a changing local magnetic field due to orientation of the N-H bond with respect to the B_0 field. This will vary due to molecular motion and the oscillating local magnetic field experienced by the nuclei can stimulate relaxation. The CSA for a nucleus X is given by $c_X = \Delta \left(\frac{\omega_X}{\sqrt{3}} \right)$, where ω_X is the chemical shift of the nucleus X. DC is the through-space interaction between pairs of nuclei, d_{HX}

for the heteronucleus X with proton H is described in equation 1.12.

$$d_{HX} = \left(\frac{\mu_0 h \gamma_H \gamma_X}{8\pi^2} \right) \frac{1}{r_{HX}^3} \quad (1.12)$$

Here μ_0 is the permeability of free space, h is Planck's constant, γ is the nuclear gyro-magnetic ratio, r_{HX} is the length of the bond between H and X. As the orientation and distance of these pairs changes with motion, the effect they have on the other's local magnetic field will change, also leading to oscillating fields that will induce relaxation. Quadrupolar interactions can occur with nuclei that have spin of 1 or more, such as ^2H , ^{14}N and ^{17}O , so is not involved in relaxation processes looked at here.

To quantify the relation between these molecular motions on the ps-ns time-scale and NMR observables, an orientation correlation function for each bond vector with respect to time, $C(t)$, can be introduced. This function describes the time taken for each bond vector to "forget" previous orientations and become randomised. Taking the Fourier transform of this gives the spectral density function $J(\omega)$, which describes how efficiently oscillation at frequency ω will induce relaxation.

The motion of the protein molecule as a whole through the liquid, including rotational diffusion and collisions with other molecules, is described as tumbling. This can be described with a rotational correlation time, τ_m , where the correlation function is $C(t) = e^{(-t/\tau_m)}$ and τ_m is the time taken for the root mean squared average orientation of the molecules to move by 1 radian or 60° .

$$J(\omega) = \frac{\tau_m}{(1 + \omega^2 \tau_m^2)} \quad (1.13)$$

Equation 1.13 defines the spectral density function, $J(\omega)$, in relation to the rotational correlation time for molecular tumbling, τ_m . Due to the increase in tumbling speed exhibited by smaller molecules or at higher temperatures, the correlation time is shorter than for larger molecules or higher temperature environments.

The same rotational correlation time can be introduced for internal motions of the molecule, τ_e . This will also contribute to $J(\omega)$.

1.5.4.1 Heteronuclear NOE

The Nuclear Overhauser Effect (NOE) is a through-space magnetisation transfer between nuclei. It is useful in structure determination as it is very sensitive to the distance between nuclei, so can be used to determine distance through space^[119]. The Heteronuclear NOE is a measure of the NOE between a pair of two different nuclei, in this case the ^1H protons and the ^{15}N nuclei they are covalently bonded to. This is measured by taking the ratio of peak intensities recorded in equilibrium ($I_{equilibrium}$) and when the proton is saturated ($I_{saturated}$). It is saturated by hitting it with a radio-frequency pulse strong enough to cause equal numbers of nuclei in the high and low energy states, meaning there would be no NMR signal. The ratio of intensities of the nitrogen indicates how much of this saturation is transferred between the nuclei, $\{^1\text{H}\} - XhnNOE = \frac{I_{saturated}}{I_{equilibrium}}$, and for $\{^1\text{H}\} - \text{N}$ this value can range from -4 to 1 due to the negative gyromagnetic ratio of ^{15}N , γ_N . When the proton spin population is saturated by irradiation it will affect the population of the ^{15}N through the dipole-dipole interaction, so this ratio describes the strength of this interaction. The relation if this interaction to fast dynamics is described by equation 1.14.

$$\{^1\text{H}\} - XhnNOE = 1 + \left(\frac{\gamma_H}{\gamma_X} \frac{N}{R_1} \right) \frac{d_{HX}^2}{4} (6J(\omega_H + \omega_X) - J(\omega_H - \omega_X)) \quad (1.14)$$

Here γ is the nuclear gyromagnetic ratio, N is the number of ^1H atoms attached to the nucleus X (Here $N=1$), R_1 is the longitudinal relaxation rate described in section 1.5.4.3, d_{HX} is the dipolar coupling interaction (equation 1.12) and $J(\omega)$ is the spectral density function.

1.5.4.2 The Longitudinal Relaxation Rate R_1

The net magnetisation is lost to two forms of relaxation. One is longitudinal relaxation, and this is the component of magnetisation parallel to the B_0 field, or the z -direction, returning to alignment with the field. This form of relaxation is shown in figure 1.30a. The change in intensity, I , with time, t , due to this relaxation is described in equation

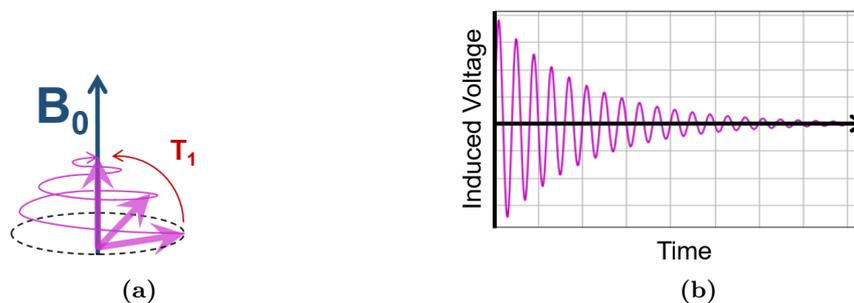


Figure 1.30: Diagrams showing the effect of relaxation on the Free Induction Decay (a) After the RF pulse of resonant frequency, the magnetisation M (pink) is initially out of alignment with the B_0 field. Longitudinal relaxation, R_1 (red), is the process of this magnetisation returning to alignment with the B_0 field. (b) This rotation of a net magnetisation produces a signal in the receiver that follows a sinusoidal wave pattern. The intensity of this signal decays due to both R_1 and R_2 relaxation, this signal is the FID.

1.15 with the relaxation rate R_1 . This usually ranges from 0.5 - 5 Hz for most proteins.

$$I(t) = I(0)(1 - e^{-R_1 t}) \quad (1.15)$$

The result of both longitudinal and transverse relaxation on the signal is shown in figure 1.30b. This is the signal recorded on the receiver coil after the RF pulse, or the FID. Relaxation accounts for the decay of this signal, as it describes the energy lost to surroundings.

1.5.4.3 The Transverse Relaxation Rate R_2

The transverse relaxation rate is the loss of magnetisation in the x-y dimension, as described in the rotating reference frame, perpendicular to the static field B_0 . This is represented in figure 1.31. The source of this magnetisation is the coherence of the collective spins, this relaxation rate is due to these spins rotating at different rates. The rate used to describe this, R_2 , is described in equation 1.16. For most proteins this value ranges from 5 - 50 Hz.

$$I(t) = I(0)(e^{-R_2 t}) \quad (1.16)$$

A representation of this process in the rotating reference frame is shown in figure 1.31. The relations between these relaxation rates as measured by NMR experiments, and fast

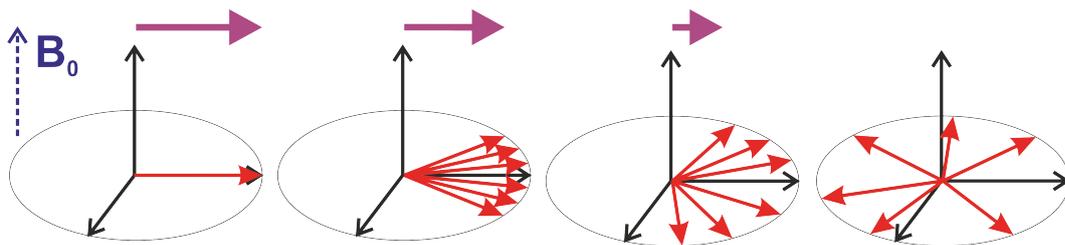


Figure 1.31: Representation of T_2 relaxation in the rotating reference frame. Time after the initial 90° pulse progresses from left to right. Red arrows represent a sample of individual nuclei magnetic vectors, the large purple arrow above represents the overall magnetisation resulting from these vectors. As time increases the spin coherence is lost and the overall magnetisation decreases.

molecular motions of the proteins, are given for a residue X in equations 1.17 and 1.18.

$$R_1(X) = \frac{d_{HX}^2}{4} (J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)) + c_X^2 J(\omega_X) \quad (1.17)$$

$$R_2(X) = \frac{d_{HX}^2}{8} (4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)) \\ + \frac{c_X^2}{6} (4J(0) + 3J(\omega_X)) + R_{ex}(X) \quad (1.18)$$

In these equations, d_{HX} is the dipolar coupling interaction described by 1.12, $J(\omega)$ is the spectral density function, and c_X corresponds to the CSA interaction, where $c_X = \Delta\sigma_X/\sqrt{3}$ and $\Delta\sigma_X$ is the effective CSA for nucleus X.

In terms of the molecular motions that contribute to each relaxation rate, $J(\omega)$ can be modelled by the overall molecular motions τ_m , the internal motions τ_e and the square of the generalised order parameter S^2 which describes the amplitude of internal motions, on the nano-picosecond time-scale. R_1 and $\{^1H\}$ -N hnNOE are comprehensively described by equations 1.17 and 1.14, so are determined by dynamics at the ps-ns times-scales. R_2 has an extra term R_{ex} included to account for the influence of μ s-ms dynamics, as described in section 1.5.5.

1.5.5 Relaxation Dispersion Measured by Carl-Purcell-Meiboom-Gill (CPMG) Measurements

Intermediate exchange (μs – ms) between states can be measured using the CPMG experiment^[118]. This technique is applicable when there is a measurable difference in the chemical shift of the different states, as shown in the intermediate exchange region of figure 1.32.

$$A(\omega_A) \frac{k_{AB}}{k_{BA}} B(\omega_B) \quad (1.19)$$

Here ω_A and ω_B are the chemical shifts of state A and B respectively. k_{AB} is the rate of exchange from state A to B, and k_{BA} is the reverse.

$$k_{ex} = k_{AB} + k_{BA} \quad (1.20)$$

The difference in chemical shift between the two states $\Delta\omega = (\omega_A) - (\omega_B)$, is a frequency. Relative to this frequency, $|\Delta\omega|$, if the exchange k_{ex} is fast then the signals will average out to one peak. Only relatively slower exchange leads to separate peaks (figure 1.32), and when $|\Delta\omega|$ is close to the value of k_{ex} the signal is obscured by line broadening, which can lead to the signal being undetectable. The exchange process directly affects the exponential decay of the FID intensity, resulting in an increase in the measured transverse relaxation rate, so $R_2^{Eff} = R_2^0 + R_{ex}$ where R_2^0 is the relaxation rate if no exchange is occurring. CPMG uses repeated refocusing pulses to reduce the effect of exchange with the secondary state on the peak intensity. In the measurements of the transverse relaxation rate the effect of this exchange is an increase the measured R_2 rate, as accounted for by R_{ex} in equation 1.18. By increasing the number of 180° pulses in a set time, T_{Relax} , the contribution of R_{ex} is found as it provides the asymmetry in the spin echo measurement. The CPMG pulse sequence is $[\tau_{CPMG} - \pi - \tau_{CPMG}]_N$ so that N π pulses are applied so that $T_{Relax} = 2 \times N \times \nu_{CPMG}$ where ν_{CPMG} is the CPMG frequency.

$$\nu_{CPMG} = \frac{1}{4\tau_{CPMG}} \quad (1.21)$$

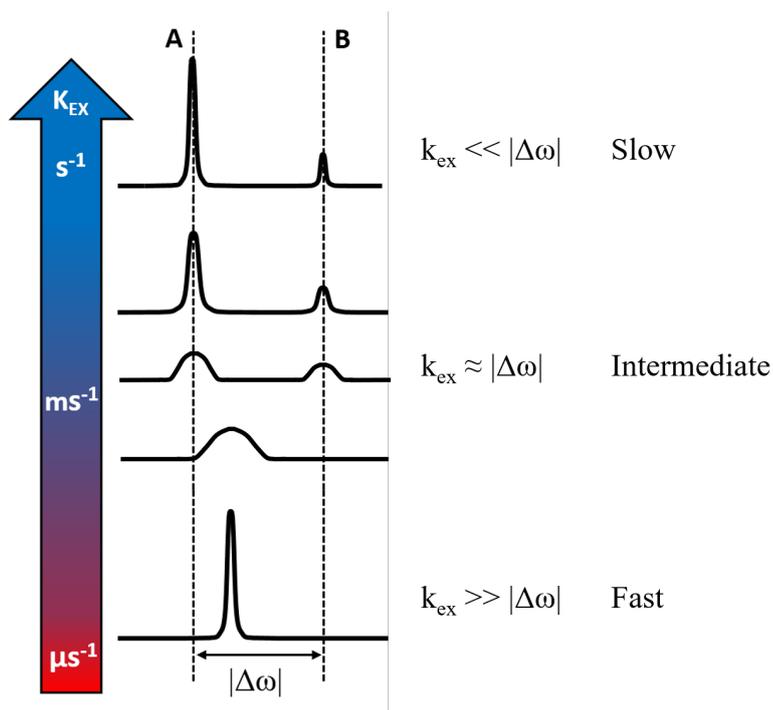


Figure 1.32: Plots demonstrating the signals generated by two states A and B, of populations $P_A=75\%$ and $P_B=25\%$, undergoing various rates of exchange, k_{ex} . When this rate is lower than the difference in chemical shift between the two states $|\Delta\omega|$ two separate peaks with intensities corresponding to the state populations are clear, this is known as slow exchange. As the exchange rate increases these become more obscured, until there is only one peak representing an average of the two states, known as fast exchange. Between fast and slow exchange is intermediate exchange where peaks are distorted. Adapted from Kleckner 2011^[120]

Residues experiencing a high R_{ex} exchange will experience a reduction in signal as ν_{CPMG} increases.

$$R_{2,Eff} = \frac{1}{T_{Relax}} \ln \frac{I}{I_0} \quad (1.22)$$

I is the intensity of the peak and I_0 is the intensity of a reference spectrum recorded without the relaxation delay. Plotting $R_{2,Eff}$ against ν_{CPMG} will show a relaxation proportional to R_{ex} if exchange is occurring and no dependence if there is no exchange, as demonstrated in figure 1.34. This curve can be fit with the two-site chemical exchange in the Carver-Richards equation^[126,127]. As an example, this is described approximately

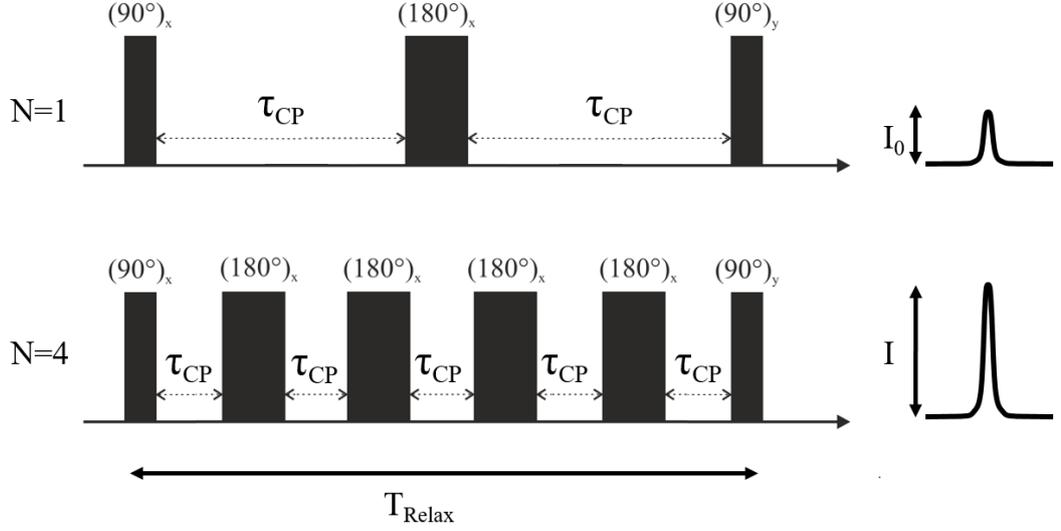


Figure 1.33: Example pulses from the CPMG sequence to record $R_{2,Eff}$. I_0 is recorded with a single spin echo pulse, then the delay between 180° pulses τ_{cp} is reduced so that over the same time period, T_{Relax} there are increasing number, N , of 180° pulses. If chemical exchange R_{ex} is occurring then the effect of this on $R_{2,Eff}$ will be reduced with increasing N so that the intensity, I , will increase due to the repeated focusing of this frequency.

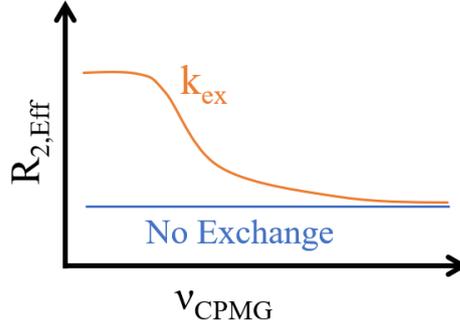


Figure 1.34: Example Plot of the measured $R_{2,Eff}$ value with increasing numbers of CPMG pulses. This differentiates the effect of intermediate exchange (k_{ex}) from R_2 .

for faster ($k_{ex} \gg \Delta\omega$) two-site chemical exchange in equation 1.23^[126].

$$R_{ex} = \frac{p_A p_B (\Delta\omega)^2}{k_{ex}} \left[1 - \frac{2}{k_{ex} \tau_{cp}} \tanh \left(\frac{k_{ex} \tau_{cp}}{2} \right) \right] \quad (1.23)$$

Here p_A and p_B are the fractional populations of the two states involved in the exchange, $\Delta\omega$ is the chemical shift difference between the two states, k_{ex} is the exchange rate between the two states and finally τ_{cp} is the delay between 180° pulses in the CPMG pulse sequence. As more 180° pulses are included in the CPMG sequence (figure 1.33), the

effect of R_{ex} is reduced, leading to an exponential decay if exchange is occurring.

Without these extra pulses, R_{ex} will obscure measurements of R_2 . The Carver-Richards equation contains a large number of variables, which potentially provide a lot of information about the exchange process. However if data from only one experiment is being fit with this relation, $|\Delta\omega|$ can not be uncoupled from the populations, p_A and p_B , of each state, so reliable outputs of these values from the fit are not possible. To counter this the experiment is usually performed using at least two B_0 field strengths, as these will change the value of $|\Delta\omega|$ while all other variables should remain fixed.

1.5.6 ZZ-Exchange

ZZ-exchange experiments, also known as EXchange SpectroscopY (EXSY) is an NMR method suitable for the slower dynamics (seconds) where distinct signals are seen for each state as $k_{ex} \ll |\Delta\omega|$, as described in figure 1.32^[120].

During the ZZ-exchange experiment, magnetisation is stored in the direction of the B_0

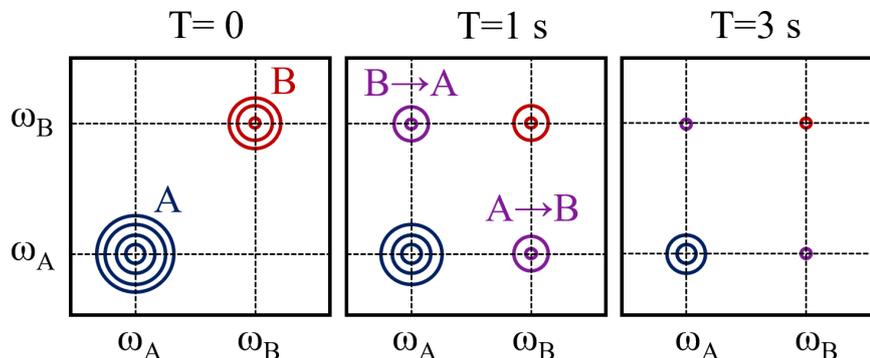


Figure 1.35: Example of the signal from a probe undergoing slow exchange as exchange time is increased during the ZZ-Exchange experiment. The two states A and B produce clear signals in the HSQC spectrum. As the exchange time, T , is increased in the EXSY experiment, exchange occurs from state A to state B and vice versa. This produces cross-peaks for each exchange and reduces the intensity of the main states A and B. If T is increased further, the intensity of all the signals decreases. Adapted from Kleckner 2011^[120].

field, or z-direction, as part of the evolution stage of the pulse sequence. During this time T , if $k_{ex} \geq \frac{1}{T}$, conversion will occur between states A and B. This results in some nuclei being in state B that were initially in state A and vice versa. These conversions during

the time T result in cross-peaks in the spectrum, as shown in figure 1.35. ^1H - ^{15}N nuclei that have transitioned during this time will produce signals that have a chemical shift of their initial state in the Nitrogen dimension, but a chemical shift signal corresponding to their final state in the hydrogen dimension. This gives an indication of pairs of peaks in the HSQC spectrum that correspond to a single set of ^1H - ^{15}N nuclei experiencing slow exchange between two protein states.

1.5.7 Relation of Protein Dynamics to Temperature Measured by Nuclear Magnetic Resonance (NMR)

Localised dynamics that only affect partial regions of the protein can be studied by NMR. For studying proteins from psychrophilic and psychrotropic organisms this is suitable for exploring the the folded energy landscape shown in figure 1.11. Fast dynamics defined by the order parameter usually follow a linear temperature dependance^[128], which incorporates the effect of increasing solvent viscosity at lower temperatures, leading to a change in tumbling rate, and the lowering of thermal energy available for local molecular motions. Chemical exchange between two distinct states, or the relaxation dispersion described in section 1.5.5 can have a more complex temperature dependence. An in depth study of the temperature dependence of Ribonuclease H measured N-H dynamics across multiple time-scales at a range of temperatures^[129]. The temperature dependence of these motions were shown to depend strongly on the structural element the residue was a part of.

1.6 Discussion and Aims

The CSPs from the mesophilic organism *Bacillus Subtilis* and the psychrotropic organism *Psychrobacter* sp. 6 provide a good model system to explore the mechanisms for the function of RNA chaperoning at low temperatures.

Can the effect of nucleic acids binding to the *Bs*-CSP be measured in the protein unfolding force?

The first aim of this work is to establish by SMFS how the binding of ssDNA affects the mechanical unfolding properties of the *Bs*-CSP and this is the focus of chapter 3. There are few studies of the effect of single stranded nucleic acid binding on protein mechanics^[112], and most studies on the effect of ligand binding have found that stabilisation usually involves a decrease in malleability, or increase in spring constant on a previously flexible protein^[105]. The *Bs*-CSP provides a scaffold that already demonstrates mechanical resistance to unfolding^[83].

What are the mechanical properties of the *Bs*-CSP when bound with nucleic acids?

In the nucleic acid bound state, maintained flexibility should be inherent to the protein function as an RNA chaperone^[20]. In particular, the role of the CSP as a response to cold stress implies that this function must be maintained at low temperatures, and a temperature dependent malleability has already been recorded in a hyperthermophilic CSP using SMFS^[7]. By including the nucleic acid bound state in these studies, more insight into the mechanisms that aid function of this particular protein can be attained and these studies are described in chapter 3. With a more general understanding of ligand binding affects on protein mechanics the potential for future protein and nucleic acid based biotechnologies can be explored.

Is the binding affinity of the CSPs to nucleic acids regulated relative to the optimal growth temperature of the organism?

In order to study the bound state, nucleic acid binding studies will be conducted to explore other properties of the binding in chapter 4. Focusing on the temperature dependence of the affinity to ssDNA, *Bs*-CSP and *PB6*-CSP will be used to compare the two proteins adapted to function in different temperature environments. The ssDNA binding of *Bs*-CSP has been studied previously^[58,130] so this will be used as a model for comparison with the *PB6*-CSP.

Can the unbinding of ssDNA from *Bs*-CSP be measured by SMFS?

The possibility of SMFS of the unbinding interaction between *Bs*-CSP and ssDNA is explored in chapter 4 to establish if it is possible to perform such experiments.

Is there a difference in the way localised dynamics of the *PB6*-CSP and *Bs*-CSP change with temperature?

The temperature dependent dynamics of the *Bs*-CSP and *PB6*-CSP are investigated using protein NMR spectroscopy in chapter 5. It is expected that the protein from the psychrotropic organism will display more dynamic behaviour as other cold adapted CSPs have demonstrated^[68,69], but the effect of native temperature conditions on these dynamics have not been explored before.

2

Materials and Methods

2.1 Materials

2.1.1 General Chemicals

Granulated tryptone and agar were purchased from Fisher Scientific and LB Broth Miller granulate was purchased from VWR. Carbenicillin Disodium, diThioThreitol (DTT) and Isopropyl- β -D-ThioGalactoside (IPTG) was purchased from Formedium. Virkon powder, Tris, sodium phosphate, sodium chloride, potassium phosphate, glycerol, chloroform and Sodium Dodecyl Sulfate (SDS) were purchased from Fisher Scientific, USA. 30% (w/v) bisacrylamide was purchased from Severn Biotech. EthyleneDiamineTetraacetic Acid (EDTA)-free Protease Inhibitor Tablets were purchased from Thermo Scientific, USA. DNase I, calcium chloride, urea, tetramethylethylenediamine (TEMED) and all other reagents unless otherwise stated were purchased from Sigma Chemical Company. Ammonium ^{15}N chloride and D-glucose U- $^{13}\text{C}_6$ was purchased from Goss Scientific and D_2O was purchased from Fluorochem.

2.1.2 Molecular Biology Materials

PureYield Plasmid Miniprep, DNA markers for agarose gels and $6 \times$ loading buffer, restriction enzymes with associated buffers, Q5 Site-Directed Mutagenesis Kits and Gel Extraction kits were purchased from Promega, USA or Qiagen, Germany. All synthesised genes and primers were designed and purchased from Eurofins MWG Operon and final plasmids were sequenced by Beckman Coulter Genomics. *E. coli* strains BL21 (DE3) pLysS, XL1-Blue and Sure2 competent cells were purchased from Stratagene.

2.1.3 AFM Materials

MLCT cantilevers and MTFML fluid cells were purchased from Bruker. Custom 4 inch Si wafers coated with a 100 nm gold layer with no titanium adhesion layer were purchased from Platypus Technologies. Magnetic disks, glass slides and stickers for attaching the gold were purchased from Agar Scientific. 76 mm PB doped silicon wafers, single side polish were purchased from IDB Technologies. 0.4 μm pore size syringe driven filters were purchased from GE HEALTHCARE WHATMAN

2.1.4 Buffers Used

All buffers were made up using double distilled H_2O (dd H_2O) and filter sterilised using a 0.4 μm filter.

Component	Final concentration in $50\times$
Tris	2 M
Glacial Acetic Acid	1 M
EDTA	50 mM

Table 2.1: Composition of 50 times Tris-acetate-EDTA (TAE) buffer

Component	Final Concentration
Bromophenol Blue	0.1 % (w/v)
Tris HCl, pH 6.8	50 mM
DTT	100 mM
SDS	2 % (w/v)
Glycerol	10 % (v/v)

Table 2.2: Composition of 6× gel loading buffer

Stock Solution	Final Concentration
1M Tris, pH 8.0	20 mM
5M NaCl	300 mM
2M Imidazole	5 mM
25 % Triton X-100	0.15 %

Table 2.3: Composition of lysis buffer

Stock Solution	Final Concentration
1M Tris, pH 8.0	0.25
5M NaCl	300 mM
2M Imidazole	10 mM

Table 2.4: Composition of wash buffer

Stock Solution	Final Concentration
1M Tris, pH 8.0	0.25
5M NaCl	300 mM
2M Imidazole	250 mM

Table 2.5: Composition of elution buffer

Stock Solution	Final Concentration
1M sodium phosphate, pH 8.0	20 mM
5M NaCl	1 M (High Salt) 50 mM (Low Salt)
EDTA	1 mM

Table 2.6: Composition of high and low salt buffers for ion exchange chromatography

2.2 Molecular Biology

Plasmids for the polyproteins used in Single Molecule Force Spectroscopy (SMFS) experiments were provided by Toni Hoffman and monomers of the two CSPs with a his-tag were provided by Michael Wilson. Genetic alterations were made to these plasmids as

needed for particular experiments. Each of these alterations and the section describing the molecular biology methods used are shown in table 2.7. The his-tag is an extra 6 histidine amino acids engineered onto the N-terminus of the protein for the purification step described in section 2.3.6.1, commonly used in molecular biology. The presence of a his-tag could affect NMR measurements of protein dynamics, so for these measurements the *PB6*-CSP and *Bs*-CSP monomers were prepared with a cleavable his-tag. Monomers containing a cleavage site between the CSP and the his-tag, shown in figure 2.1, were prepared.

The *Bs*-CSP plasmid was mutated to include a new protease cleavage site between the

Plasmid Source	Construct	Description	Genetic Alteration	Section
Provided by Michael Wilson	His-<i>Bs</i>-CSP	Histidine tagged WT CSP from <i>Bacillus Subtilis</i>	NA	NA
	<i>Bs</i> -CSP	CSP with a TEVp recognition site to remove the histidine tag during purification	Insertion of TEVp recognition site	2.2.2
	<i>Cys-Bs</i> -CSP	Histidine tagged CSP containing one Cysteine residue	Substitution replacing one amino acid with Cysteine	2.2.3
Provided by Michael Wilson	<i>PB6</i> -CSP	Histidine tagged WT CSP from <i>Psychrobacter sp. B6</i>	NA	NA
	MBP- <i>PB6</i> -CSP	CSP with a Histidine tag and MBP label. TEVp recognition site to remove the histidine tag and MBP label during purification	Insertion of the <i>PB6</i> -CSP gene into the pMal-c5x plasmid	2.2.5
Provided by Toni Hoffman	(I27-<i>Bs</i>CSP)₃-I27	Histidine tagged chimeric polyprotein containing 3 <i>Bs</i>-CSP and 4 I27 domains	NA	NA

Table 2.7: Table of the plasmids used throughout this project to produce the protein constructs. The source of the original plasmid (highlighted in bold) is shown in the first column and the sections describing the molecular biology used to produce the final plasmid is identified in the final column. Sequences for each construct can be found in tables A.2 and A.3

protein and the his-tag. As this only required the insertion of 15 nucleotides, for the 5 extra amino acid residues, the plasmid was engineered using an insertion Q5 mutagenesis.

PB6-CSP was expressed fused to Maltose Binding Protein (MBP) to increase protein expression in minimal medium. MBP is a highly soluble protein that expresses well in *E. coli* so is often used as a tag to increase expression and solubility^[131,132]. After purification of the MBP-*PB6*-CSP construct the MBP and his-tag could be removed. For this construct the *PB6* gene was inserted into the pMal-c5x plasmid, which contains the MBP gene.

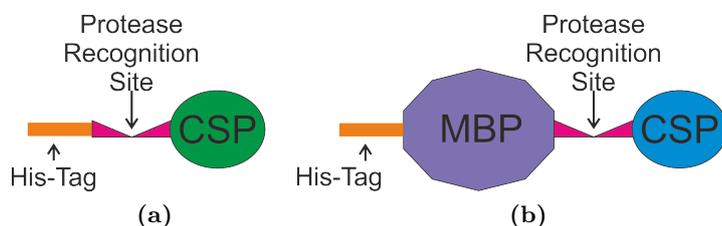


Figure 2.1: Diagram of the (a) new *Bs*-CSP construct and (b) *PB6*-CSP construct for NMR experiments.

Another variant was used for the single molecule force spectroscopy experiments measuring the unbinding force between dT7 ssDNA and the *Bs*-CSP. This required a *Bs*-CSP protein with a single cysteine residue to tether the protein to the surface.

2.2.1 Polymerase Chain Reaction (PCR)

Each PCR reaction was carried out in triplicate in thin walled PCR tubes containing reagents, described in table 2.8.

Component	Volume Added	Final Conc.
Q5 Hot Start High-Fidelity 2× Master Mix	12.5 μL	1 ×
10 μM Forward Primer	1.25 μM	0.5 μM
10 μM Reverse Primer	1.25 μM	0.5 μM
Template DNA	1 μM	1-25 ng
Nuclease-free water	9 μL	

Table 2.8: Reagents used in PCR reactions for insertion and substitution mutations

Step	Temperature (°C)	Time	Cycles
Initial 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

Table 2.9: Thermocycling conditions used for PCR. Annealing temperature was 5 °C below calculated melting temperature of the primers.

2.2.2 Insertion of the TEVp Restriction Site to the *Bs*-CSP Gene

PCR was used to insert DNA encoding the recognition site for Tobacco Etch Virus protease (TEVp) into the *Bs*-CSP plasmid. TEVp has a specific cleavage site, ENLYF\Q, where it cleaves protein constructs. This site was inserted into the *Bs*-CSP protein between the the his-tag and desired monomer using the Q5 Site-Directed Mutagenesis Kit from NEB labs. Primers used for this reaction are given in table 2.10

Primer	Sequence
Forward	tactttcagAGCCTGGAAGGCAAAGTG
Reverse	caggttttcGTGATGGTGTGATGGTGTGATGC

Table 2.10: Primers used for the mutation of the *Bs*-CSP plasmid. Uppercase letters show the target specific primer, designed to overlap with the vector.

2.2.3 Insertion of a Cysteine Residue into the *Bs*-CSP Gene

In order to tether the *Bs*-CSP to the surface for SMFS protein-DNA experiments, a cysteine substitution mutation was generated to provide a site where the PolyEthylene Glycol (PEG) linker could be attached through a maleimide group. The site was chosen to be on the opposite surface of the protein to the ssDNA binding site as identified in the crystal structure^[62], shown in figure 2.2. Sequence of the mutated protein with Histidine tag:

MHHHHHSSLEGKVKWFNSEKGFIEVEGQDDVVFVHFSAIQGEGFKTLEEGQ
AVCFEIVEGNRGPQAANVTKEA

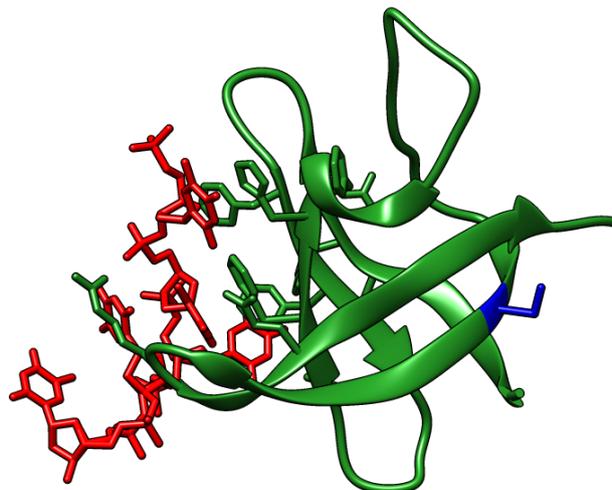


Figure 2.2: Crystal structure of *Bs*-CSP-dT7 complex (protein in green and ssDNA in red), PDB code 2ES2^[62], showing the residue (blue) mutated from a serine to a cysteine residue

Primer	Sequence
Forward	CCAGGCGGTGtGCTTTGAAATTG
Reverse	CCTTCTTCCAGGGTTTTAAAG

Table 2.11: Primers used for the replacement of a serine residue with a cysteine in the *Bs*-CSP

2.2.4 Agarose Gel Electrophoresis

Agarose gels provide a matrix through which DNA (which has an overall negative charge) can move by electrophoresis. The DNA movement will be impeded to a greater extent for longer strands, leading to the separation of DNA by length. This mechanism is used to separate DNA so that the length of DNA plasmids can be measured and to separate components that have been cut from other plasmids. Gels were made up using 2.25 g of agarose added to 150 ml of TAE buffer (table 2.1). This was microwaved in a conical flask for 1 minute with gentle mixing. After cooling, 15 μ L of ethidium bromide was

added from a 10 mg/ml stock solution and then the final mixture was poured into a cast. Once the gel was set, the comb was removed and the gel was placed in into the running tank and covered with 1 times TAE buffer. The DNA samples and markers had gel loading buffer, see table 2.2, added with a 5:1 ratio. The gel was run at 100 V for about an hour until the bands were distinct. The resulting gel was then imaged using the UV light and camera.

If part of the DNA needed to be extracted from the agarose gel this was cut out using a scalpel. The DNA was then purified using the Quiagen Gel Extraction Kit.

2.2.5 Insertion of the *PB6*-CSP Gene into the pMal-c5x Plasmid

The gene for the *PB6*-CSP was provided by Michael Wilson. To increase protein expression in minimal medium and solubility during purification, an MBP tag was used. This necessitated the introduction of a protease cleavage site to remove the MBP from the CSP. PCR was used to amplify the *PB6*-CSP gene from the original plasmid and to introduce restriction sites designed to insert into the MBP expression plasmid pMal c5x. Using primers described in table 2.12 a PCR amplification using Vent polymerase was performed with reagents shown in table 2.13. The temperature cycle was the same as shown in table 2.9

Primer	Sequence
Forward	ataccattg <u>GGATCC</u> GATAAAAGTGGAAGGCACCGTG
Reverse	AGGCGGAACAGATTGAAGCGATT <u>TAA</u> GAATTCataccattg

Table 2.12: Primers used for the amplification of the *PB6*-CSP gene, with EcoRI and BamHI recognition sites underlined

The products of the amplification PCR of the *PB6*-CSP gene were run on an agarose gel, shown in figure 2.3. The bands corresponding to the *PB6*-CSP gene, expected to be 219 bp long, are clear in all three reactions, just below the marker for 250 bp. This band was cut from the gel and the insert extracted using the Quiagen Gel Extraction Kit.

Once the *PB6*-CSP insert had been extracted from the gel, the insert and the pMal-c5X plasmid were digested with BamHI and EcoRI enzymes. Restriction digests

Component	Volume Added μL	Final amount
Plasmid DNA 25 ng/ μL	2	50 ng
Forward Primer	2	50 ng
Reverse Primer	2	50 ng
MgSO ₄	0-6	0-6 mM
DMSO	1	
Vent Polymerase	0.5	1U
4 \times PCR Buffer	25	1 \times
Nuclease-free Water	To made up 100 μL total	

Table 2.13: Reagents used in PCR reactions for amplification of *PB6-CSP* gene

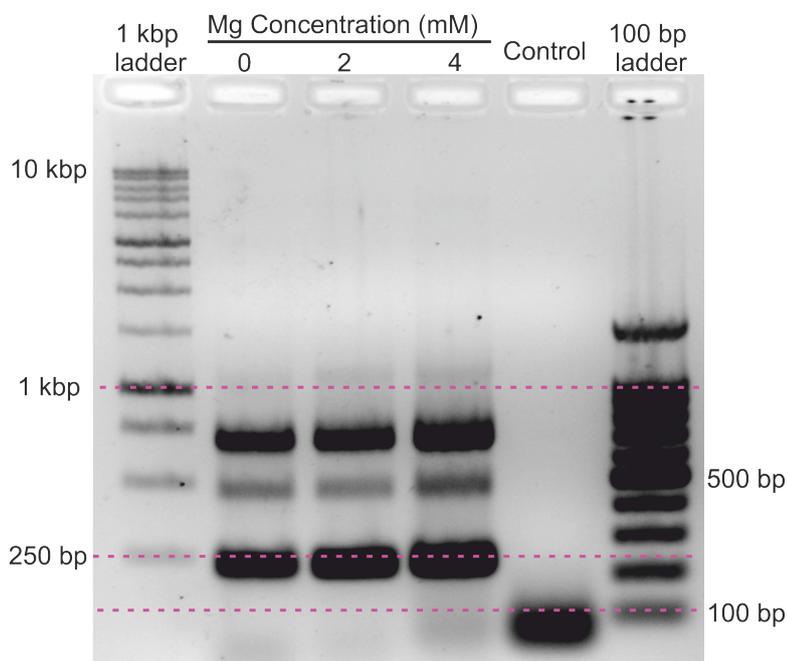


Figure 2.3: 1.5 % (w/v) agarose gel of the products of the PCR reaction amplifying the *PB6-CSP* gene from the provided vector and adding 5' EcoRI and 3' BamHI recognition sites. Three concentrations of magnesium sulphate were used to optimise the yield of the the PCR reaction. The control contains no plasmid.

were carried out with the extracted insert and with 1 mg of the pMal-c5x plasmid. These were mixed with 1 unit of each enzyme in a reaction volume of 10 μL 1 \times NEBuffer 3.1. Control reactions were also carried out with BamHI and EcoRI only and neither enzyme. These mixtures were incubated at 37 $^{\circ}\text{C}$ for 1 hour. The vector was then dephosphorylated to prevent the two overhanging ends of the digested strand self-ligating. To do this, the PCR mix from the restriction digest of the plasmid was mixed with 1:10 volume of 10 \times *Antarctic Phosphatase* reaction buffer and 1 μL of the

Antarctic Phosphatase enzyme, in a 40 μ L total reaction volume. This mixture was incubated at 37 °C for 15 minutes and then 65 °C for 5 minutes to heat inactivate the enzyme. The products of this reaction and the restriction digest of the insert were run on an agarose gel as shown in figure 2.4.

After cutting out the vector and insert and extracting from the gel, a ligation reaction

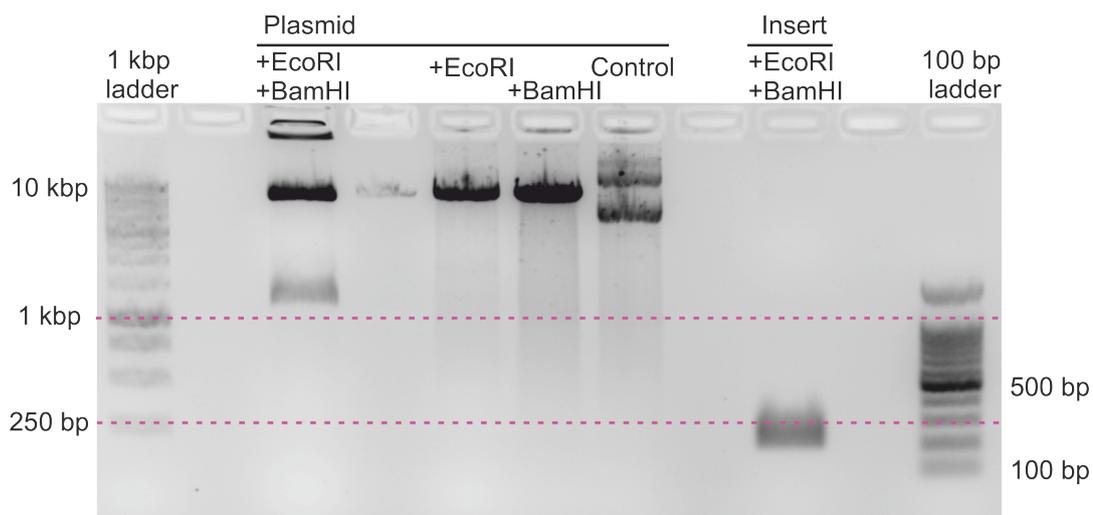


Figure 2.4: 1.5 % (w/v) agarose gel of the products of the digestion with enzymes EcoRI and BamHI

was performed using the T4 DNA Ligase. In a total reaction volume of 10 μ L of 1 \times ligase buffer, the vector DNA was mixed with a molar excess of the insert and 1 μ L of T4 DNA ligase. A control with only the vector DNA, that did not contain the insert DNA, was also prepared. These were held at 15 °C for 3 hours before being left at 5 °C overnight. The products were then transformed into SURE-2 competent cells and spread onto agar plates for overnight incubation at 37 °C.

Colonies were observed on ligation plates but not on the control plate. Five colonies were selected, grown up to 5 ml and plasmids purified using the PureYield Plasmid Miniprep. These were tested by another restriction digestion with BamHI and EcoRI enzymes. The agarose gel confirming the new plasmid is shown in figure 2.5. A band appears around 219 bp identifying the *PB6-CSP* gene has been separated from the plasmid by the restriction enzymes, confirming that it has been correctly inserted. The

plasmid was also sent for sequencing to confirm the insertion of the *PB6*-CSP gene into the pMal-c5X vector, primers used for this sequencing are shown in the appendix, table A.1.

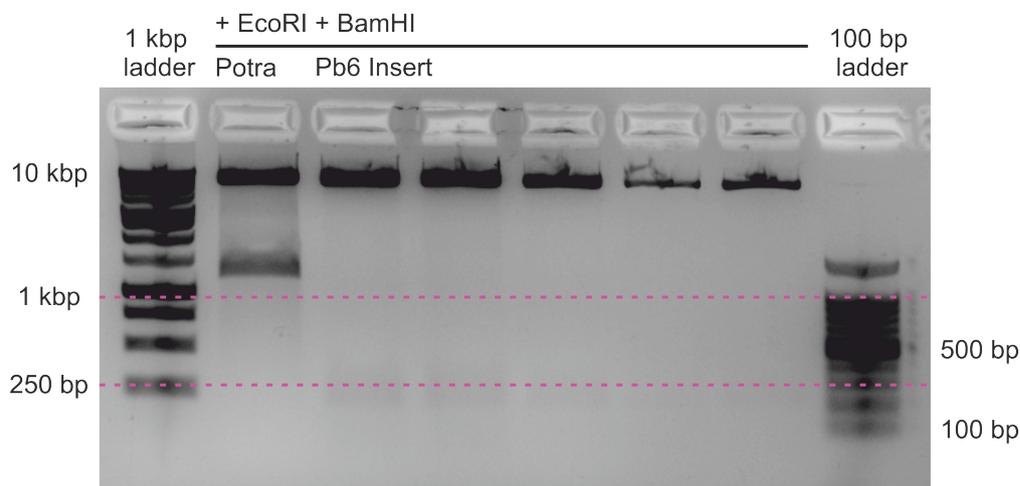


Figure 2.5: 1.5 % (w/v) agarose gel of the products of the digestion with enzymes EcoRI and BamHI of plasmid with the *PB6*-CSP inserted into the pMal-c5x vector. pMal-c5x with a potra domain (1.2 kbp) insert was used as a control.

2.3 Production and Purification of the Cold Shock Proteins

2.3.1 Cell Culture Medium

Proteins and polyproteins were produced using Lysogeny Broth (LB), described in table 2.14, with 100 $\mu\text{g ml}^{-1}$ carbenicillin added after autoclaving. For the NMR experi-

Component	Quantity added per Litre LB (g)
Bactotryptone	10
Yeast Extract	5
NaCl	10

Table 2.14: Composition of Lysogeny Broth (LB) used as growth medium for cell culturing

ments, proteins were produced using M9 Minimal medium, table 2.16 enriched with ^{15}N ammonium salts and also ^{13}C glucose for the backbone assignment experiments.

Component	Quantity added per Litre M9 salts (g)
Na ₂ HPO ₄ .H ₂ O	34
KH ₂ PO ₄	15
NaCl	2.5
¹⁵ NH ₄ Cl	5

Table 2.15: Composition of 5× M9 minimal salts.

Stock Solution	Final Concentration	
5×M9 salts	100 ml/500 ml (1×)	
1 M MgSO ₄	2 mM	
1 M CaCl ₂	0.1 mM	
20% Glucose (w/v)	0.4 %	<i>Added post-autoclave</i>
100 mg/ml Carbenicillin	100 µg/ml	

Table 2.16: Composition of M9 Minimal Medium. For ¹³C enriched protein, labelled glucose was used.

2.3.2 Large Scale Growth

After transformation into BL21 (DE3) pLysS competent cells, samples were plated on LB agar plates supplemented with carbenicillin antibiotic. Colonies were selected and used to inoculate 15 ml of LB medium with 100 µgml⁻¹ carbenicillin for overnight incubation at 37 °C. 10 litres of LB medium was prepared and incubated at 37 °C. This medium was then inoculated with the culture and grown at 37 °C to an OD₆₀₀ of 0.6 before induction with IPTG (final concentration 1mM) and the reduction of the temperature to 26 °C. These cultures were then harvested the next morning by centrifugation at 6500 rpm and pellets were stored in the freezer.

2.3.3 Large Scale Growth in ¹⁵N and ¹³C Labelled Minimal Medium

Proteins produced for NMR measurements requiring ¹⁵N labelling were grown in 5 or 10 litres of M9 minimal medium. This medium was prepared using 10g of ¹⁵N ammonium sulphate and incubated in 2 litre flasks each containing 500 ml. After transformation into BL21 (DE3) pLysS competent cells, samples were plated on LB agar plates supplemented with 100 µgml⁻¹ carbenicillin and incubated overnight at 37 °C. Single colonies were used to inoculate 15 ml of LB medium with 100 µgml⁻¹ carbenicillin for overnight incubation

at 37 °C. 3 ml of this was used to seed 200 ml of M9 minimal medium, which was again incubated at 37 °C overnight. This culture was used to inoculate the 10 litres of labelled medium. This was grown at 37 °C to an OD₆₀₀ of 0.6 before induction with IPTG (final concentration 1 mM) and the reduction of the temperature to 26 °C. These cultures were then harvested the next morning by centrifugation at 6500 rpm and pellets were stored in the freezer.

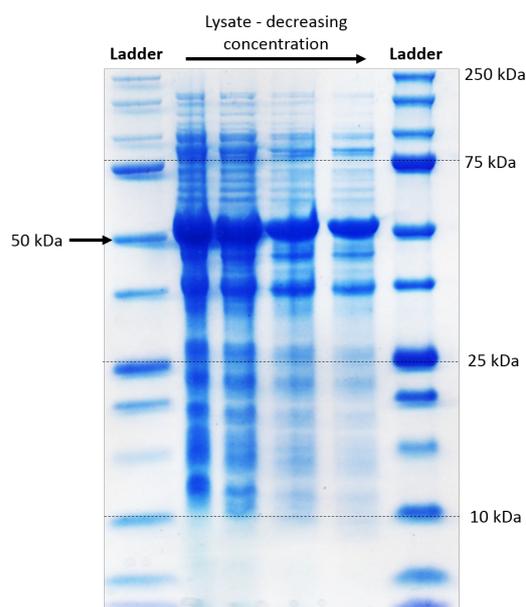


Figure 2.6: SDS-PAGE of the lysate used in growing double labelled MBP-*PB6*-CSP . With an expected mass of 50 kDa

2.3.4 Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

Expression was checked using SDS-PAGE to detect proteins of the expected size in the lysate. SDS-PAGE was also used to check the purity of the protein sample after each purification step.

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

Table 2.17: Compositions of the Sodium Dodecyl Sulfate PolyAcrylamide Gel. Amounts make two gels with the lower Resolving gel for separating particles of different size and the Stacking gel for aligning samples above.

2.3.5 Lysing the Cells

After defrosting, the cells were resuspended in lysis buffer (table 2.3) at a ratio of 10 ml of buffer for each litre of growth medium. Once homogenised, the solution was sonicated using a tip sonicator at 40% power for 10 repeats of 30 s, before being centrifuged at 16,000 rpm to pellet insoluble cell components. The remaining solution was then filtered using a 0.22 μm (0.4 μm for polyproteins) syringe filter to clarify the lysate prior to chromatography.

2.3.6 Columns used for Protein Purification

2.3.6.1 Nickel Affinity Chromatography

The filtered lysate solution was applied to a 5 ml Histrap FF Nickel-NTA affinity column (GE Healthcare, Sweden) pre-equilibrated with wash buffer (table 2.4). In this type of chromatography the histidine tag on the protein constructs binds to the resin (figure 2.7). Wash buffer (table 2.4) was run through to clear any unbound components of the cell lysis, before a stepped gradient of elution buffer (table 2.5) was applied to elute the bound protein. Fractions containing protein were collected and buffer exchanged into low salt buffer for the anion exchange step.

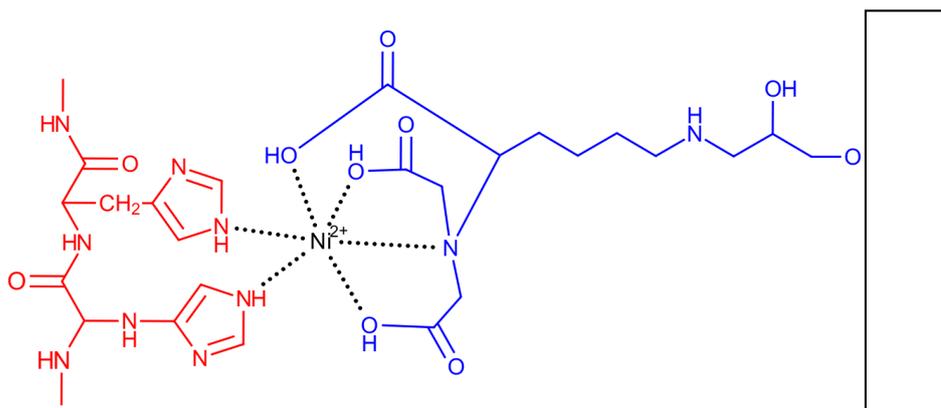


Figure 2.7: Diagram showing the chemical interaction between histidine residues on the protein of interest (red) and the Nickel loaded column resin (Blue).

2.3.6.2 Anion Exchange Chromatography

A 6 ml ResourceQ anion exchange column (GE Healthcare, Sweden) was used to remove DNA bound to the protein. Anion exchange chromatography uses a positively charged material to bind the negatively charged particles in the solution. A gradient of buffer ionic strength (table 2.6) is applied to remove particles and separate them by ionic charge, as shown in figure 2.8. Weakly bound components elute from the column at lower salt concentrations while tightly bound components, such as negatively charged DNA, stick until a high salt concentration counters the electrostatic interactions.

2.3.6.3 Size Exclusion Chromatography

For the final purification step, a gel filtration column was used to separate any other particles, smaller or larger than the protein of interest. This technique uses porous beads to slow down the passage of smaller particles while larger particles pass through with relative ease, as shown in figure 2.9. The column was equilibrated with 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 4 mM DTT. 4.5 ml of highly concentrated protein solution was injected onto the column. Buffer was run through the column at a rate of 3 ml min^{-1} until the protein was eluted and collected, separate from contaminants of different size.

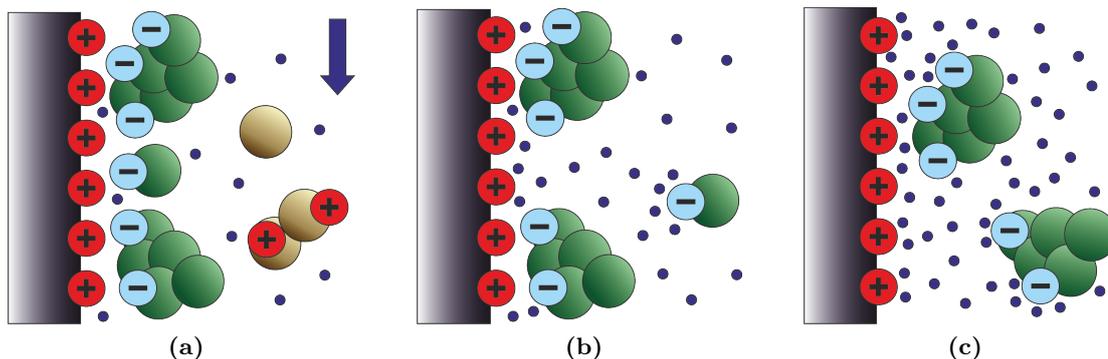


Figure 2.8: Diagram of the anion exchange column separating particles of differing electrostatic charge by application of an ionic strength gradient. (a) At low ionic strength buffer, the negatively charged particles bind to the column while neutral or positive charges flow through. (b) As buffer ionic strength increases particles with a lesser negative charge are displaced and flow through column. (c) At high ionic strength particles with strong negative charge are removed from the column.

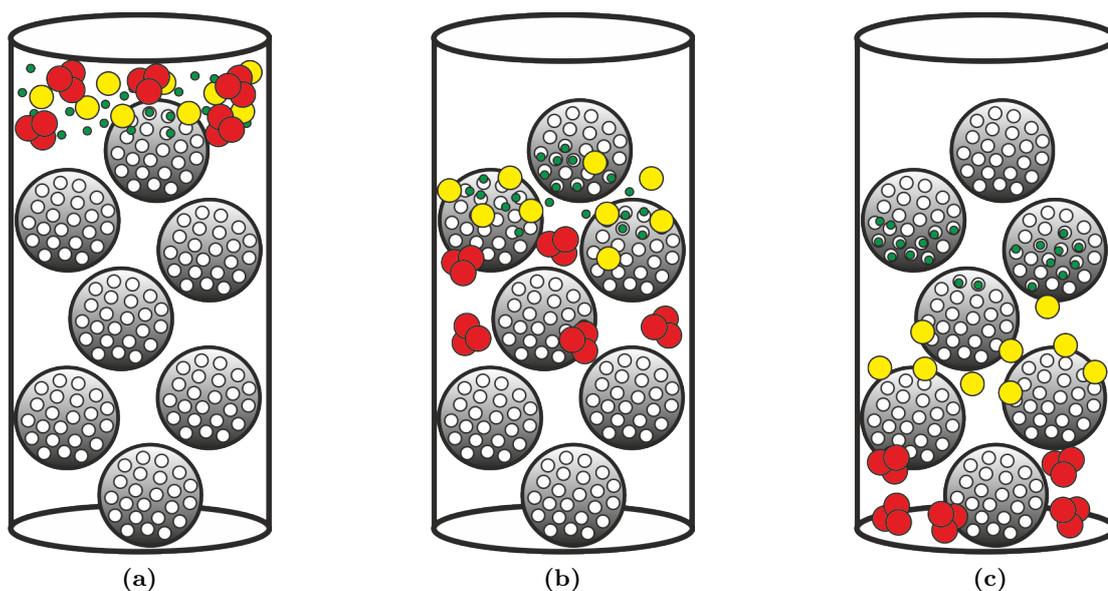


Figure 2.9: Diagram of the size exclusion column separating particles by size. (a) Sample is injected onto the column which is filled with porous beads. (b) The flow is moving from top to bottom and the larger particles (red) are passing through easily while small particles (green) are slowed down by getting stuck in the pores of the beads. (c) The largest particles leave the column to be collected while smaller particles are being separated according to relative size.

2.3.7 Cleavage using Tobacco Etch Virus protease (TEVp)

TEVp was produced using a plasmid provided by Robert Schiffrin and methods described in Blommel 2007^[133]. The plasmid was transformed into *E. Coli* BL21-Codonplus RIPL cells (Stratagene) according to the Stratagene manual. Transformed cells were grown overnight on agar plates supplemented with 50 μgml^{-1} kanamycin. Single colonies were used to inoculate 100 ml of LB medium (with kanamycin at 50 μgml^{-1}) which was then incubated at 37 °C overnight. This culture was used to inoculate 10 litres of LB medium (with kanamycin at 50 μgml^{-1}) which was incubated at 37 °C. When $\text{OD}_{600} = 0.6$, expression was induced with IPTG at a final concentration of 1mM. Expression was allowed for 4 hours at 30 °C, then the culture centrifuged at 12,000 rpm. The protease was then purified with Ni affinity chromatography (section 2.3.6.1) and size exclusion chromatography (section 2.3.6.3) in buffer containing 25 mM sodium phosphate, 0.2 M NaCl 10 % glycerol and 5 mM β -mercaptoethanol. The resulting protease solution was stored at -80 °C.

For the cleavage digestion, the CSP solution collected from the anion exchange step (section 2.3.6.2) was concentrated down to a volume of 4.5 ml. 0.1 mg of TEVp was added to the solution which was incubated at room temperature for 20 minutes and then left overnight at 4 °C.

The TEVp, undigested protein and cleaved MBP were separated from the CSP during the size exclusion step (section 2.3.6.3). The *Bs*-CSP had only the his-tag removed by the TEV protease digestion, which was too similar in size to the protein to be resolved by size exclusion. The cleaved his-tag and undigested *Bs*-CSP were removed by Ni affinity chromatography (section 2.3.6.1).

2.3.8 Purification of Each System

Examples from the purification of each type of protein are described here, with sections corresponding to each construct identified in table 2.18. The purification of the polypro-

tein (I27-*Bs*-CSP)₃-I27, the *Bs*-CSP monomer and the *PB6*-CSP monomer all include nickel affinity chromatography (section 2.3.6.1), the anion exchange (section 2.3.6.2) and the size exclusion chromatography step (section 2.3.6.3). The purification of the *PB6*-CSP and the *Bs*-CSP when being prepared for NMR measurements includes digestion with the TEVp. Example chromatograms and corresponding SDS-PAGE are shown for each step.

Construct	Variation	Experimental Purpose	Purification Section
<i>Bs</i> -CSP Monomers	His- <i>Bs</i> -CSP Cys- <i>Bs</i> -CSP <i>Bs</i> -CSP	Binding studies SMFS unbinding Binding and NMR Studies	2.3.8.2 page 78 2.3.8.2 page 78 2.3.8.3 page 82
<i>PB6</i> -CSP Monomer	<i>PB6</i> -CSP	Binding and NMR studies	2.3.8.4 page 82
<i>Bs</i> -CSP Polyprotein	(I27- <i>Bs</i> CSP) ₃ -I27	SMFS unfolding	2.3.8.1 page 75

Table 2.18: Table of the main constructs used throughout this project with the sections describing the purification of each.

2.3.8.1 (I27-*Bs*CSP)₃-I27

The lysate containing the polyprotein (I27-*Bs*CSP)₃-I27 was applied to the nickel column as described in section 2.3.6.1. The resulting chromatogram and SDS-PAGE are shown in figure 2.10. The expected size of the polyprotein construct is 66 kDa, and a clear band is seen in all 3 peaks of this size (figure 2.10b). Fractions associated with each of these peaks were pooled and concentrated in low salt buffer for the anion exchange step. 4.5 ml of the protein solution, collected from the fractions corresponding to peaks seen in the chromatogram (figure 2.10) was loaded onto the anion exchange column. A gradient of increasing ionic strength was then applied. As the (I27-*Bs*CSP)₃-I27 has a lower calculated pI (4.97) than that expected for most nucleic acids^[134], the protein is expected to be eluted before the DNA. The separation of components by anion exchange is clear (figure 2.11). The first peak corresponds to contaminants that did not bind to the anion exchange column, the second is the (I27-*Bs*CSP)₃-I27 as shown in the SDS-PAGE by a

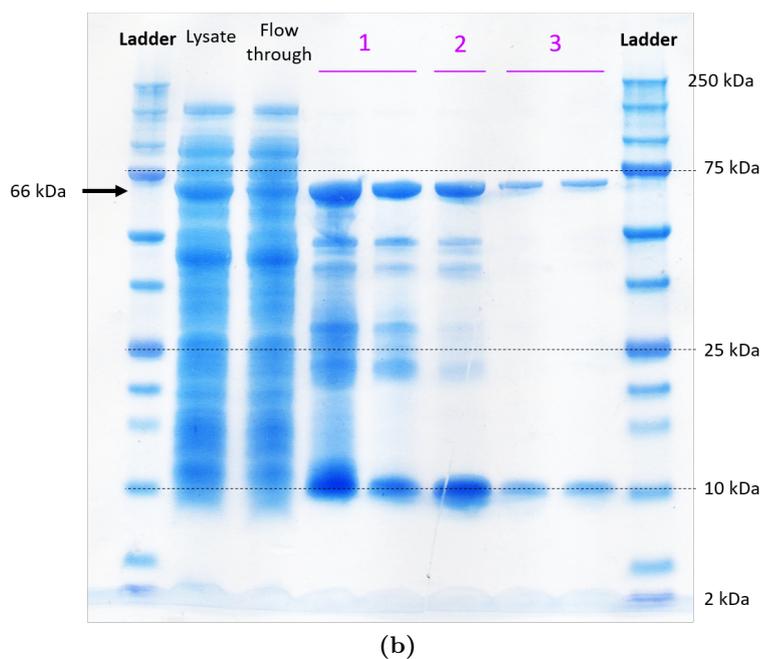
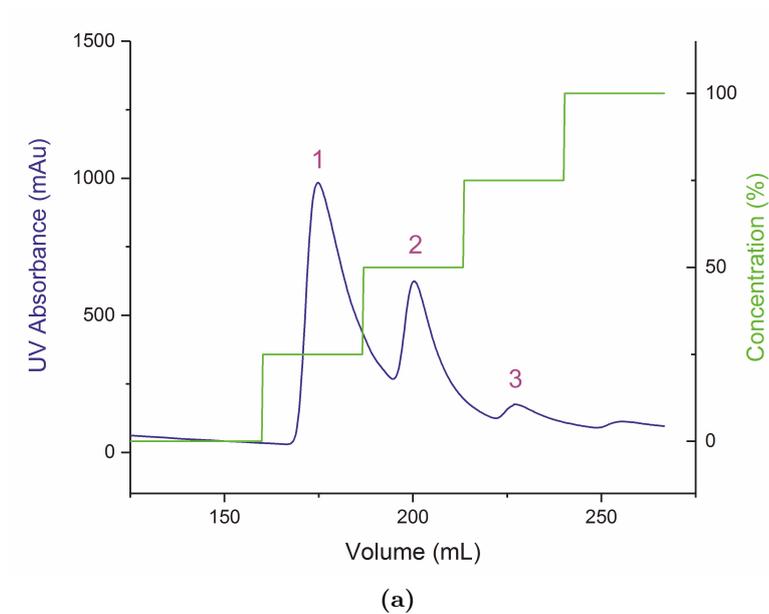
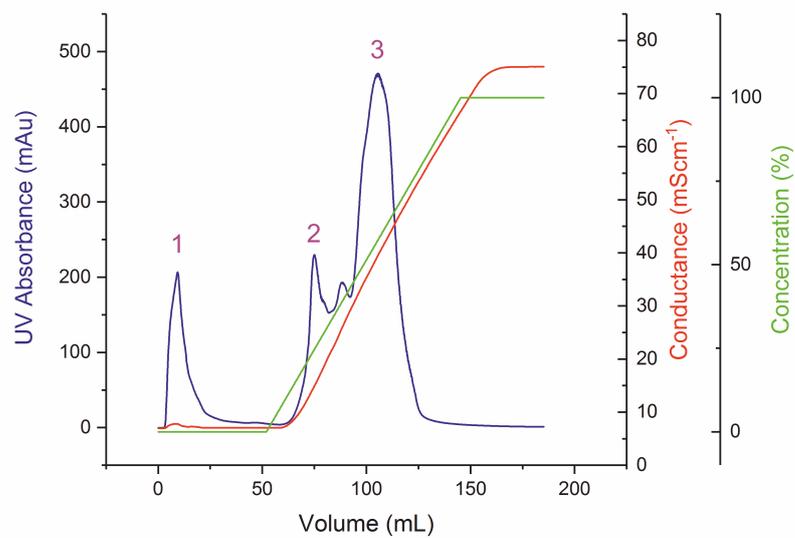
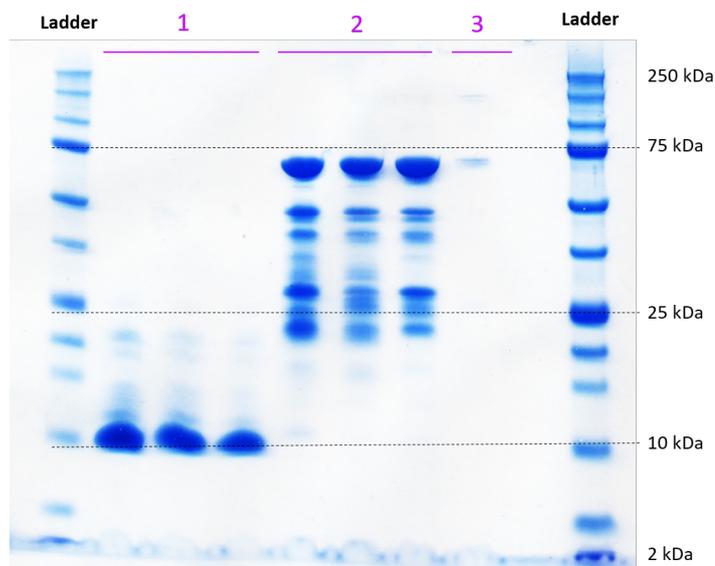


Figure 2.10: (a) Chromatogram of elution of His-tagged polyprotein by an increase in imidazole concentration, by increasing the concentration of elution buffer (green) in steps, after initial loading of column with lysate (not shown). (b) SDS-PAGE of peaks seen in the chromatogram. Bands of the expected size, 66 kDa, are seen in all three peaks so fractions associated with the peaks 1, 2 and 3 were pooled.

band at 66 kDa. The final peak has a large UV absorbance but does not appear on the SDS-PAGE, indicating this peak is caused by the DNA. Fractions from the second peak were collected and concentrated for the next step.



(a)



(b)

Figure 2.11: (a) Chromatogram of protein being eluted from the anion exchange column by an increase in concentration of high salt buffer (green), which increases the conductance (red). Peak 1 corresponds to the sample component that did not bind to the anion exchange column. The increasing concentration of high ionic strength buffer releases the protein of interest (peak 2) followed by the DNA peak (peak 3). (b) SDS-PAGE of peaks seen in the chromatogram. Only peak 2 results in bands of the expected size for the polyprotein, 66 kDa. There is no band associated with peak 3 despite a high UV absorbance, indicating this is the peak corresponding to the DNA.

The final step in purification of the polyprotein (I27-*B_s*CSP)₃-I27 was the size exclusion chromatography step. The chromatogram and SDS-PAGE from this step are shown in

figure 2.12. The first peak produces a band in the SDS-PAGE where one would be

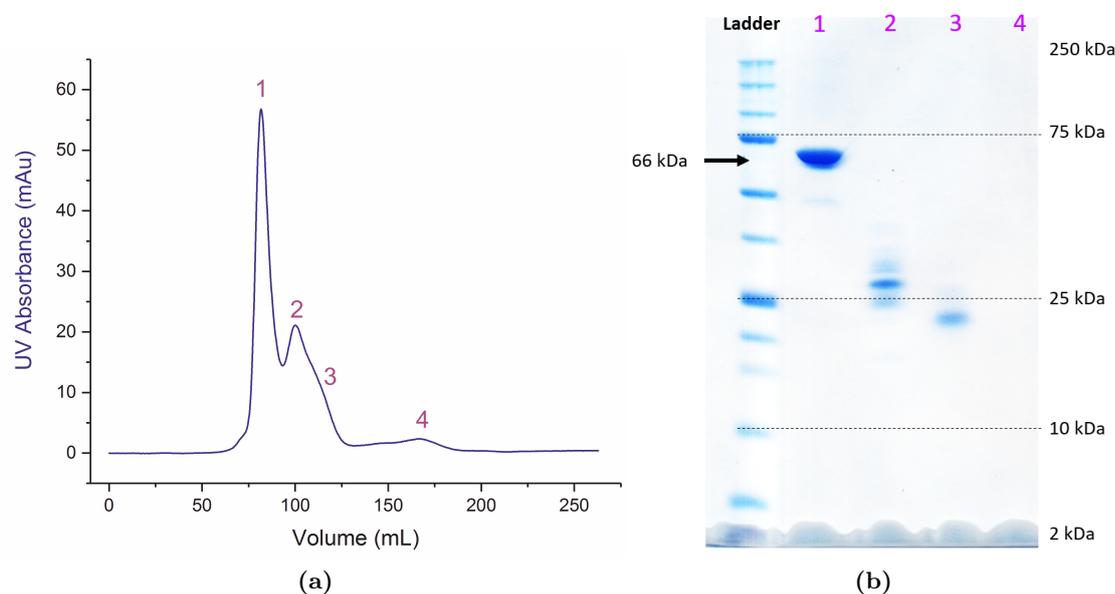


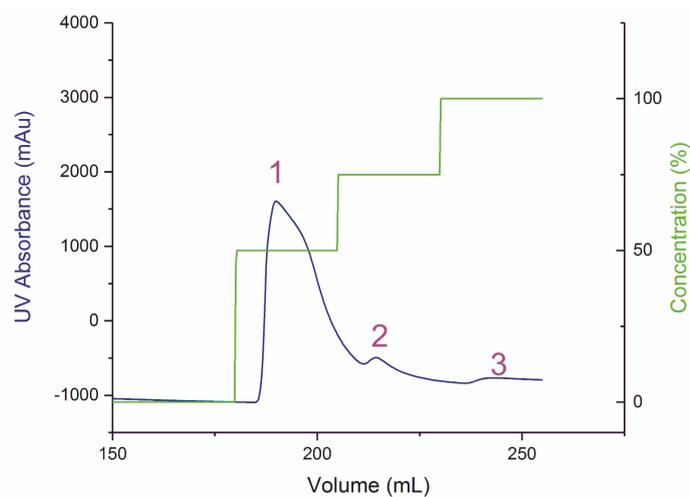
Figure 2.12: (a) Chromatogram of size exclusion column. (b) SDS-PAGE of peaks seen in the chromatogram. The band corresponding to the first peak is of the expected size for the $(I27-BsCSP)_3-I27$ at 66 kDa.

expected for the pure $(I27-BsCSP)_3-I27$ polyprotein. Fractions corresponding to this peak were pooled.

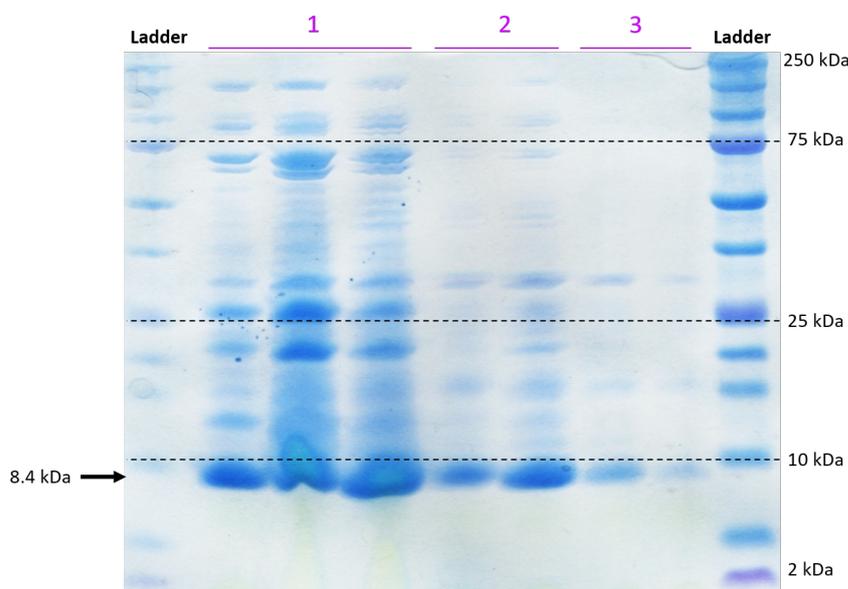
2.3.8.2 *Bs*-CSP monomer

Monomers of the *Bs*-CSP were prepared the same way. Here cysteine mutated *Bs*-CSP (*Cys-Bs*-CSP) was purified using the same procedure as established for the wild type *Bs*-CSP. The lysate of the *Cys-Bs*-CSP was applied to the nickel column as described in section 2.3.6.1. The resulting chromatogram and SDS-PAGE are shown in figure 2.13. The expected size of the *Cys-Bs*-CSP with the his-tag is 8.4 kDa, and a clear band is seen in all 3 peaks of this size (figure 2.13b). Fractions associated with each of these peaks were pooled and concentrated in low salt buffer for the anion exchange step.

4.5 ml of the protein solution, collected from the fractions corresponding to peaks seen in the chromatogram (figure 2.13) was loaded onto the anion exchange column. A gradient of increasing ionic strength was then applied. As the *Cys-Bs*-CSP has a higher



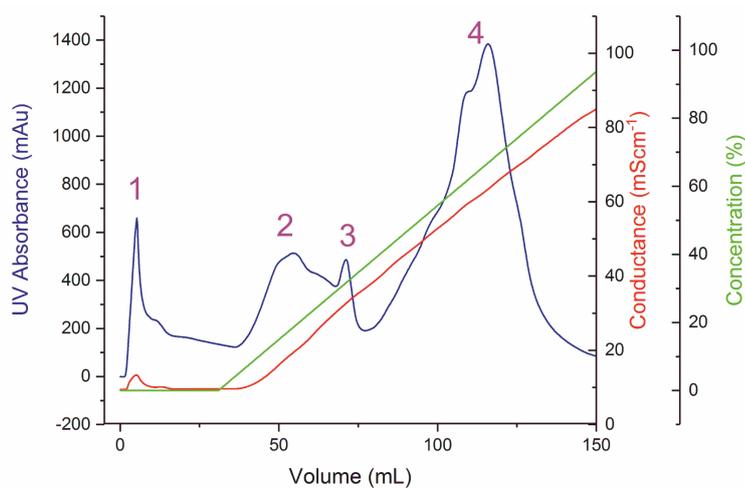
(a)



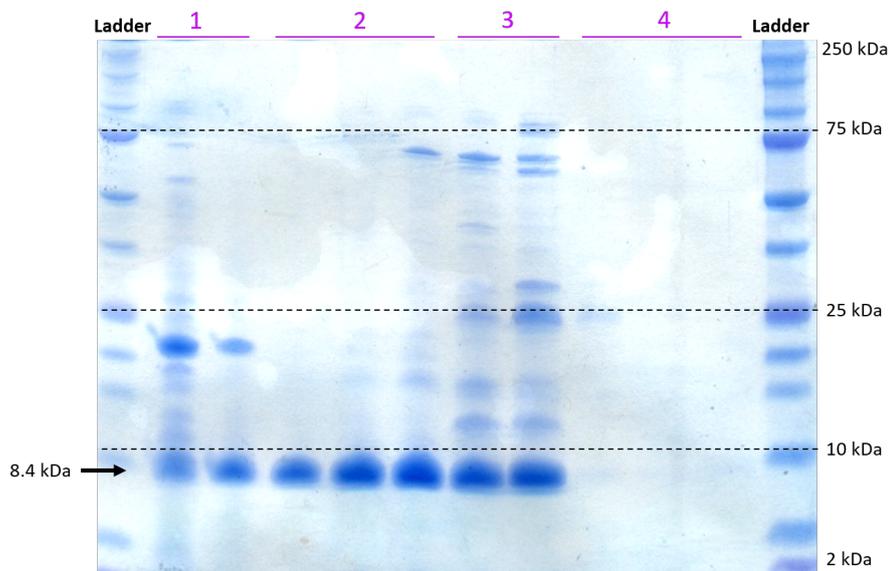
(b)

Figure 2.13: (a) Chromatogram of elution of His-tagged *Bs*-CSP monomer by an increase in elution buffer concentration (green) in steps, after initial loading of column with lysate (not shown). (b) SDS-PAGE of peaks seen in the chromatogram. Bands of the expected size, 8.4 kDa, are seen in all three peaks so fractions associated with the peaks 1, 2 and 3 were pooled.

calculated pI (5.53) than that of DNA, the protein is expected to be eluted before the DNA. The separation of components by anion exchange is clear (figure 2.14). The first peak corresponds to contaminants that did not bind to the anion exchange column, the second and third are the Cys-*Bs*-CSP as shown in the SDS-PAGE by a band at 8.4



(a)



(b)

Figure 2.14: (a) Chromatogram of protein being eluted from the anion exchange column by an increase in concentration of high salt buffer (green), which increases the conductance (red). Peak 1 corresponds to the components of the sample that did not bind to the anion exchange column. The increasing concentration of high ionic strength buffer releases the protein of interest (peaks 2 and 3) followed by the DNA peak (peak 3). (b) SDS-PAGE of peaks seen in the chromatogram. Peaks 2 and 3 cause bands of the expected size for the Cys-Bs-CSP of 8.4 kDa. There is no band associated with peak 4 despite a high UV absorbance, indicating this is the peak corresponding to the DNA.

kDa. The final peak has a large UV absorbance but does not appear on the SDS-PAGE, indicating this peak is caused by the DNA. Fractions from the second and third peak were collected and concentrated for the next step.

The final step in purification of the Cys-*Bs*-CSP was the size exclusion chromatography step. The chromatogram and SDS-PAGE from this step are shown in figure 2.15. The

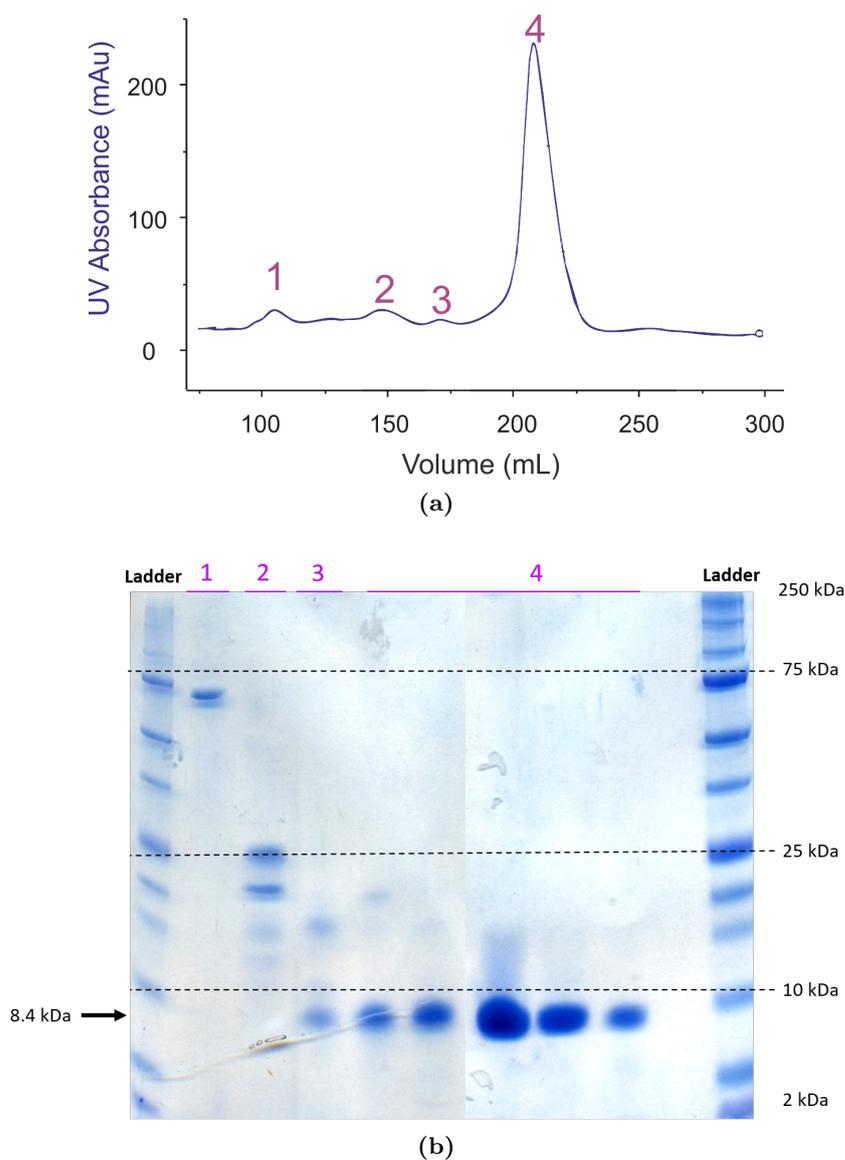


Figure 2.15: (a) Chromatogram of size exclusion column. (b) SDS-PAGE of peaks seen in the chromatogram. The band corresponding to the fourth peak is of the expected size for the Cys-*Bs*-CSP at 8.4 kDa.

fourth peak produces a band in the SDS-PAGE where one would be expected for the pure Cys-*Bs*-CSP monomer. Fractions corresponding to this peak were pooled.

2.3.8.3 *Bs*-CSP monomer for NMR relaxation Studies

In the case of the *Bs*-CSP monomer prepared for NMR measurements, with the removable his-tag, a cleavage digestion was performed with the TEVp (section 2.3.7). Products of this digestion were separated from the *Bs*-CSP monomer by application of nickel resin beads. This removed the TEVp, the cleaved his-tag and any undigested *Bs*-CSP with the his-tag still remaining.

2.3.8.4 *PB6*-CSP

The lysate of the *PB6*-CSP was applied to the nickel column as described in section 2.3.6.1. The resulting chromatogram and SDS-PAGE are shown in figure 2.16. The

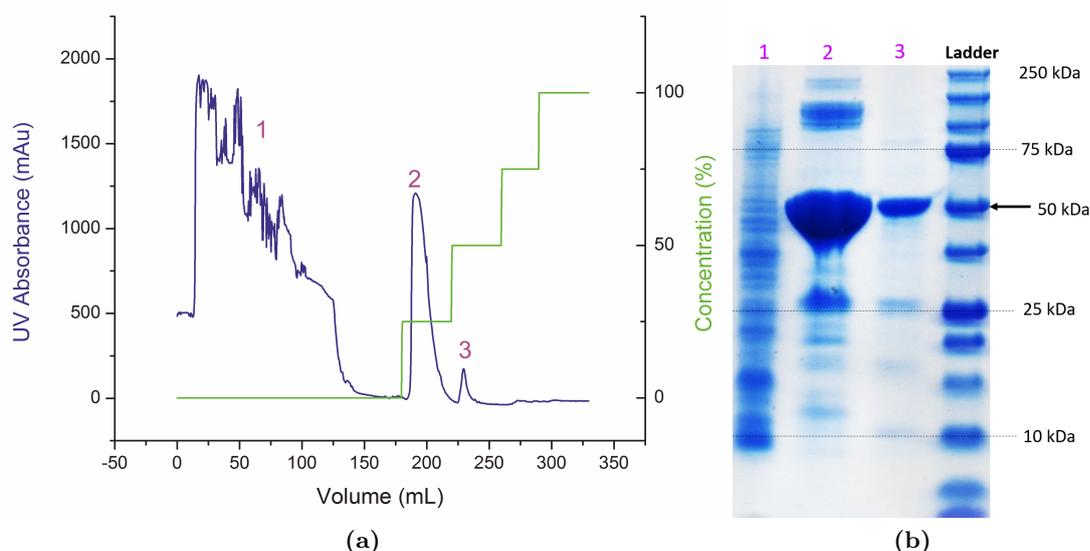


Figure 2.16: (a) Chromatogram of elution of His-tagged MBP-*PB6*-CSP by an increase in elution buffer in steps, with initial loading of column with lysate (peak 1). (b) SDS-PAGE of peaks seen in the chromatogram. Bands of the expected size, 50 kDa, are seen in peaks 2 and 3, while the lysate that did not bind the column in peak 1 does not show a band of this size, indicating that the majority of the MBP-*PB6*-CSP construct has been eluted in peaks 2 and 3.

expected size of the MBP-*PB6* construct is 50 kDa, and a clear band is seen in peaks 2 and 3 of this size (figure 2.16b). Fractions associated with each of these peaks were

pooled and dialysed into low salt buffer for the anion exchange step.

The anion exchange column was loaded using a low salt buffer, 150 mM NaCl with Tris-HCl pH 8.0. The MBP-*PB6*-CSP construct was found to be insoluble at low ionic strength solutions so concentrating in this buffer led to a significant loss of protein. With a theoretical pI of 5.36, the construct was expected to bind to the anion exchange column, however after a few trials this was found to not be the case. This meant that it was possible for the protein construct to be passed through the anion exchange column at low concentration in the low salt buffer and be collected in the flow-through immediately. When a gradient to a high salt (1M NaCl) buffer was applied a high absorbance was observed in the elution, due to DNA. The output of this is shown in figure 2.17. The purity of the protein was tested by measuring the 260/280 absorbance ratio. Using this method the ratio was found to be less than 0.8 indicating less than 5% contamination from nucleic acids^[135].

At this stage the *PB6*-CSP was cleaved from the MBP by the TEVp (section 2.3.7) in an overnight incubation. The reactants and products of this digestion are shown in the first and second lanes of the SDS-PAGE shown in figure 2.18b. These products were then separated by size exclusion chromatography (section 2.3.6.3) as shown in the chromatogram and SDS-PAGE of figure 2.18. The large third peak seen in the chromatogram produces a band in the SDS-PAGE at 45 kDa, which is the expected size of the, now separated, MBP. This peak is distinct from the smaller fourth peak, which produces a band in the SDS-PAGE at 7.5 kDa, the expected weight of the *PB6*-CSP monomer.

2.3.9 Dialysis and Lyophilisation

The concentration of the purified protein solution was calculated by measuring the UV absorbance at 280 nm and applying the Beer-Lambert law. The Beer-Lambert law is given in equation 2.1 where A is the absorbance, l is the cuvette length, c is the concentration and ϵ is the extinction coefficient calculated from the protein sequence.

$$A = \epsilon lc \tag{2.1}$$

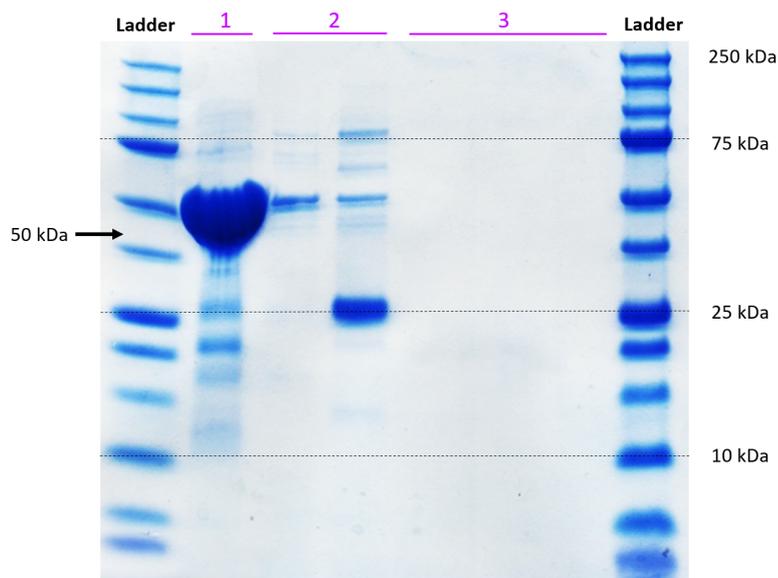
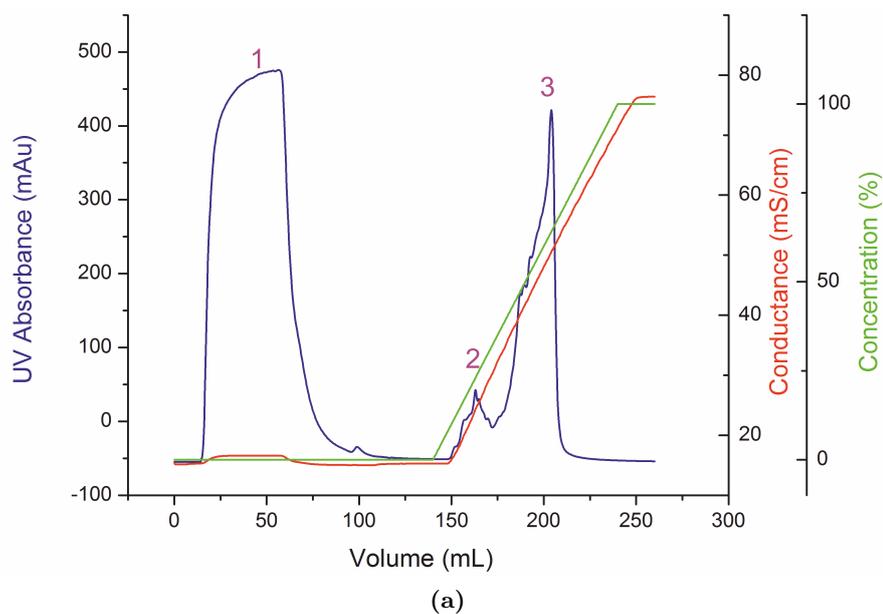
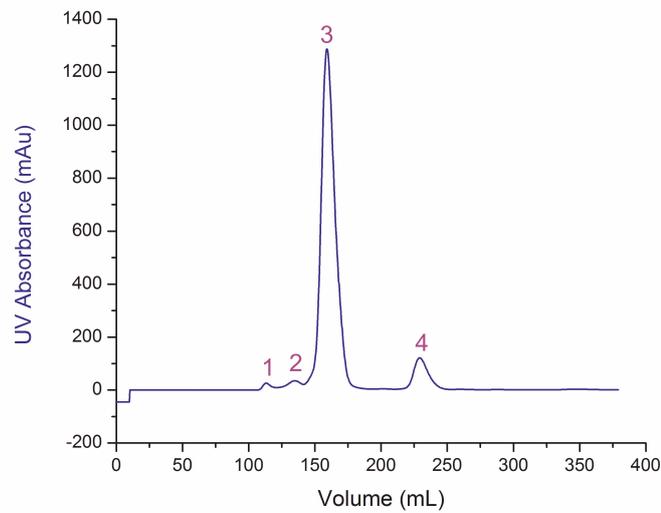
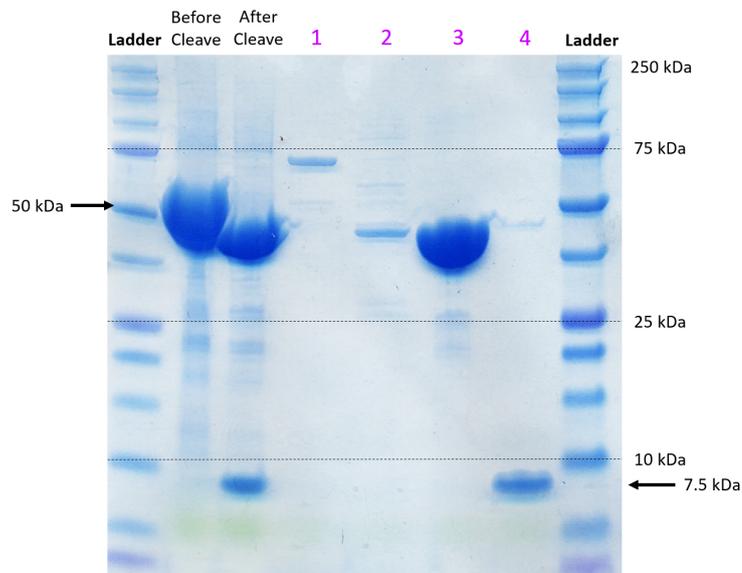


Figure 2.17: Chromatogram of the anion exchange purification step of the MBP-*PB6*-CSP construct. The concentration of high salt buffer (green) was increased to elute protein and nucleic acids (b) SDS-PAGE of numbered peaks from the anion exchange step. Most of the MBP-*PB6*-CSP construct is found in the flow-through of the column, peak 1. The high UV absorbance in peak 3, does not produce a band on the SDS-PAGE, indicating this peak is produced by the nucleic acids.

The *Bs*-CSP and all polyproteins were dialysed into milliQ water, with 50 mM ammonium bicarbonate during the first overnight step. Each sample was then snap frozen in a mixture of dry ice and ethanol, then put on to the freeze dryer for three days. This



(a)



(b)

Figure 2.18: (a) Chromatogram of the products of the digestion of the MBP-*PB6*-CSP by TEVp. (b) SDS-PAGE of the reactants and products of the digestion with TEVp and peaks of the separated products in the chromatogram. A band of the expected size for MBP, 45 kDa, is seen in the 3rd peak and a band of the expected size for the *PB6*-CSP monomer is seen in the 4th peak so fractions associated with the peak 4 were pooled.

lyophilised protein could then be stored long term in the -20 °C freezer.

After his-tag removal, *PB6*-CSP became insoluble after lyophilisation so was stored at 4 °C in buffer 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 4 mM DTT. Stored this way it could remain in solution for 2-4 weeks.

2.3.10 Circular Dichroism Spectroscopy

Protein folding and unfolding can be monitored using Circular Dichroism (CD) spectroscopy. This technique provides a measure of the degree of secondary structure features in the protein^[136]. CD uses circularly polarised light of left and right handed-ness, which interact differently with asymmetric protein components, specifically, alpha helices and β -sheets. Protein samples were prepared at a concentration of 0.3 mg/ml in 300 μ L. This was pipetted into a 1 mm path-length cuvette and placed into the Chirascan CD instrument. Ellipticity was measured using a Jasco J715 spectropolarimeter between 190 and 280 nm to assess the degree of protein secondary structure, with a bandwidth of 2 nm, a rate of 1 nm/s and 2 repeats.

When using CD to monitor protein unfolding during a temperature ramp, the following equation was used to fit the data:

$$f(T) = \frac{(a_F + b_FT)e^{-\frac{\Delta H}{R}(\frac{1}{T_m} - \frac{1}{T})} + (a_U + b_UT)}{1 - e^{-\frac{\Delta H}{R}(\frac{1}{T_m} - \frac{1}{T})}} \quad (2.2)$$

The signal $f(T)$ is dependent on the fraction of folded, F, and unfolded, U, protein. a_F and b_F define the baseline applied to correct for the folded protein signal, and a_U and b_U define the baseline applied to correct for the unfolded protein signal. T is the absolute temperature, T_m is the melting temperature of the protein, ΔH is the change in enthalpy on unfolding and R is the gas constant.

2.3.11 Monitoring Protein Aggregation with SEC-MALS

The histidine tag free protein had a low solubility, where it would precipitate out of solution after around a week at concentrations needed for NMR measurements. This meant it was necessary to monitor whether the protein remained monomeric at these concentrations. For this purpose, Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS) was used. This incorporates the size exclusion step as described in section 2.3.6.3 on a superdexTM 75 10/300 GL column, followed by the static laser light scattering (LS, miniDAWN TREOS from Wyatt technology, Refraction

Index (RI), Optilab T-rEX (refractometer with EXtended range) and ultraviolet (UV) detection^[137], set up with an ÄKTAmicroTM system. The size exclusion step separates the injectant into peaks of differing molecular size, then the molecular weight (M_w) is calculated from the LS, UV and RI measurements^[138]. The LS intensity $I(w, \theta)$ for a sample at concentration w and scattering angle θ is used to determine the Rayleigh ratio R by equation 2.3.

$$R(w, \theta) = k_{inst}[I(w, \theta) - I(0, \theta)] \quad (2.3)$$

k_{inst} is an instrumental constant determined during calibration of the LS element with a scattering standard and $I(0, \theta)$ is the LS intensity recorded in pure solvent. The molecular weight M_w of the protein can then be calculated using the Zimm-Debye formulation of static LS, equation 2.4

$$R(w, \theta) = K_{opt} \left(\frac{dn}{dw} \right)^2 M_w w P(\theta) [1 - 2A_2 P(\theta) w + O(w^2)] \quad (2.4)$$

K_{opt} is a constant, $K_{opt} = \frac{4\pi^2 n_0^2}{\gamma_0^4 N_A}$ where n_0 is the refractive index of the solvent, γ_0 is the wavelength of incident light, N_A is Avogadro's number. $\left(\frac{dn}{dw} \right)$ is the specific refractive increment of solute, A_2 is a measure of the effective interaction between the scattering molecules and $P(\theta)$ is a structure factor that is roughly equal to 1 for small (less than 10 nm) molecules. The concentration, w , is then calculated from the UV measurements and calculated according to the Beer-Lambert law, equation 2.1.

2.4 Measurements of Protein-DNA Interactions

2.4.1 Tryptophan Fluorescence Quenching

Both *Bs*-CSP and *PB6*-CSP contain a tryptophan amino acid as part of the nucleic acid binding surface. When excited by light with a wavelength of 285 nm this residue fluoresces at around 345 nm. A change in this fluorescence signal can indicate either unfolding of the protein or binding of a ligand to that site. The fluorescence emission signal from all three cold shock proteins is quenched on the binding of ssDNA.

3 ml of protein at a concentration of 300 nM, 90 nM or 30 nM was prepared in a cuvette with a stirrer bar. The Fluorescence Spectrometers used were a PTI Fluorimeter and an Edinburgh Instruments Fluorimeter. The excitation and emission slits were widened to maximise the fluorescence signal of the tryptophan residue without exceeding limits of the instrument. Each point was measured after the addition of 10 μ L of ssDNA solution at high concentration, followed by equilibration for 2 minutes. The fluorescence signal was corrected for dilution, background signal and for the inner filter effect, using equation 2.5^[139], caused by the molecules involved in the titration absorbing signal at the excitation wavelength and reabsorbing at the emission wavelength.

$$C_i = \frac{1 - 10^{A_i}}{2.303A_i} \quad (2.5)$$

C_i is the correction factor at titration point i , and A is the absorbance given by equation 2.6

$$A_i = (\varepsilon_{P,ex} + \varepsilon_{P,em})[P] + (\varepsilon_{L,ex} + \varepsilon_{L,em})[L] \quad (2.6)$$

$\varepsilon_{P,ex}$, $\varepsilon_{P,em}$, $\varepsilon_{L,ex}$ and $\varepsilon_{L,em}$ are the extinction coefficients of the protein P and ligand L at the fluorescence excitation and emission wavelengths. This correction is applicable as long as the absorbance, A_i , remains less than 0.3 which is the case for all experiments performed here. Any larger absorbance value requires more complex corrections^[140].

The fluorescence intensity of the quenching curve was fit using the tight binding equation, equation 2.7 to obtain a measure of the dissociation constant K_D .

$$I = I_M \times \left([P_T] - \frac{([P_T] + [L_T] + K_D) - \sqrt{([P_T] + [L_T] + K_D)^2 - 4[P_T][L_T]}}{2} \right) + I_{Base} \quad (2.7)$$

In equation 2.7, I_M is the fluorescence intensity per mole of unbound protein P, $[PL]$ is the concentration of the complex, $[P_T]$ and $[L_T]$ are the total concentrations of the protein and ligand respectively, K_D is the dissociation constant and I_{Base} is the baseline fluorescence intensity of the complex.

2.4.2 Isothermal Titration Calorimetry

The experiments were performed on a MicroCal 200 ITC instrument at 25 °C. Two cells are used in a thermally isolated system, one for the protein and ssDNA titration and one as a reference cell filled with MilliQ water. The energy put into, or taken out of, the titration cell in order to keep the two at equal temperature is the measure of heat released or absorbed in the reaction^[141]. An initial titration was performed injecting the ssDNA solution into buffer, this measured the energy change caused by diffusion of the ssDNA and these readings were removed from the final titration data. After an initial injection of a very small volume to clear the needle, 20 injections of 80 µM solution of ssDNA were added to a 10 µM solution of CSP and stirred gently. The power needed to equilibrate this cell with the reference cell was then integrated with time to get the overall energy change per injection, and a binding curve produced. This was then fit by the 1 to 1 binding equation provided in the MicroCal Origin software provided with the instrument^[142].

2.4.3 MicroScale Thermophoresis (MST)

Thermophoresis occurs on length scales smaller than 1 millimetre, and the thermodiffusion of particles is labelled positive when they move from a hot to a cold region (thermophobic) and negative when they move from cold to hot (thermophilic). Typically, larger molecules are expected to display positive thermophoresis behaviour while smaller molecules are expected to show negative behaviour. Other factors that affect the behaviour are the heat conductivity and heat absorption of the particles, plus the charge and entropy of the hydration shell of molecules. By introducing a temperature gradient of a few degrees using an infra-red (IR) laser, MST can detect changes in these properties caused by ligand binding. Lyophilised ssDNA samples with an Alexa Fluor 488 fluorescent label were re-suspended in the same buffer as the protein (63 mM sodium phosphate, pH 7.4).

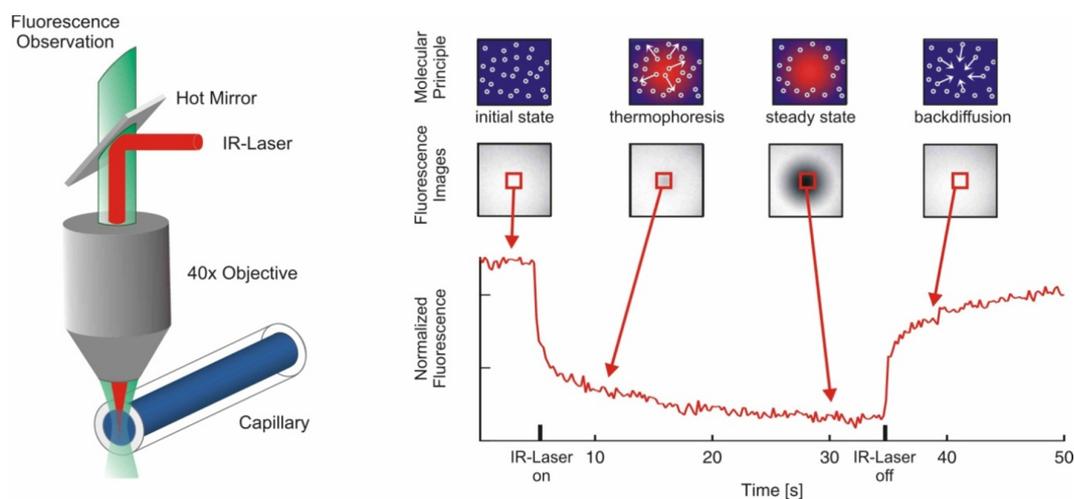


Figure 2.19: Diagram of the MST equipment with the excitation light and IR laser shone at the same point in the capillary. On the right is an example of the fluorescence signal against time obtained when molecules thermodiffuse away from the heated spot (thermophobic). Each capillary has a different concentration of unlabelled protein and produces a different curve depending on the binding of the protein to the labelled ssDNA. Image from Nanotemper-technologies website^[143]

16 capillaries were prepared containing a 25 nM solution of labelled CT2 ssDNA and *Bs*-CSP concentrations starting at 10 μ M diluted down to around 0.3 nM by progressively halving the concentration. The experiment was carried out at 22 °C. Each capillary was then excited by a blue LED so that the label on the ssDNA fluoresced and an initial scan of the fluorescence signal in each capillary taken to ensure consistency. If a clear fluorescent signal was measured across all the capillaries the MST measurements were started. The samples were exposed to the excitation light again and then after a short time an infra-red laser was turned on aimed at the same spot as the excitation light introducing a temperature gradient as shown in figure 2.19.

After a short time that allowed the solution to reach equilibrium, this IR laser was switched off and the temperature gradient stopped. The initial fluorescence signal before switching on the IR laser was used to normalise the signal so that the movement of the labelled ssDNA in response to the temperature gradient was measured by the change in fluorescent signal.

2.5 Single Molecule Force Spectroscopy (SMFS)

SMFS applies the Atomic Force Microscope (AFM) to the measurement of unfolding or unbinding forces. AFMs consist of a piezoelectric positioning stage and a small cantilever that is deflected by small forces. This deflection is monitored by a laser directed by mirrors to the end of the cantilever, which is then reflected onto a photodetector. By measuring the position of the laser on this detector, the deflection of the cantilever and hence force on the tip can be measured. When this is used in imaging mode the cantilever is scanned over the surface to map out the forces. For SMFS the AFM cantilever pulls on molecules in the vertical direction to probe molecular forces between them. The set up for a polyprotein unfolding experiment is shown in figure 2.22.

2.5.1 Polyprotein Unfolding Experimental Set-Up

To prepare a surface that will bind to the cysteine of the polyprotein through a thiol bond, a glass slide was glued to a thin (100 nm) layer of gold on a silicon wafer. The glass slide with gold attached was then cleaved from the silicon wafer, ensuring a clean gold surface. This was attached to a magnetic disk and placed on the piezo-electric positioning stage. The (I27-*Bs*-CSP)₃-I27 polyprotein was re-suspended to around 0.1 mg/ml in 63mM sodium phosphate buffer at pH 7.4. This was filtered through a 0.44 μm filter to remove any aggregated protein. 40 μL of this solution was pipetted onto the gold surface and left to equilibrate for 30 minutes.

The fluid cell was prepared with a rubber o-ring, syringe of buffer and a new MLCT silicon nitride cantilever. Once the cantilever position was checked under a microscope, the fluid cell was placed in the multimode head and tightened into place. A webcam was then used to position the laser onto the appropriate cantilever (the V-shaped cantilever labelled D, of length $225 \pm 5 \mu\text{m}$ and spring constant 30 pN/nm with a 50 % error, according to manufacturer specifications was used for all experiments).

After equilibration the protein solution was removed from the gold surface and replaced

with buffer to remove any polyprotein not attached to the surface. The AFM head, with fluid cell secured, was positioned over the piezo-electric positioning stage, so that the fluid cell was filled with the droplet of buffer. This was left to equilibrate for 10 minutes before calibration.

2.5.2 Calibration of the AFM Cantilever

The Cantilever was calibrated using the thermal equilibration method^[144]. After equilibration, the frequencies of thermal oscillation were measured by performing a Fast Fourier Transform (FFT) on the voltage signal measured by the photo-detector over a period of time sitting at equilibrium, as shown in figure 2.20. The output of the FFT

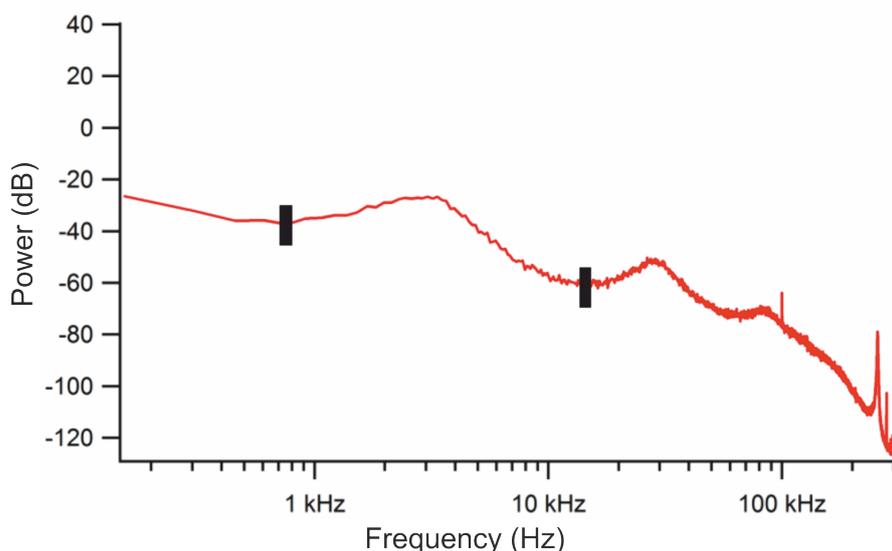


Figure 2.20: Output of the FFT as used for selection of the fundamental resonant mode in the thermal equilibration method of calibrating the AFM cantilever. Power is displayed in decibels (dB) calculated from the voltage signal by $20\log_{10}(V/V_t)$ where V_t is the sum total of all voltage signals in the spectrum as a reference. Power is plotted against the frequency of oscillation

calculates the power (in units of V^2) associated with each frequency. This allows the main vibrational mode of the cantilever to be isolated from any background noise. The first normal mode is isolated as shown by the black markers in figure 2.20. Integrating to find the area under this peak gives the mean power of the resonance, which is initially in units of V^2 . To convert the voltage signal from the photo-detector to the deflection of

the cantilever the relation of voltage signal to tip deflection must next be calibrated. For this the surface is raised into contact with the tip and then moved away at a constant velocity recording a force-extension trace on a clear surface area, as shown in figure 2.21.

As the surface is rigid, the vertical motion of the positioning stage is directly correlated

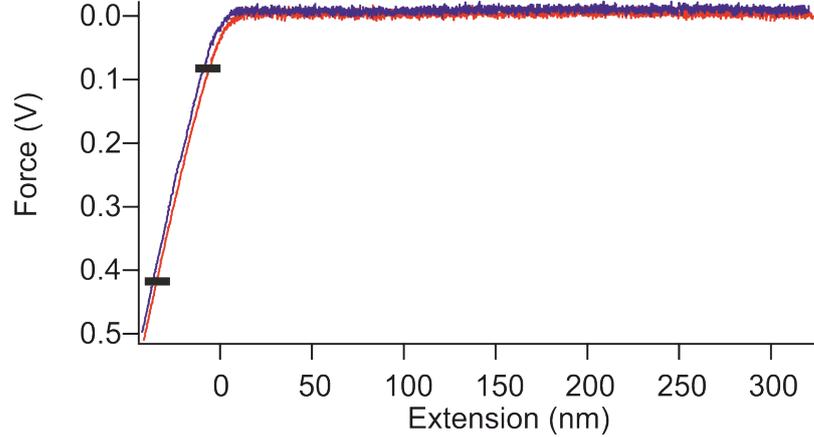


Figure 2.21: Slope calibration showing the Photo-detector voltage plotted against vertical position of the surface for a single approach (Red) and retraction (Blue). Black markers indicate edges of the region used to calculate the slope from pushing the tip into the rigid surface.

with the motion of the tip of the cantilever when they are in contact and the tip is being pushed into the surface. Each nanometre of vertical movement of the positioning stage corresponds to a nanometre movement of the end of the cantilever. Using this relation for this range of motion, as identified by the two black markers in figure 2.21, the signal can be converted from a photo-detector voltage to a distance of cantilever tip deflection (x).

Treating the cantilever as a simple harmonic oscillator the energy of the system is described in equation 2.5.2.

$$\left\langle \frac{m\omega_0^2 x^2}{2} \right\rangle = \frac{k_B T}{2} \quad (2.8)$$

Here m is the mass of the oscillator, ω_0 is the resonant angular frequency of oscillation, x is the tip displacement, k_B is the Boltzmann constant and T is the absolute temperature. This applies the equipartition theorem to assume that the average of each quadratic term in the Hamiltonian of the harmonic oscillator, here the kinetic energy, is equal to $k_B T/2$. As ω_0 is equal to $\frac{k}{m}$ where k is the spring constant and m is the mass of the oscillator, equation 2.5.2 can be simplified to $k = \frac{k_B T}{\langle x^2 \rangle}$. $\langle V^2 \rangle$ is the value calculated

by integration of the power spectrum in figure 2.20, and by applying the conversion of V to x calculated from the slope in figure 2.21 the value of $\langle x^2 \rangle$ is obtained, and hence the value of the spring constant k .

2.5.3 Polyprotein Pulling Experiments

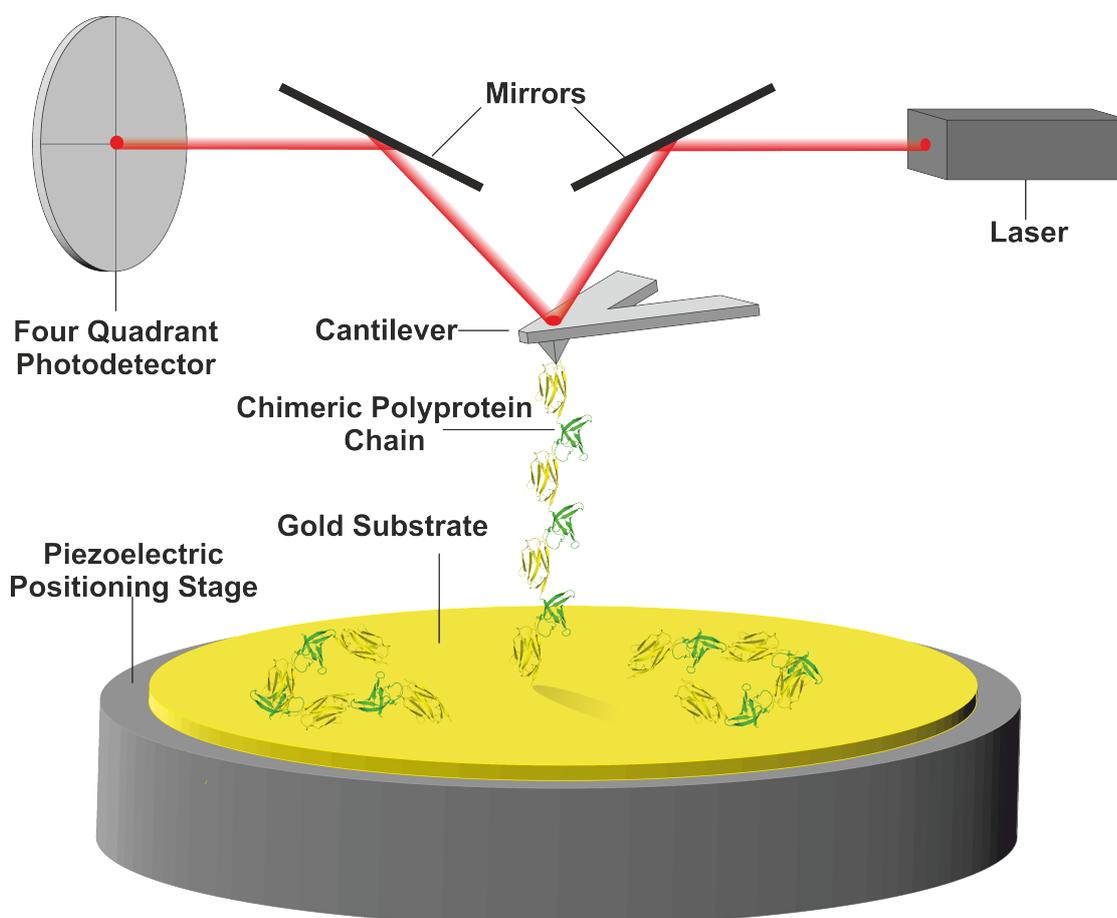


Figure 2.22: Diagram of the polyprotein unfolding experiment set up on the AFM. The mirrors are positioned so that the laser beam shines onto the end of the cantilever and is reflected to the centre of the four quadrant photodetector. Polyproteins are covalently bound to the gold surface at one end via a cysteine residue. The cantilever is shown bound to the other end of a chimeric polyprotein chain. The SMFS experiments are performed as the piezoelectric positioning stage moves the gold surface away from the cantilever at a constant velocity, increasing the force across the polyprotein bound to the tip until a protein domain unfolds or the polyprotein chain unbinds from the tip.

Once the cantilever spring constant had been found, the experiment was started. The

positioning stage was raised into contact with the cantilever, held for a small amount of time to allow the polyproteins to attach to the tip, and then lowered at the speed chosen for that particular experiment. The apparatus is shown in figure 2.22, with one full (I27-*Bs*CSP)₃-I27 construct picked up by the AFM tip. This was repeated in the same location on the surface 10 times, then the surface was moved 10 nm horizontally and the approach retract cycle repeated. In this way the gold surface was scanned to maximise the number of polyprotein unfolding events recorded. This was continued until the end of the experiment, with adjustments made throughout if there was too little protein being picked up or too much protein obscuring the measurements.

2.5.4 The Worm Like Chain (WLC) Model

The peaks seen in the force-extension (FX) traces obtained were fitted using the worm-like chain (WLC) model, (equation 2.9):

$$F(x) = \frac{k_B T}{p} \left(\frac{1}{4 \left(1 - \frac{x}{L_c}\right)^2} - \frac{1}{4} + \frac{x}{L_c} \right) \quad (2.9)$$

where p is the persistence length, L_c as the contour length, k_B is Boltzmann's constant and T is the absolute temperature. This allows the proteins to be identified by their contour length which can be predicted from their amino acid sequence. For I27 this is around 28 nm and for the *Bs*-CSP and *PB6*-CSP it is around 23 nm, with a persistence length for all between 0.35 nm and 0.40 nm.

2.5.5 Protein Unfolding Analysis

Multiple FX traces are recorded during a SMFS experiment and most of these will not report a polyprotein unfolding event. Each polyprotein is covalently bound to the gold surface through a cysteine residue engineered into one end of the chain. The interaction between the AFM tip and the polyprotein chain is a non-specific interaction so it can

bind at any part of the polyprotein construct, or not bind at all. The source of this interaction force is not fully understood, usually it is thought to be due to physioadsorption and that by holding the tip on the surface for a more extended period of time the chance of binding a polyprotein increases^[145]. Some evidence suggests that the interaction is electrostatic, with charged amino acid side chains measured showing a stronger interaction with inorganic materials than neutral side chain residues^[146].

FX measurements recorded throughout the experiment were checked manually for unfolding peaks corresponding to at least 2 I27 unfolding events and a clear detachment peak. An example of the production of an unfolding trace is shown in figure 2.23. Fig-

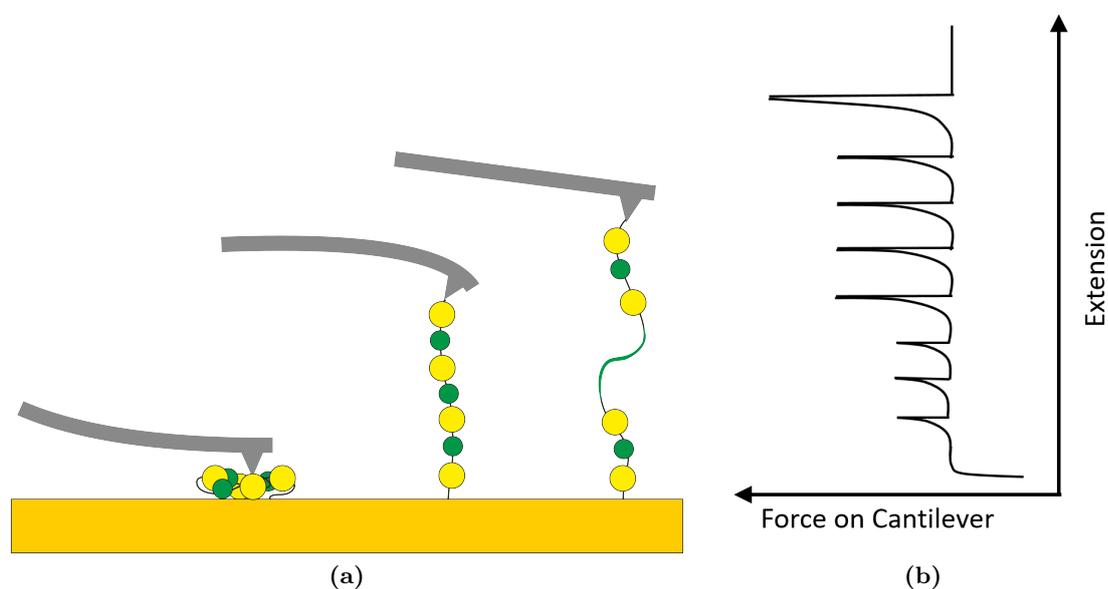


Figure 2.23: Production of an unfolding trace (a) Three cantilever positions in the initial retraction from the surface, first pushed into the surface, then at high force through the attached polyprotein and finally at no force when the domain unfolding has slackened the tension between protein and tip. (b) The trace produced by this polyprotein fully unfolding, with three smaller unfolding peaks for the CSP (green) domain, four larger unfolding peaks for the I27 (yellow) domain and a final detachment peak.

ure 2.23a shows how a chimeric polyprotein containing three CSP domains and four I27 domains is picked up by the tip and each domain is progressively unfolded, from lowest unfolding force to highest. Figure 2.23b shows the corresponding trace produced from this set-up. There are 3 lower force unfolding peaks, one for each CSP domain, four higher force unfolding peaks for each I27 and then a final detachment peak. Traces were

rejected if there were more than 7 unfolding events, peaks appeared split, obscured by noise or if they did not fit with the WLC model (equation 2.9). Once each peak is identified as a protein unfolding peak by a WLC model, the unfolding force is measured by taking the height of the peak. The peak-to-peak distance is also recorded, giving an approximation of the increase in length caused by the domain unfolding. Gathering these measurements for a whole experiment allows statistical analysis of the unfolding events of many single molecule unfolding events.

2.5.6 Protein Unfolding Energy Landscape

During SMFS unfolding experiments both thermal and mechanical forces influence the protein unfolding process. As the pulling velocity increases the time window for thermal fluctuations while the protein is under mechanical stress decreases, leading to a higher unfolding force. This effect can be used to explore the underlying energy landscape of protein unfolding as zero force, by determining the pulling velocity dependence of the force at which the protein unfolds^[145].

Assuming a two-state unfolding process the unfolding rate k_U can be described as a diffusion by Brownian motion over a 1D energy profile described by the Bell model, (figure 1.18). The effect of force on this energy landscape is to reduce the energy of the transition state barrier (ΔG_{TS}). The force dependent unfolding rate ($k_U(F)$) can be described using the Bell-Evans Ritchie model^[147]:

$$k_U(F) = A \exp\left(\frac{-(\Delta G_{TS}^0 - F\Delta x_U)}{k_B T}\right) = k_U \exp\left(\frac{F\Delta x_U}{k_B T}\right) \quad (2.10)$$

where F is the applied force, Δx_U is the distance to the transition state along the reaction coordinate, A is the attempt frequency, k_B is Boltzmann's constant, T is the temperature, ΔG_{TS}^0 is the unperturbed transition state free energy and k_U is the unfolding rate at zero force. This predicts an exponential increase in $k_U(F)$ as F is increased that is dependent on the constants k_U and Δx_U .

2.5.7 Monte Carlo Simulations of Protein Unfolding

Software developed previously^[148] for the Monte Carlo simulation of unfolding heteropolyproteins was used to match experimental data with underlying energy landscape parameters. This software mimics the stochastic nature of the unfolding process and generates histograms of unfolding forces for set values of k_U and x_U at different pulling velocities. The values of k_U and x_U are adjusted until the simulated F_U values match those of experimental values at multiple pulling velocities.

Each simulation begins with a pick-up event in which the AFM tip attaches to the (I27-*BsCSP*)₃-I27 at one of the protein domains. This pick-up point is randomly selected between domains 3-7 to ensure the inclusion of 2 I27 unfolding events as this was part of the selection criteria for experimental data. This point defines the number of folded domains, N , of each type of protein domain, I27 and CSP. In constant velocity pulling experiments the force across the polyprotein changes as a function of time and this was simulated using the WLC model described in equation 2.9, where the distance x increases with each timestep Δt according to the pulling velocity, v_C . The contour length, L_C is determined by the sum of the folded lengths of each domain. Based on previously determined structures the folded length of the I27 was set as 4.1 nm, and the folded length of the CSP was set to 1.4 nm. All parameters used in simulations are summarised in table 2.19. Based on the sequence of the (I27-*BsCSP*)₃-I27 the linker length between each domain was set to 2.3 nm.

The probability of a domain unfolding, P_U during each timestep, Δt , is given in equation 2.11.

$$P_U = Nk_U(F)\Delta t \quad (2.11)$$

Here N is the number of folded domains remaining in the polyprotein and $k_U(F)$ is the unfolding rate determined by the Bell-Evans-Ritchie model of stochastic unfolding given in equation 2.10. The probability P_U is then compared to a random number between 0-1, to determine whether the domain unfolds or not. Δt was set to 0.0001 s to ensure that $P_U \ll 1$ during each timestep.

If the simulation indicates an unfolding event the force, F_U , is recorded and the number

of protein domains N reduced by 1. Then the overall L_C length is increased by the difference between the folded and unfolded lengths of the domain. The simulation then continues until all of the domains are unfolded. This is repeated to produce a distribution of unfolding forces for the range of pulling speeds used in experiments.

To obtain the best representation of all of the experimental data collected for each

Parameter	Value
Timestep, Δt	0.0001 s
Number of Iterations	1000
Persistence length, p	0.38 nm
Average Spring constant	34 pNnm ⁻¹
Average Linker Length	2.3 nm
Length of folded <i>Bs</i> -CSP	1.4 nm
Length of unfolded <i>Bs</i> -CSP	23 nm
Length of folded I27	4.1 nm
Length of unfolded I27	28 nm

Table 2.19: Parameters used in Monte Carlo simulations of (I27-*Bs*CSP)₃-I27 unfolding events, based on experimental values and previous studies^[83]

set of conditions, a linear fit of the median unfolding forces against the natural log of the pulling speed was determined. MC simulations were then performed until values matching this fit were obtained, within experimental error. The standard error of the linear fit was used to determine uncertainties, with simulations performed to match the steepest and shallowest linear relationships included in the standard error.

2.5.8 SMFS Experiments using Functionalised ssDNA Tips and *Bs*-CSP Functionalised Surfaces

Silicon nitride AFM cantilevers and surfaces were functionalised with the cysteine mutated *Bs*-CSP (section 2.2.3) and thiolated ssDNA (table 2.20) using NHS-PEG₂₄-maleimide linkers through four preparation steps.

1 Oxidisation of cantilevers and surfaces

Silicon surfaces and cantilevers were cleaned using piranha solution (3:1 0.5 M H₂SO₄ with 30% H₂O₂) before being placed under a UV lamp set to 254 nm for 30 minutes.

2 Aminosalinisation of silicon nitride

Primer	Sequence
Short	Thiol-GCTTTTTTTT
Long	Thiol-CACCATCACCATCACCATCACCATCACCAT- -CACCATCACCATCACCATCACTTTTTTTT

Table 2.20: Sequences of ssDNA strands used for SMFS unbinding experiments

Both cantilevers and surfaces were then placed in a desiccator next to 80 μL of (3-aminopropyl)triethoxysilane (APTES) and 20 μL of N,N-diisopropylethylamine (DIPEA) in small plastic dishes. A vacuum pump was used to evacuate the desiccator which was then left to incubate for 2 hours before the removal of the APTES and DIPEA solutions. The desiccator was then filled with nitrogen and left for at least 48 hours to cure.

3 Attachment of NHS-PEG₂₄-Maleimide linkers

The cantilevers and surfaces were submerged for one hour in a solution of 1 ml chloroform with 15 μL of the NHS-PEG₂₄-maleimide linkers. Then they were washed with chloroform and dried with nitrogen gas.

4 Attachment of the cysteine mutated protein and thiolated ssDNA

The dried cantilevers were then incubated in a solution of thiolated ssDNA containing 2-4 μM of tris(2-carboxyethyl)phosphine (TCEP) for 30 minutes. 60 μL of protein solution (0.5mgml⁻¹) with 0.001% TWEEN was added to the surface for 30 minutes. The TCEP is used to break any disulphide bonds formed between the thiolated ssDNA strands and the TWEEN is a detergent used to reduce protein surface interactions which cause the protein to stick to the surface. After this incubation period both the cantilever and the surface were rinsed using 63mM sodium phosphate buffer at a pH of 7.4.

The cantilever was then loaded into the fluid cell of the force clamp AFM and the surface was raised by the piezoelectric device so that the surface and the tip were in contact. After a short delay, usually 0.2 s, the surface was retracted and any unbinding event measured as a peak in the force-extension trace, as shown in figure 2.24b. This was repeated at least 2000 times for each experiment and the data was then analysed by fitting all clear peaks (traces where only one unbinding peak was recorded) with a WLC model and the rupture forces and fitted contour lengths at unbinding were recorded.

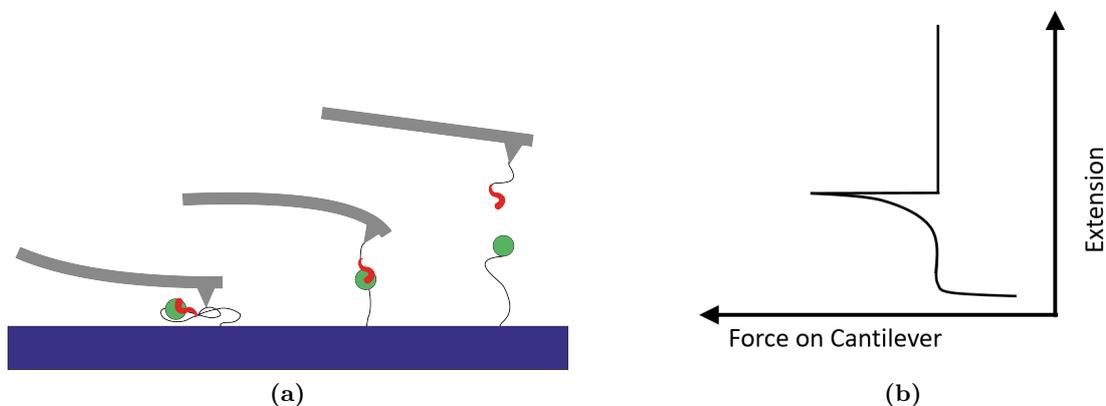


Figure 2.24: Production of an unbinding trace between a surface attached protein (green) and a binding ligand (red) (a) Three cantilever positions in the initial retraction from the surface, first pushed into the surface, then pulled back by the attached complex and finally at no force when the ligand has detached. (b) The trace produced by this unbinding, with one detachment peak.

Software to carry out these analysis steps was written by Yun Chen. The maximum contour length at the point of rupture is the sum of the length of the two NHS-PEG₂₄-maleimide linkers, each 9.52 nm plus the total stretched length across the *Bs*-CSP-dT7 complex. Folded, the *Bs*-CSP is around 5 nm across and the thiolated ssDNA strand is around 3 nm, so the contour length is expected to be around 25 nm depending on the geometry and stretching of the complex. With the longer DNA strand described in table 2.20, this was extended to 45 nm. For these longer ssDNA strands, the complementary strand for the whole sequence excluding the final 7 thymine nucleotides was added to reduce flexibility and interactions between the protein and other regions of the DNA. If the density of ssDNA on the tip or protein on the surface is too high, then multiple bonds between the ssDNA and protein will form while the tip is on the surface and peaks of a higher force and shorter measured contour length will appear as noise in the data^[149].

The unbinding interaction and resulting trace are shown in figure 2.24. Unlike the polyprotein unfolding experiments there is not a clear fingerprint to distinguish an event from normal surface interactions other than the position of the peak. This is why con-

trolling the contour length and having a control experiment are vital requirements to obtaining meaningful results. As a control, after this experiment was complete, ssDNA with a sequence AAAAAAA (dA7) was added to the solution in the fluid cell. As this is the complementary sequence to the dT7 on the functionalised tip this should form double stranded DNA which the *Bs*-CSP does not bind to. This will block the interaction being measured and so should lead to a reduction in measured unbinding traces.

2.6 Protein NMR Methods

All NMR experiments were performed using Varian Inova spectrometers with cryogenic probes and proton resonance frequency of 600 MHz or 750 MHz, under the guidance of Dr Lars Kuhn and Dr Arnout Kaldera.

2.6.1 Acquiring HSQC Spectra

^{15}N labelled proteins were used to obtain HSQC Spectra. 600 μl samples were prepared at a protein concentration of 200 μM in 20 mM HEPES, 150 mM NaCl, pH 7.4 in 10% D_2O and with 0.1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) for the *PB6*-CSP and in 20 mM HEPES, 50 mM NaCl, 0.1 mM DSS, pH 7.5 in 10% D_2O for the *Bs*-CSP. 2D ^1H - ^{15}N HSQC spectra were acquired at 5, 10, 20, 29 or 37 $^\circ\text{C}$ using 128 complex points in the indirect (^{15}N) dimension and 1024 in the direct dimension over 16 scans. Spectral widths were adjusted to contain all the chemical shifts in each experiment. All the data was processed with NMRpipe^[150].

2.6.2 Backbone assignment experiments

A uniformly ^{13}C and ^{15}N labelled *PB6*-CSP sample was prepared at a protein concentration of 200 μM in 20mM HEPES, 150mM NaCl, pH 7.4 in 10% D_2O and with 0.1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Experiments were performed at 20 $^\circ\text{C}$. HNCA, HNC(O), HN(CO)CA and HN(CA)CB spectra were used in the assignment, these

experiments expand the HSQC in a third dimension (^{13}C) through the spin coherence transfer as described in section 1.5.2. Processing of the spectra was performed using NMRpipe^[150] and assignment using ccpn Analysis^[151].

2.6.3 Relaxation measurements

The ^{15}N labelled *PB6*-CSP sample was prepared at 200 μM in 20 mM HEPES, 150 mM NaCl, pH 7.4 in 10% D_2O and with 0.1 mM DSS. $\{^1\text{H}\} - XhnNOE$ proton saturated and unsaturated spectra were recorded. For T_2 determination, 9 experiments were performed with relaxation delays increasing in steps of 16.963 ms, and 2 duplicate spectra were recorded for uncertainty determination. $^1\text{H} - ^{15}\text{N}$ HSQC spectra were recorded before and after relaxation experiments to monitor sample quality. For the relaxation dispersion CPMG experiments, 16 spectra were obtained with ν_{CPMG} (equation 1.21) increased from 0 to 1000 Hz in 14 steps, again with 2 duplicate spectra for determination of uncertainty.

Peak volumes were calculated using Peak INTeGration (PINT) software^[152,153]. This

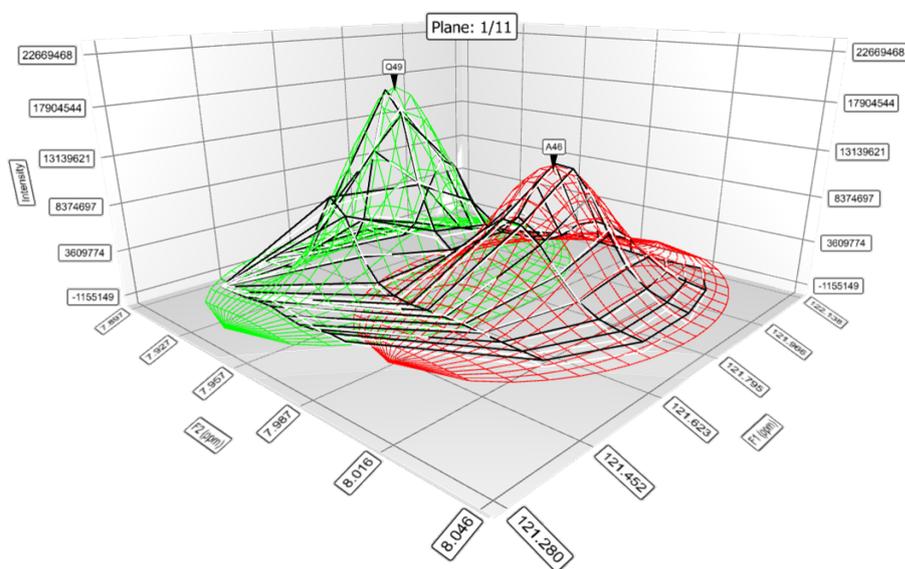


Figure 2.25: Peak integration of two overlapped peaks. Model Gaussian peaks are placed where each peak is identified, then the experimental data (Black) is fit with the simulated data (White) and volumes are calculated from this shape.

fits the experimental data with predefined lineshapes which can then be integrated to find the volume. Overlapped peaks can be separated (figure 2.25) and the peak volumes calculated used in further analysis of relaxation rates. For the relaxation dispersion measurements, the peak volumes were then loaded into the software Graphical User-friendly Analysis of Relaxation Dispersion Data (GUARDD)^[154]. This program uses MATLAB software and allows the global fitting of multiple RD curves to the Multiple Quantum Carver-Richards-Jones all-timescales dispersion equation^[155], to extract kinetic details.

3

Impact of Nucleic Acid Binding on Mechanical Stability of the Cold Shock Protein

Cold Shock Proteins (CSPs) are understood to function as RNA chaperones^[63]. When organisms experience a drop in temperature, CSPs aid in the melting of RNA secondary structures unfavourable for translation, helping acclimate organisms to the new conditions. For this role, the protein is functional when it is bound to nucleic acids. The mechanic properties of the *Bs*-CSP have not been investigated previously in this bound state. In this chapter the impact of ssDNA binding on the mechanical unfolding of the Cold Shock Protein B from *Bacillus Subtilis* (*Bs*-CSP) is investigated with Single Molecule Force Spectroscopy (SMFS). For this purpose, a chimeric polyprotein consisting of 4 I27 domains interspersed with 3 CSP domains ((I27-*Bs*CSP)₃-I27) was prepared as described in methods section 2.3.8.1^[156]. Mechanical unfolding of this structure was performed on the Atomic Force Microscope (AFM) in the absence and presence of binding ssDNA oligonucleotides. The effect of ssDNA on the mechanical stability of the *Bs*-CSP was measured using different ssDNA oligonucleotides, at a range of pulling speeds and at two temperatures.

SMFS provides the ability to measure the behaviour of individual molecules as opposed to bulk measurements which record signal from averages of the whole population. This allows a far more detailed view of protein systems, and can elucidate variation in individual molecule behaviour. In proteins, which are dynamic molecules that sample a range of conformations through stochastic fluctuations, this can elucidate outliers that would otherwise be lost in the average.

Mechanical unfolding of proteins is dependent on the kinetics of the unfolding pathway, which is influenced by the energy of the transition state. Other methods to measure protein unfolding such as with a chemical denaturant or temperature ramp are can only measure the difference in energies of the initial and final state.

3.0.1 Incorporating I27 in the Polyprotein Chain as a Marker Protein

The CSP monomer is incorporated into a chimeric polyprotein chain with I27 as a marker protein of well characterised mechanical properties^[157].

The giant muscle protein titin has a function that requires mechanical stability and elasticity^[158], and the region of this protein containing repetitions of Immunoglobulin-like domains provides both of these properties at the molecular scale^[13]. One of these domains is I27, and the mechanical unfolding of this structure has been very well characterised under a range of conditions^[159,160]. The successful incorporation of this monomer into a chimeric polyprotein chain with I27 as a marker with new Gibson assembly techniques allowing for efficient production of a chimeric polyprotein gene has been demonstrated^[156].

3.1 Determining the Polyprotein-ssDNA Binding Association

Tryptophan fluorescence quenching titrations were performed on the polyprotein (I27-*Bs*-CSP)₃-I27 to ensure that the polyprotein structure did not obstruct binding of ssDNA

with individual domains. This uses the fluorescence signal observed from the tryptophan that forms part of the *Bs*-CSP nucleic acid binding region^[59]. When nucleic acids bind to the *Bs*-CSP then this fluorescence is quenched, providing a signal proportional to the free protein in solution. Figure 3.1b shows the resulting quenching effect caused

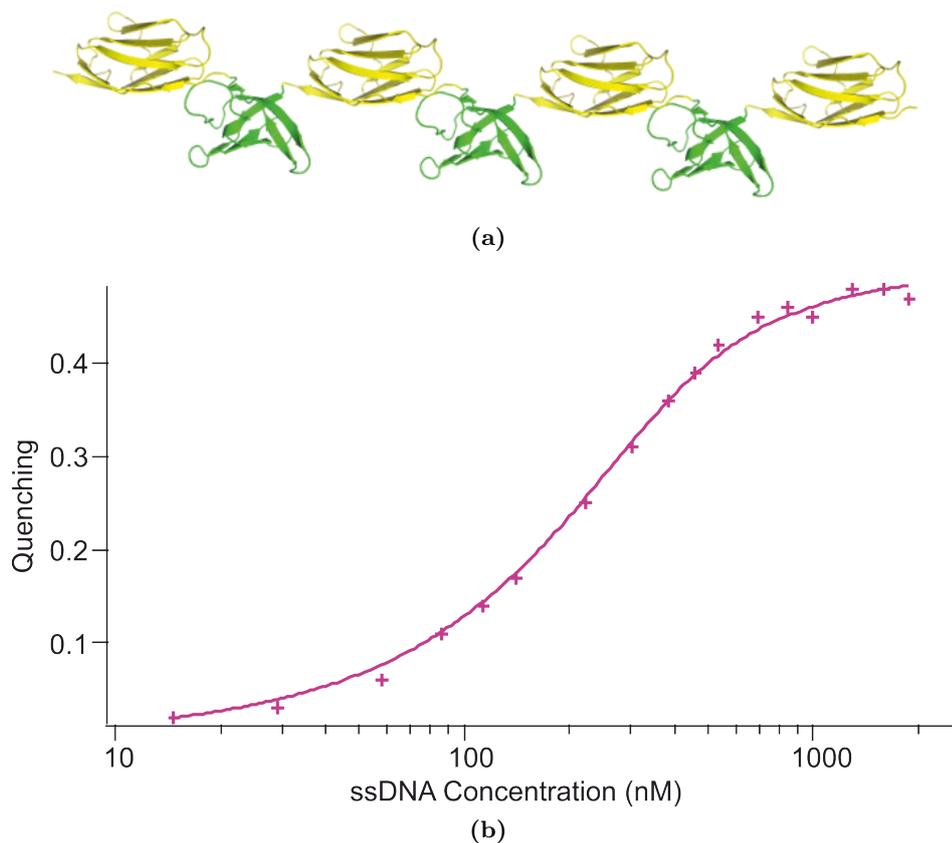


Figure 3.1: (a) $(I27-BsCSP)_3-I27$ consisting of 4 I27 domains (yellow) and 3 *Bs*-CSP domains (green) (b) Quenching of tryptophan fluorescence in the polyprotein $(I27-BsCSP)_3-I27$ as CT2 is titrated into the sample. Experimental data shown by markers for experiments recorded at 15 °C (pink). Lines show the fits to the tight binding equation, with resulting K_D value of 68 ± 14 nM

by this titration with the ssDNA oligo CTTTTTC (CT2). The CSP domains show a clear reduction of fluorescence in the presence of CT2, following the trend typical of quenching by binding at the tryptophan site. This is a clear indication that the CSP domains are still capable of binding ssDNA while in the polyprotein. The maximal quenching reached is lower than that seen in the monomer quenching titrations, only reaching a value of 0.5, but this is likely due to the signal from the I27, which also contains a tryptophan residue. The signal from this protein should not be affected by

the presence of ssDNA. The dissociation constant, K_D derived from the fit to the tight binding equation (equation 4.9) was 68 ± 14 nM. For CT2, which has a slightly lower affinity than dT7, this is in the same order of magnitude as measurements made at 15 °C, where the K_D was calculated as 33.7 ± 4.1 nM^[62]. This oligo was chosen early in the experimental process as an alternative to dT7 in order to avoid cross-linking of the CSP domains in the polyprotein, as observed in the crystal structure of *Bs*-CSP bound to dT6^[62]. This potential cross-linking might obscure the unfolding peak of the *Bs*-CSP recorded in FX experiments. CT2 contains the high affinity thymine region surrounded by the two cytosine residues which were shown to have a lower affinity^[61,130]. It was thought that this sequence would be less likely to bridge domains in the polyprotein structure, while still binding with a high affinity to the *Bs*-CSP.

3.2 Obtaining a Mechanical Fingerprint for (I27-*Bs*CSP)₃-I27 unfolding

SMFS experiments were carried out using a constant pulling speed to retract the gold surface on which the polyprotein was bound away from the AFM tip. This results in a Force Extension (FX) trace with the force measured on the cantilever plotted against the distance between surface and tip, or extension.

An example trace showing the full (I27-*Bs*CSP)₃-I27 structure unfolding is shown in figure 3.2. The *Bs*-CSP is mechanically weaker as well as being smaller than the I27 domain so as the distance between the tip and the surface increases, extending the polyprotein chain, the weakest domain unfolds first. After each unfolding event the force of the cantilever is reduced to zero, as the tension of the chain has been reduced by the additional difference in contour length (ΔL_C).

Each unfolding peak is identified by the fit to the Worm-Like Chain (WLC) model, described by equation 2.9 in the methods section 2.5.4. The ΔL_C of the polypeptide chain upon unfolding is equal to the unfolded length of the protein minus the length between the N and C termini while the protein is folded. The persistence length is typically 0.35-0.4 nm for polyprotein chains. The (I27-*Bs*CSP)₃-I27 construct is shown

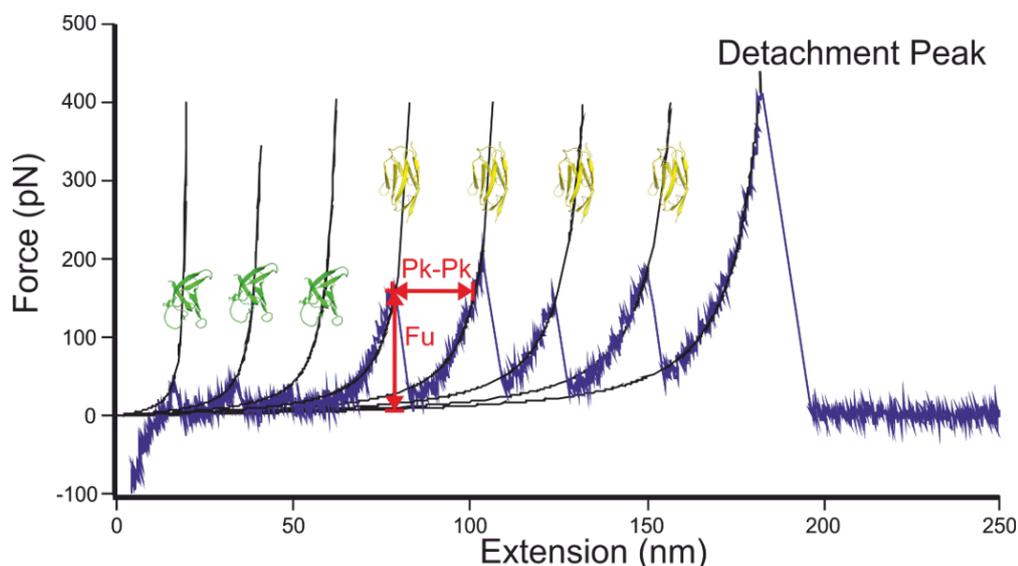


Figure 3.2: Force Extension trace of the chimeric polyprotein. An example FX trace measured at 200 nm s^{-1} contains unfolding peaks for all 7 domains and a detachment peak. Each peak has been identified with a corresponding domain by fits to the WLC (black). The peak to peak distance (Pk-Pk) and unfolding force (F_U) are shown in red for the first I27 unfolding peak.

in 3.1a, with the unfolding peaks for each domain identified in 3.2. For I27 the folded length is around 4 nm while unfolded it is around 32 nm (98 amino acid residues each adding around 0.32 nm to the length, $98 \times 0.32 \text{ nm}$), leading to a ΔL_C of 28 nm (32 nm - 4 nm). The smaller *Bs*-CSP has a folded length of around 1.4 nm and an unfolded length of around 24 nm, so the ΔL_C of *Bs*-CSP is around 23 nm. Incorporating these values into the WLC model, shown by the black lines in figure 3.2 and described in equation 2.9, each unfolding peak of the FX trace can be identified as a domain of a particular size.

From each unfolding peak in the FX trace, the unfolding force (F_U) and distance between peaks, or peak to peak distance (P_2P), can be identified. These values are identified in figure 3.2 in red, showing F_U as the height of the peak maximum and the P_2P distance as the distance between the top of the peak and a point in the following peak of equal height. The P_2P distance is correlated with the ΔL_C and can be obtained from each FX trace. These values are used in the statistical treatment of these single molecule studies.

3.3 Triplicate Experiments show that a Robust Mechanical Fingerprint can be Obtained

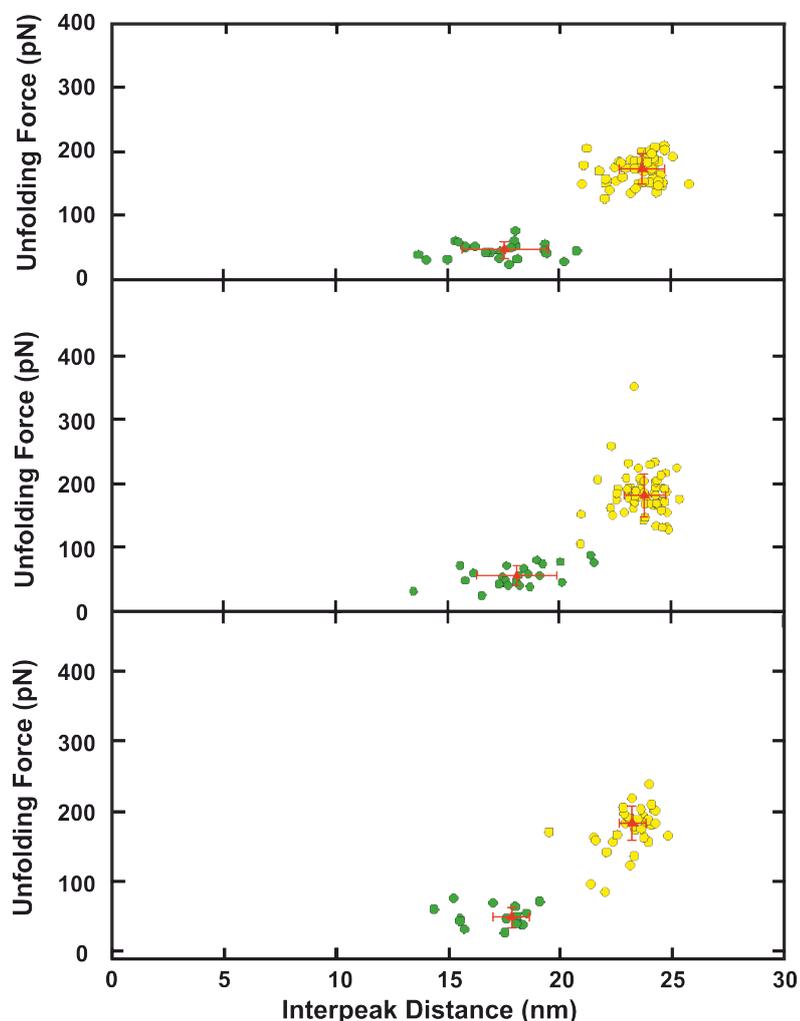


Figure 3.3: Scatter plots of the Unfolding force against the inter-peak distance for three FX experiments performed on the $(I27-BsCSP)_3-I27$ at 600 nms^{-1} and at $23 \text{ }^\circ\text{C}$. Peaks were identified as either $Bs\text{-CSP}$ (green) or $I27$ (yellow) by fits to the WLC model. The median values calculated for each domain are shown (red) with error bars representing the standard deviation.

SMFS experiments in each set of conditions are performed in triplicate and statistical analysis is performed on each experiment separately. This helps confirm that cantilever calibration has been consistent and that no outlier experiments have been recorded which might skew results.

FX traces are examined and the values for F_U and the P_2P distance for each peak are

collected. Scatter plots of F_U against the P₂P distance are shown in figure 3.3. This demonstrates results from three FX experiments performed using the (I27-*BsCSP*)₃-I27 at 23 °C, with a pulling speed of 600 nms⁻¹

The graph plots each peak recorded in the experiment according to the F_U and P₂P values. The median values are shown in red, with error bars showing the standard deviation of each parameter.

These results are plotted as histograms in figure 3.4. The bin size was selected based on the range of the thermal noise in the FX experiments, here 10 pN was used for the unfolding force and 0.5 nm for the interpeak distance. Fitting the histograms with a

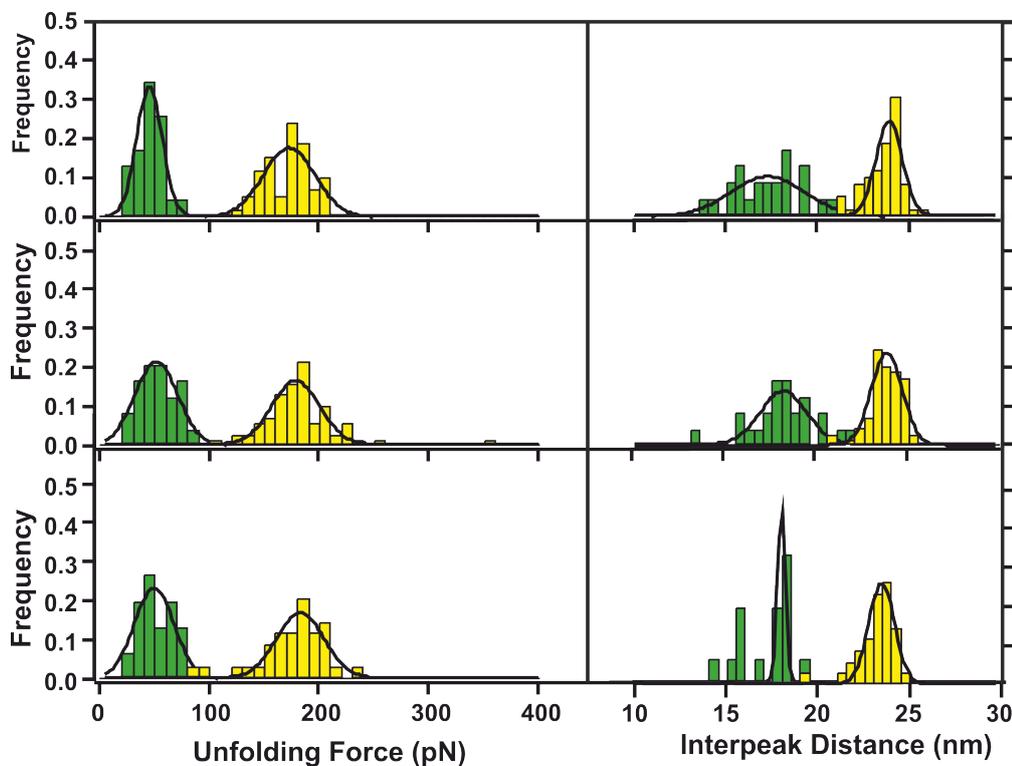


Figure 3.4: Histograms of the unfolding forces and inter-peak distances of (I27-*BsCSP*)₃-I27 unfolding events from three experiments, measured at 23 °C and with a pulling speed of 600 nms⁻¹. Bin sizes of 10 pN were chosen based on the thermal noise of the experiment and the data for each domain parameter was fit to a Gaussian distribution (black).

Gaussian distribution allows assessment as to whether enough data was collected during the experiment for a statistically significant representation of the stochastic process. The interpeak distance of the *Bs*-CSP measured in the third experiment (shown bottom right of figure 3.4) does not fit well with the Gaussian distribution, however this is due to the

low forces recorded and as the other properties do follow a normal distribution, data from this experiment was accepted.

In this way, information from single protein molecule unfolding events were collated.

Pulling Speed (nms^{-1})	Rep. No.	<i>Bs</i> -CSP			I27		
		No. of peaks	Median F_U (pN)	P_2P (nm)	No. of peaks	Median F_U (pN)	P_2P (nm)
600	1	31	46 ± 13	18 ± 2	75	173 ± 24	24 ± 1
	2	24	52 ± 16	18 ± 2	71	180 ± 34	23.2 ± 0.9
	3	14	49 ± 14	17.8 ± 0.8	57	183 ± 25	23.2 ± 0.6
	Av.		49 ± 3	17.8 ± 0.3		178 ± 5	23.5 ± 0.3

Table 3.1: Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27-*Bs*CSP)₃-I27 at 600 nms^{-1} and at 23°C . Experiments in each set of conditions were performed in triplicate and the averages from these triplicates are shown in bold.

Data was obtained from at least three experiments, performed with different protein surfaces and AFM cantilevers, with median values that agreed within error in each set of conditions. This is shown in table 3.1, where all of the data shown in figures 3.3 and 3.4 is shown. This controlled for potential systematic error in single experimental set ups. The agreement between the three experiments verifies the measurement of a robust mechanical fingerprint of the (I27-*Bs*CSP)₃-I27 unfolding.

Once this analysis has been performed and the results of each experiment verified, the

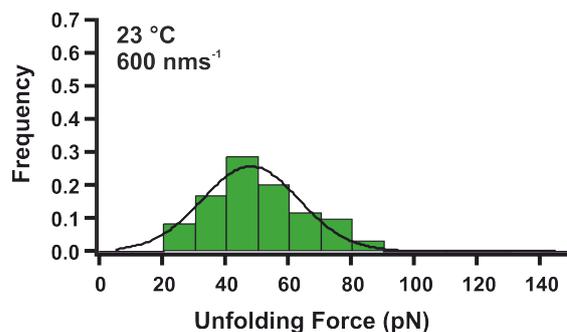


Figure 3.5: Histogram showing all 69 of the unfolding forces measured in *Bs*-CSP unfolding events in the three experiments performed at 23°C with a pulling speed of 600 nms^{-1} . Data is fit with a Gaussian distribution (black).

data can be pooled to look for features of the overall distribution. The histogram for the F_U of all the unfolding events measured in these three experiments is shown in figure 3.5. The unfolding force measured fits well with a Gaussian, shown in black, and the

centre of this Gaussian fit, 48 ± 1 pN, matches well with the average value of the three experiments, 49 ± 3 pN.

3.4 Obtaining a Mechanical Fingerprint of the *Bs*-CSP Unfolding in the Presence of ssDNA

ssDNA was used in this study as an alternative to RNA, having been shown to bind in the same manner to the same region of the protein^[58]. By adding ssDNA to the buffer in the fluid cell at an appropriate concentration, the experiments were performed in conditions that should lead to ssDNA bound to the *Bs*-CSP in the polyprotein, as shown in figure 3.6a. To ensure that the majority of *Bs*-CSP proteins would be in the bound state, a concentration of ssDNA that exceeded the *Bs*-CSP concentration by at least 10 times the K_D was used, as described in chapter 4. The protein concentration in solution was known before being bound to the gold surface, and as a large portion of this protein will not bind to the surface, the surface concentration is unknown, but must be less than the original concentration in solution. This concentration was used to calculate the sufficient ssDNA concentration required, usually around 3 μ M.

FX experiments were carried out as before, with only the ssDNA concentration in the buffer changed. An example trace is shown in figure 3.6b. The WLC fits are the same as for the free *Bs*-CSP and the data is processed in the same way as described in section 3.3. The similarity of the fits to WLC in the absence and presence of ssDNA shows that the binding of ssDNA does not change the position of the region of the protein between the N and C termini that is pulled apart, termed the mechanical clamp, of the *Bs*-CSP.

3.4.1 Identification of Appropriate ssDNA Oligonucleotide Sequence for further SMFS Experiments

Three different ssDNA oligonucleotides with varying affinities were used in these measurements, as described in table 3.2. The dT7 and CT2 oligo both contain the high

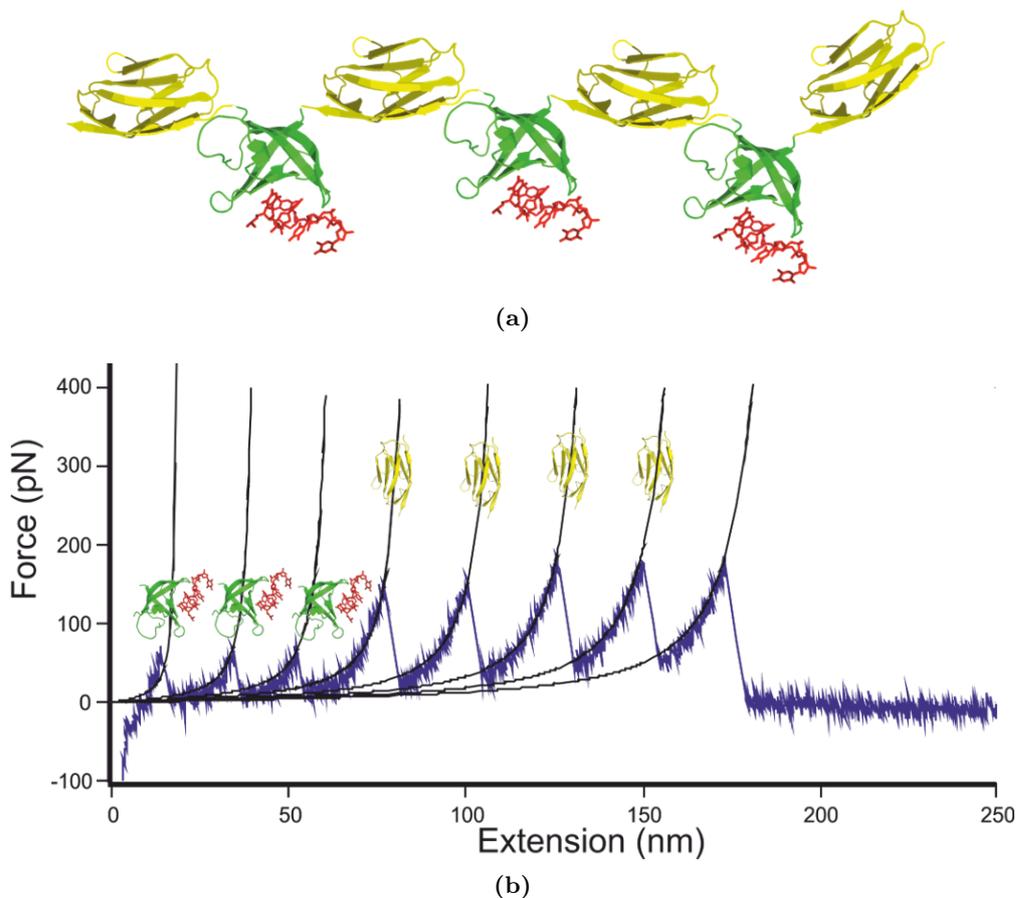


Figure 3.6: Force Extension trace of the chimeric polyprotein (a) $(I27-BsCSP)_3-I27$ consisting of 4 I27 domains (yellow) and 3 *Bs*-CSP domains (green) in the presence of ssDNA (red) (b) An example FX trace measured at 200 nms^{-1} contains unfolding peaks for all 7 domains and a detachment peak. Each peak has been identified with a corresponding domain by fits to the WLC (black).

Oligo Label	Sequence	K_D (at 15 °C)
dA7	AAAAAAA	NA
CT2	CTTTTTC	$33.7 \pm 4.1 \text{ nM}$
dT7	TTTTTTT	$1.8 \pm 0.4 \text{ nM}$

Table 3.2: ssDNA oligonucleotides used in SMFS unfolding experiments. K_D values are taken from values measured by fluorescence quenching by Max 2006^[62]. dA7 has too low an affinity for it to be determined

affinity thymine region, but CT2 has two cytosines at each termini reducing the overall affinity to the *Bs*-CSP. dA7 contains purely adenine nucleotides which do not bind strongly with the *Bs*-CSP. This oligo was used to assess whether any effects observed in the presence of ssDNA are due to binding or macromolecular crowding^[103].

The measured unfolding forces and interpeak distances of the domains in the (I27-

ssDNA	Rep. No.	<i>Bs</i> -CSP			I27		
		No. of peaks	Median F_U (pN)	P_2P (nm)	No. of peaks	Median F_U (pN)	P_2P (nm)
CT2	1	20	63 ± 19	18.3 ± 0.9	30	176 ± 30	23.6 ± 0.9
	2	19	68 ± 14	18 ± 1	37	172 ± 30	23.5 ± 0.6
	3	39	59 ± 13	18 ± 1	58	169 ± 30	23.3 ± 0.9
	Av.		63 ± 4	18.1 ± 0.2		172 ± 3	23.5 ± 0.2
dT7	1	11	62 ± 11	18.4 ± 0.6	21	172 ± 26	23.3 ± 0.5
	2	19	56 ± 11	18.3 ± 0.9	27	172 ± 25	23.4 ± 0.4
	3	15	64 ± 18	18 ± 1	25	189 ± 27	23.4 ± 0.5
	Av.		61 ± 4	18.3 ± 0.1		180 ± 12	23.4 ± 0.1
dA7	1	37	45 ± 14	18 ± 2	41	171 ± 22	23.6 ± 0.8
	2	17	33 ± 14	17 ± 1	20	155 ± 17	23.7 ± 0.9
	3	24	47 ± 19	18 ± 1	34	176 ± 19	23.4 ± 0.7
	Av.		42 ± 8	17.6 ± 0.2		166 ± 14	23.6 ± 0.3

Table 3.3: Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27-*BsCSP*)₃-I27 at 23 °C and with a pulling speed of 600 nms⁻¹. Oligonucleotides were added to buffer at a concentration of around 3 μM. Experiments in each set of conditions were performed in triplicate and the averages from these triplicates are shown in bold.

BsCSP)₃-I27 in the presence of the ssDNA oligonucleotides are shown in table 3.3. The I27 domain acts as a second control, as it does not bind ssDNA it should not be affected by the presence of oligonucleotides, and this can be seen in the median I27 unfolding forces. These unfolding forces all agree with the value measured with no ssDNA present (table 3.1) within the error.

Figure 3.7 shows the median unfolding force measured from the *Bs*-CSP domain in the absence of ssDNA and in the presence of the three oligonucleotides described in table 3.2. The measurements recorded in the presence of dT7 and CT2, two oligonucleotides with high affinity (table 3.2), result in a higher unfolding force of the *Bs*-CSP than when it is free. This increase of 14 pN for the CT2, is almost a 30 % increase in the unfolding force of the *Bs*-CSP in these conditions. As the presence of the low affinity oligonucleotide dA7 does not result in an increase in unfolding force of the *Bs*-CSP, it can be inferred that it is the binding of ssDNA causing this effect, rather than any molecular crowding.

CT2 was selected as the *Bs*-CSP binding oligonucleotide as it demonstrated a measurable effect on the unfolding force of the *Bs*-CSP, as can be seen in the histogram of the pooled data in figure 3.8, and more data was collected from experiments with CT2

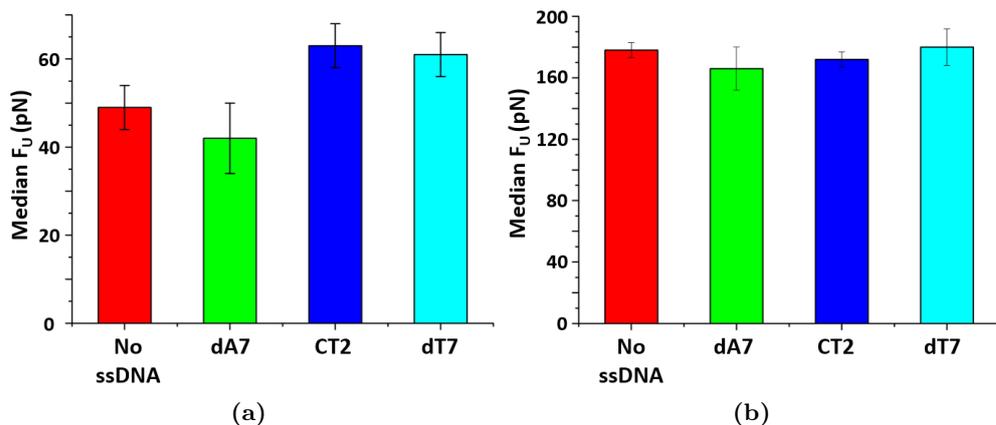


Figure 3.7: Comparison of the median (a) *Bs*-CSP and (b) I27 unfolding forces obtained from three experiments in the absence of ssDNA (red) and in the presence of dA7 (green), CT2 (dark blue) and dT7 (pale blue). FX experiments performed with a pulling speed of 600 nms^{-1} at $23 \text{ }^\circ\text{C}$. Errors represent the standard deviation of the triplicate experiments.

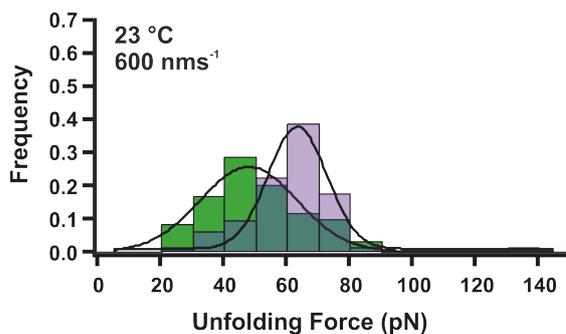


Figure 3.8: Histogram showing all 78 of the unfolding forces measured in *Bs*-CSP unfolding events in the three experiments performed in the presence of CT2 ssDNA (purple) and in the absence of ssDNA (green from figure 3.5) at $23 \text{ }^\circ\text{C}$ with a pulling speed of 600 nms^{-1} . Data is fit with a Gaussian distribution (black).

than dT7. This might be due to the bridging of CSP domains, observed in the crystal structure of the *Bs*-CSP-dT6 complex^[62], obscuring unfolding peaks in the FX traces.

3.5 Determining the Pulling Speed Dependence of the *Bs*-CSP Unfolding Force

The relationship of the unfolding force of *Bs*-CSP to the pulling speed of the SMFS experiments is determined by performing the experiment using a range of pulling speeds.

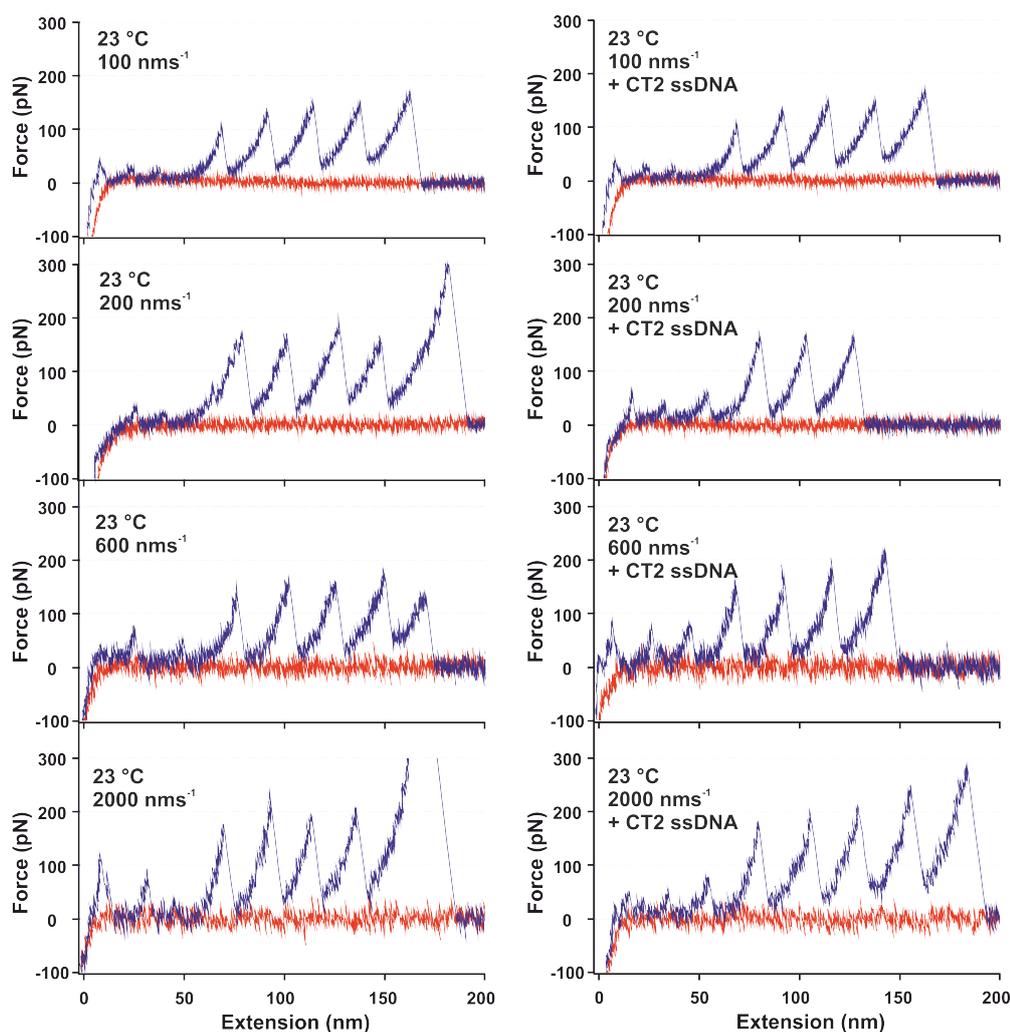


Figure 3.9: Example FX traces recorded at 100, 200, 600 and 2000 nms^{-1} at 23 °C both in the absence of ssDNA (left) and in the presence of CT2 ssDNA (right). Red line shows the approach of the cantilever to the surface and the blue line shows the retraction from the surface.

FX experiments were performed at 100, 200, 600 and 2000 nms^{-1} at 23 °C. Phosphate buffer (65 mM NaP, pH 7.4) was used and the experiments were set up in the fluid cell in the presence and absence of CT2 ssDNA at a concentration of 3 μM . Example traces from each of these experiments are shown in figure 3.9.

Traces shown demonstrate the range of unfolding events seen in the FX traces. To ensure that non-specific interactions between the AFM tip and the gold surface were not included in analysis, traces were only included in analysis if two I27 unfolding events were seen. This means that traces with 3-7 unfolding events were included if at least 2

of those unfolding events were identified as I27 domains.

All of the results from these experiments are shown in tables 3.4 and 3.5. Histograms of all the data recorded at room temperature can be found in figure B.1.

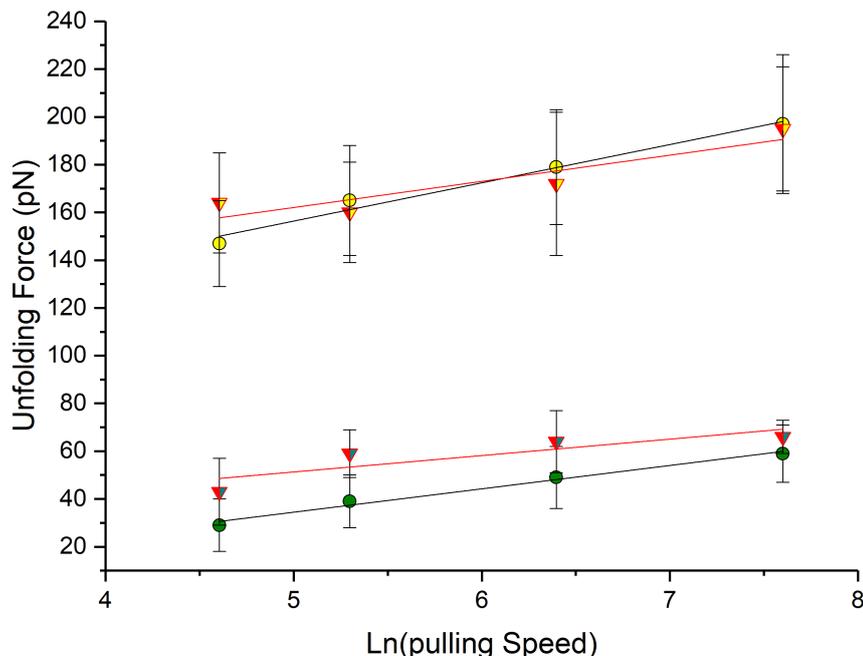


Figure 3.10: Pulling speed dependence of the unfolding force recorded for the (I27-*BsCSP*)₃-I27 at 23 °C. The F_U for both I27 (yellow) and *BsCSP* (green) are shown in the absence (black outline) and presence (red outline) of the ssDNA oligo CT2. Lines show linear unweighted fits for each domain. Error bars show the standard deviation from three experimental repeats, unless this value was below 5 pN as this is the measurement limit of the instrument in these conditions.

The median values of unfolding force shown in tables 3.4 and 3.5 is represented in figure 3.10. Both domains show the expected increase in F_U with increasing pulling speed. The I27 F_U values recorded in the absence and presence of CT2 ssDNA agree within error at every speed. The *BsCSP* domain shows an increase in F_U on addition of the ssDNA oligonucleotides, that appears to increase at slower pulling speeds. At 100 nm s^{-1} , the slowest pulling speed, this difference is 14 pN, which is almost a 50 % increase in F_U from the free *BsCSP*. The low forces at this speed means that this difference in F_U lies just outside the error.

Pulling Speed (nms ⁻¹)	Rep. No.	<i>Bs</i> -CSP			I27		
		No. of peaks	Median F_U (pN)	P_2P (nm)	No. of peaks	Median F_U (pN)	P_2P (nm)
100	1	21	31 ± 16	18 ± 1	28	147 ± 18	23.2 ± 0.5
	2	13	28 ± 11	18 ± 1	24	145 ± 18	23.4 ± 0.6
	3	18	28 ± 11	17 ± 2	27	149 ± 19	23.5 ± 0.9
	Av.		29 ± 2	17.7 ± 0.8		147 ± 2	23.3 ± 0.1
200	1	13	32 ± 14	17.9 ± 0.4	30	163 ± 25	23.0 ± 0.6
	2	18	42 ± 11	17.8 ± 0.7	28	171 ± 25	23.2 ± 0.5
	3	23	45 ± 14	18.0 ± 0.8	29	162 ± 23	23.2 ± 0.5
	Av.		39 ± 7	17.9 ± 0.6		165 ± 5	23.1 ± 0.5
600	1	31	46 ± 13	18 ± 2	75	173 ± 24	24 ± 1
	2	24	52 ± 16	18 ± 2	71	180 ± 34	23.2 ± 0.9
	3	14	49 ± 14	17.8 ± 0.8	57	183 ± 25	23.2 ± 0.6
	Av.		49 ± 3	17.8 ± 0.3		178 ± 5	23.5 ± 0.3
2000	1	20	66 ± 20	18 ± 1	47	198 ± 29	24 ± 1
	2	37	59 ± 14	18 ± 1	53	197 ± 33	23.6 ± 0.7
	3	12	52 ± 12	18.1 ± 0.7	20	198 ± 34	23.5 ± 0.6
	Av.		59 ± 7	18.0 ± 0.1		198 ± 1	23.6 ± 0.1

Table 3.4: Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27-*Bs*CSP)₃-I27 at 23 °C. Triplicate averages are shown in bold.

Pulling Speed (nms ⁻¹)	Rep. No.	<i>Bs</i> -CSP			I27		
		No. of peaks	Median F_U (pN)	P_2P (nm)	No. of peaks	Median F_U (pN)	P_2P (nm)
100	1	16	42 ± 17	18 ± 1	20	161 ± 21	23.4 ± 0.5
	2	22	44 ± 14	17.9 ± 0.9	31	163 ± 22	23.3 ± 0.5
	3	32	44 ± 14	17.8 ± 0.9	41	168 ± 21	23.5 ± 0.6
	Av.		43 ± 1	17.9 ± 0.1		164 ± 4	23.4 ± 0.1
200	1	11	62 ± 10	18.0 ± 0.6	17	153 ± 21	23.2 ± 0.4
	2	16	65 ± 10	17.8 ± 0.4	22	175 ± 22	23.3 ± 0.4
	3	23	50 ± 13	18.1 ± 0.9	23	151 ± 29	23.3 ± 0.5
	Av.		59 ± 8	18.0 ± 0.6		160 ± 13	23.3 ± 0.4
600	1	20	63 ± 19	18.3 ± 0.9	30	176 ± 30	23.6 ± 0.9
	2	19	68 ± 14	18 ± 1	37	172 ± 30	23.5 ± 0.6
	3	39	59 ± 13	18 ± 1	58	169 ± 30	23.3 ± 0.9
	Av.		63 ± 4	18.1 ± 0.2		172 ± 3	23.5 ± 0.2
2000	1	23	65 ± 7	18.1 ± 0.5	36	187 ± 30	23.2 ± 0.4
	2	14	65 ± 12	18.0 ± 0.7	33	199 ± 42	23.3 ± 0.5
	3	43	67 ± 16	18.4 ± 0.5	65	198 ± 26	23.3 ± 0.4
	Av.		66 ± 1	18.2 ± 0.2		195 ± 7	23.3 ± 0.1

Table 3.5: Unfolding forces (F_U) and Interpeak Distances (P_2P) measured for the (I27-*Bs*CSP)₃-I27 in the presence of ssDNA at 23 °C. Oligo CT2 were added to buffer at a concentration of 3 μM. Triplicate averages are shown in bold.

3.5.1 The Affect of Lower Temperature on Unfolding Force of the *Bs*-CSP

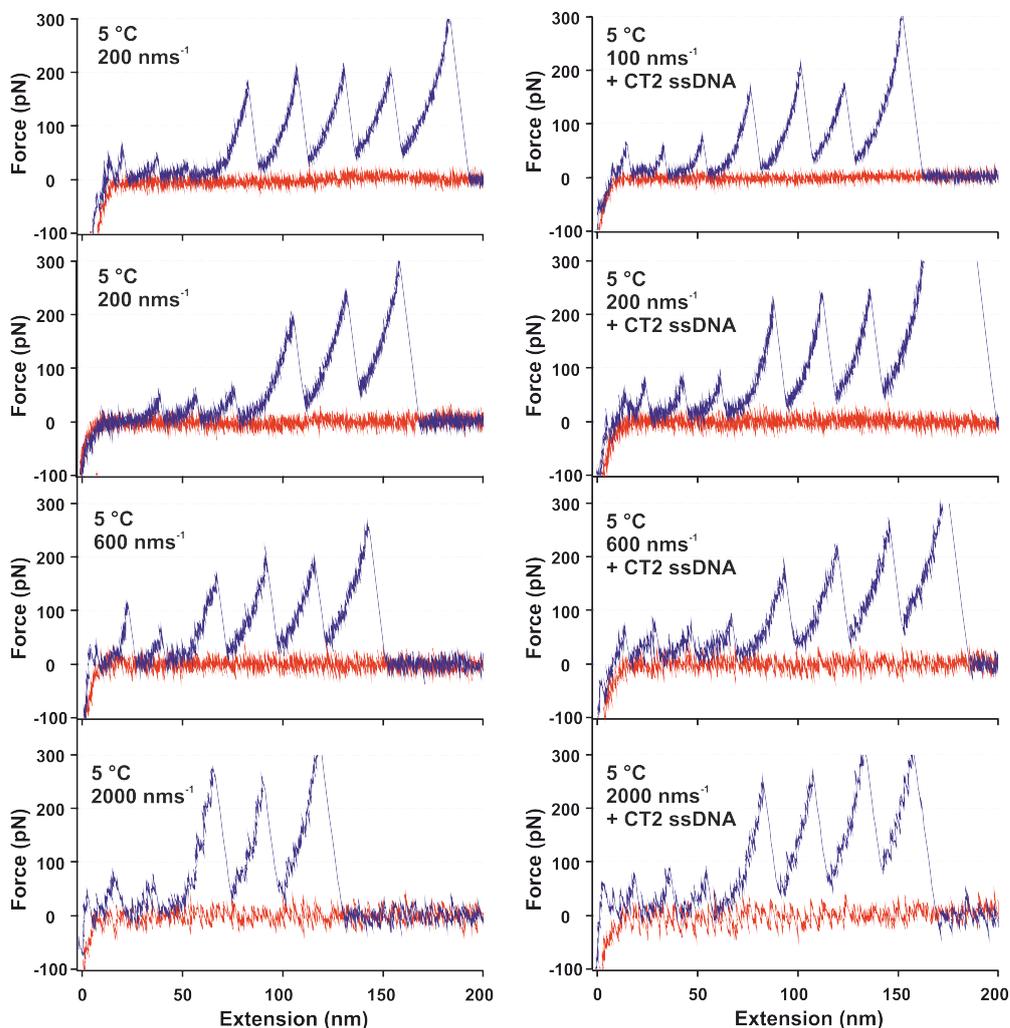


Figure 3.11: Example FX traces recorded at 100, 200, 600 and 2000 nms⁻¹ at 5 °C both in the absence of ssDNA (left) and in the presence of CT2 ssDNA (right). Red line shows the approach of the cantilever to the surface and the blue line shows the retraction from the surface.

At lower temperatures, the force at which proteins mechanically unfold is generally increased due to the lowering of probability that the thermal fluctuations will induce unfolding while the protein is under mechanical stress^[161]. This includes CSPs, with higher temperatures inducing a mechanical softening for the CSP from a hyperthermophilic organism^[7]. The binding affinity between the *Bs*-CSP and the ssDNA is also increased at lower temperatures^[59], and the thermal noise seen in the AFM is decreased.

For these reasons, the investigation was repeated at 5 °C, to explore if the change in F_U in the presence of ssDNA would increase at lower temperatures. Example traces from each pulling speed measured at 5 °C are shown in figure 3.11.

These FX traces have clearer peaks for each unfolding event due to the reduction in thermal noise and increase in protein mechanical stability. These experiments were performed in the absence (figure 3.11 left) of ssDNA, and in the presence of CT2 ssDNA oligonucleotides (figure 3.11 right). The data was processed as described previously and all of the results are summarised in tables 3.6 and 3.7. Histograms for all of the data recorded at 5 °C can be found in figure B.2.

The resulting median values of F_U shown in tables 3.6 and 3.7 are plotted against the pulling speed in figure 3.12. At 5 °C the impact of ssDNA binding on the *Bs*-CSP

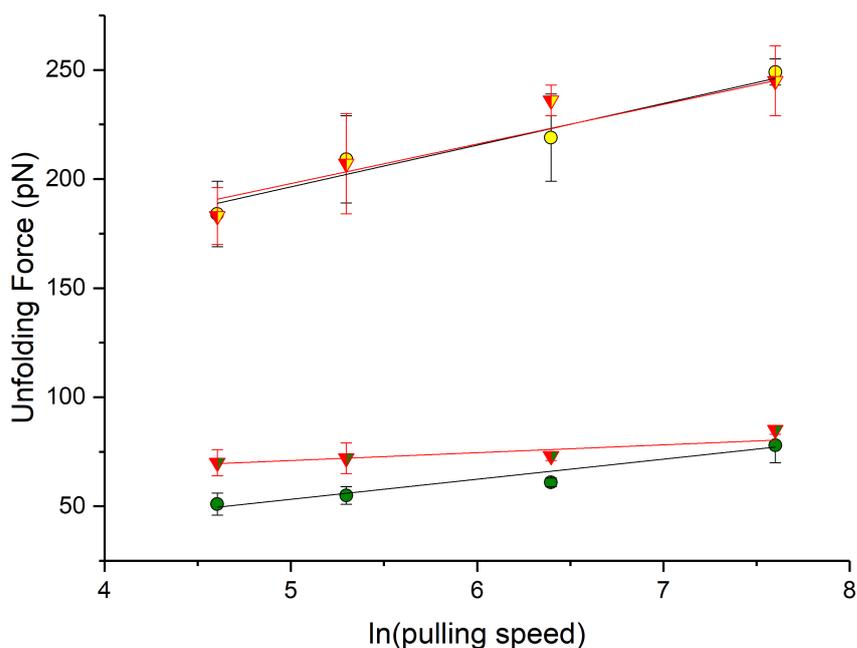


Figure 3.12: Pulling speed dependence of the unfolding force recorded for the (I27-*Bs*CSP)₃-I27 at 5 °C. The F_U for both I27 (yellow) and *Bs*-CSP (green) are shown in the absence (black outline) and presence (red outline) of the ssDNA oligo CT2. Lines show linear unweighted fits for each domain.

unfolding force is clear, as can be seen in figure 3.12. The stabilising effect increases at lower pulling speeds, so that at 100 nm s^{-1} the median F_U measured increases from 51 to 71 pN. This is a 40 % increase in unfolding force of the *Bs*-CSP, an effect which is

Pulling Speed (nms ⁻¹)	Rep. No.	<i>Bs</i> -CSP			I27		
		No. of peaks	Median F_U (pN)	P ₂ P (nm)	No. of peaks	Median F_U (pN)	P ₂ P (nm)
100	1	14	48 ± 15	17 ± 1	16	174 ± 23	23.5 ± 0.6
	2	11	48 ± 16	18.2 ± 0.8	17	178 ± 28	23.6 ± 0.7
	3	16	56 ± 15	17.8 ± 0.5	19	201 ± 20	23.4 ± 0.5
	Av.		51 ± 5	17.8 ± 0.8		184 ± 15	23.5 ± 0.6
200	1	16	51 ± 14	18 ± 1	20	187 ± 41	23.3 ± 0.6
	2	13	55 ± 20	18 ± 1	21	215 ± 32	23.6 ± 0.4
	3	22	58 ± 13	17.8 ± 0.8	32	225 ± 25	23.5 ± 0.5
	Av.		55 ± 4	17.9 ± 0.3		209 ± 20	23.5 ± 0.2
600	1	18	62 ± 17	18.0 ± 0.4	22	239 ± 45	23.9 ± 0.4
	2	12	59 ± 9	18.0 ± 0.4	20	200 ± 33	23.8 ± 0.9
	3	16	64 ± 13	18.3 ± 0.5	24	225 ± 37	23.8 ± 0.6
	Av.		61 ± 3	18.1 ± 0.2		221 ± 20	23.8 ± 0.1
2000	1	12	80 ± 19	17.8 ± 0.8	19	248 ± 48	24 ± 1
	2	23	68 ± 12	18.4 ± 0.8	37	245 ± 45	24.1 ± 0.7
	3	30	84 ± 19	18 ± 2	47	256 ± 42	23.9 ± 0.8
	Av.		78 ± 8	18.2 ± 0.3		249 ± 6	24.0 ± 0.1

Table 3.6: Unfolding forces (F_U) and Interpeak Distances (P₂P) measured for the (I27-*Bs*CSP)₃-I27 at 5 °C. Triplicate averages are shown in bold.

Pulling Speed (nms ⁻¹)	Rep. No.	<i>Bs</i> -CSP			I27		
		No. of peaks	Median F_U (pN)	P ₂ P (nm)	No. of peaks	Median F_U (pN)	P ₂ P (nm)
100	1	14	73 ± 21	17.9 ± 0.9	15	170 ± 20	23.2 ± 0.4
	2	13	63 ± 14	18.4 ± 0.9	21	182 ± 31	23.2 ± 0.6
	3	17	75 ± 11	18.0 ± 0.6	24	197 ± 30	23.4 ± 0.5
	Av.		71 ± 6	18.1 ± 0.3		183 ± 13	23.2 ± 0.1
200	1	14	76 ± 19	18.6 ± 0.4	16	199 ± 18	23.6 ± 0.6
	2	13	75 ± 7	18.2 ± 0.5	13	233 ± 25	23.7 ± 0.4
	3	12	64 ± 12	18 ± 1	13	190 ± 18	23.8 ± 0.5
	Av.		72 ± 7	18.2 ± 0.4		207 ± 22	23.7 ± 0.1
600	1	23	73 ± 13	18.1 ± 0.7	27	243 ± 33	23.5 ± 0.6
	2	20	75 ± 13	18.3 ± 0.4	26	237 ± 36	23.8 ± 0.4
	3	22	70 ± 16	18.3 ± 0.8	29	229 ± 23	23.8 ± 0.5
	Av.		73 ± 2	18.2 ± 0.1		236 ± 7	23.7 ± 0.2
2000	1	48	83 ± 24	18.4 ± 0.9	57	230 ± 46	24 ± 1
	2	25	84 ± 16	18.7 ± 0.8	44	243 ± 33	24.1 ± 0.9
	3	16	87 ± 17	18.3 ± 0.9	24	261 ± 41	23.9 ± 0.7
	Av.		85 ± 2	18.5 ± 0.2		245 ± 16	23.8 ± 0.3

Table 3.7: Unfolding forces (F_U) and Interpeak Distances (P₂P) measured for the (I27-*Bs*CSP)₃-I27 at 5 °C in the presence of the ssDNA oligo CT2. Triplicate averages are shown in bold.

not seen at the higher pulling speed of 2000 nms^{-1} , where the difference is reduced to 7 pN, just outside the thermal noise of the AFM.

3.5.2 Determining the Effect of Low Temperature on the *Bs*-CSP Mechanical unfolding in the Presence and Absence of ssDNA Binding

All of the *Bs*-CSP unfolding events were pooled and the resulting histograms are shown in figure 3.13. Due to the stochastic nature of protein unfolding and the small forces involved, there has to be a statistically significant number of unfolding events recorded. By pooling the data from all experiments and plotting histograms, the distribution can be assessed. The resulting histograms in figure 3.13 all fit the Gaussian distribution. While the experiments performed with ssDNA present do show an increase in the average unfolding force, the maximum measured unfolding force without ssDNA present is generally just as high, or higher. This suggests that the ssDNA binding could be stabilising a conformational state that the *Bs*-CSP does inhabit briefly as it fluctuates in the free state.

The medians of *Bs*-CSP unfolding force are plotted against the natural log of the pulling speed (in nms^{-1}) in figure 3.14. Both free proteins show a similar relation between F_U and pulling speed in the free state, with higher forces measured at the lower temperature. Equally, the speed dependence of F_U for the *Bs*-CSP in complex with ssDNA is similar at both temperatures, with a higher F_U at lower temperatures. These linear fits were used to match results of the Monte Carlo simulations.

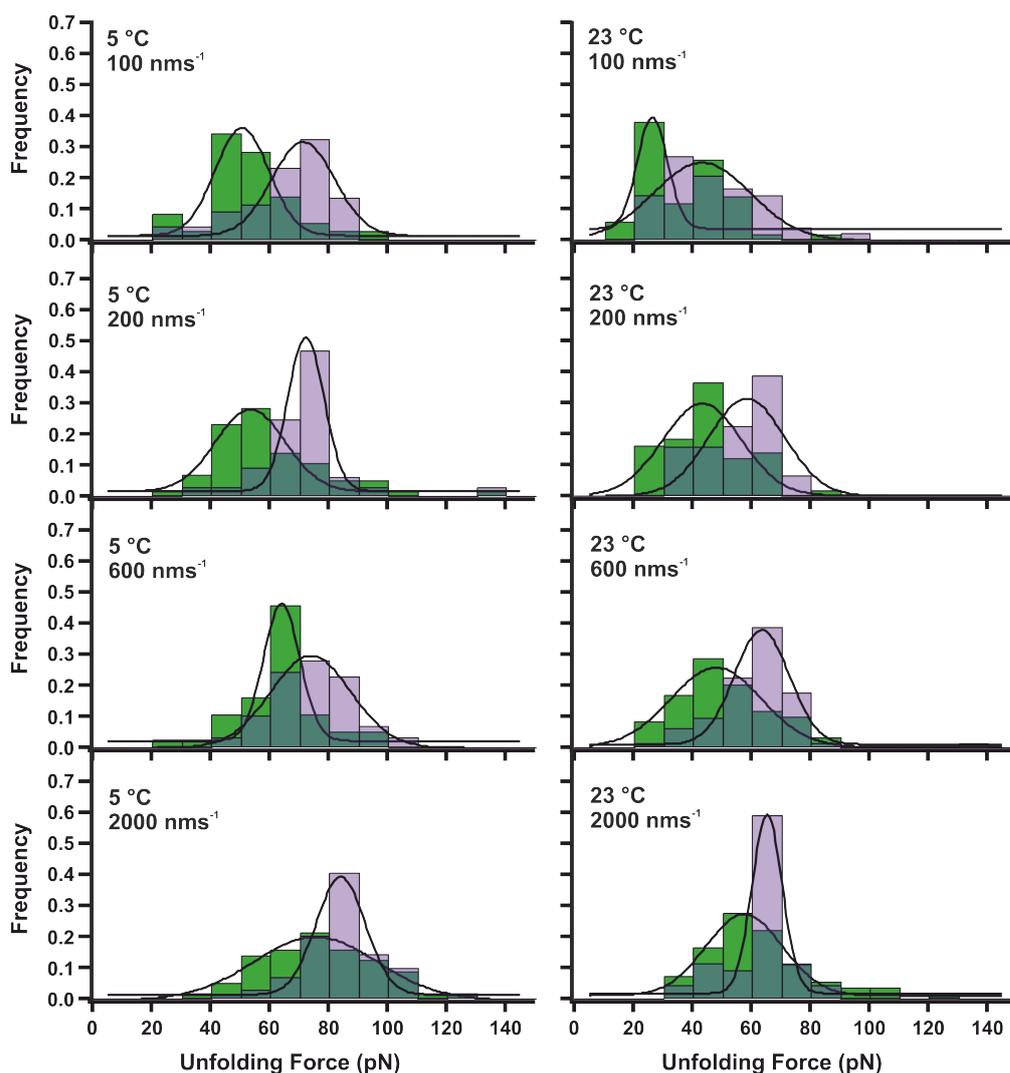


Figure 3.13: Histograms of the unfolding force of the *Bs*-CSP in all conditions, comparing free *Bs*-CSP (green) and with CT2 ssDNA bound *Bs*-CSP (purple). A bin size of 10 pN was chosen to account for thermal noise and each distribution was fit to a Gaussian (black).

3.6 Determining Parameters of the Underlying Unfolding Energy Landscapes of the *Bs*-CSP using Monte Carlo (MC) Simulations

The dependence of unfolding force on the pulling speed can be modelled by particular properties of the shape of the unfolding energy landscape described in the Bell-Evans-Ritchie Model^[147]. To explore the underlying energy landscapes of these unfolding

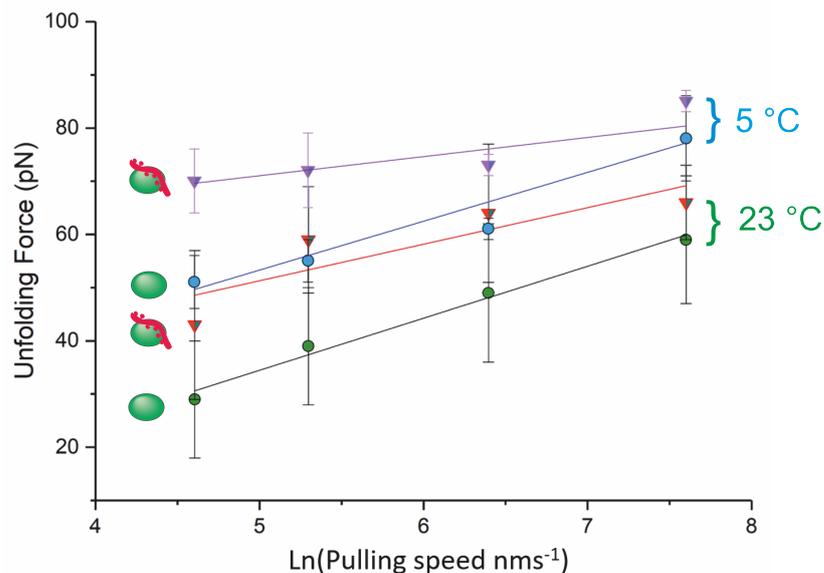


Figure 3.14: The median unfolding forces of the *Bs*-CSP in all conditions plotted against the natural log of the pulling speed. The *Bs*-CSP unfolding forces measured with no ssDNA present are shown for the 23 °C experiment (green) and the 5 °C experiment (blue) and with CT2 ssDNA present at 23 °C (red outline) and 5 °C (purple outline) with corresponding linear fits.

events, MC simulations were used as described in section 2.5.7. The parameters, k_U and Δx_U , of the energy landscape were adjusted and simulations performed to produce unfolding force values for both *Bs*-CSP and I27 domains at pulling velocities covering the range used in experiments. The resulting pulling speed dependence of the unfolding forces were used to find conditions that match the experimental results.

The linear fits shown in figures 3.10 and 3.12 were used to match the MC simulation results, and the pulling speed dependence of unfolding force for the *Bs*-CSP in all conditions is shown in figure 3.14. The standard error of these linear fits were used to determine the uncertainties of calculated parameters parameters.

The underlying energy landscapes derived from MC simulations matching data at 23 °C and 5 °C are shown in figures 3.15a and 3.15b respectively. Uncertainties calculated from the maximum and minimum linear fits within the standard error are shown in table B.1. At both temperatures the ssDNA binding shows a stabilising effect that both increases the energy difference between the folded and transition state, ΔG_U , and

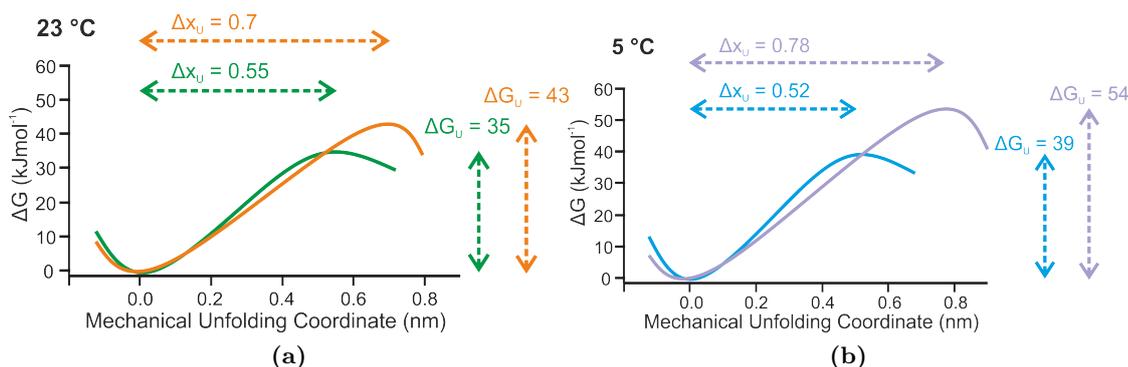


Figure 3.15: Underlying energy landscapes calculated from experiments (a) at 23 °C in the absence (green) and the presence (orange) of the CT2 ssDNA oligo. (b) At 5 °C in the absence (blue) and presence (purple) of the CT2 ssDNA oligo

increases the distance to this state, Δx_U . At 23 °C the increase in ΔG is around 8 kJmol⁻¹ and at 5 °C it is around 5 kJmol⁻¹. The effect of binding on this value does not change much with temperature, but the difference in Δx_U is clear. At 23 °C this increases by 0.15 nm, and at 5 °C it increases by 0.26 nm.

Figure 3.16 compares the energy landscapes at 5 and 23 °C of the free protein and the

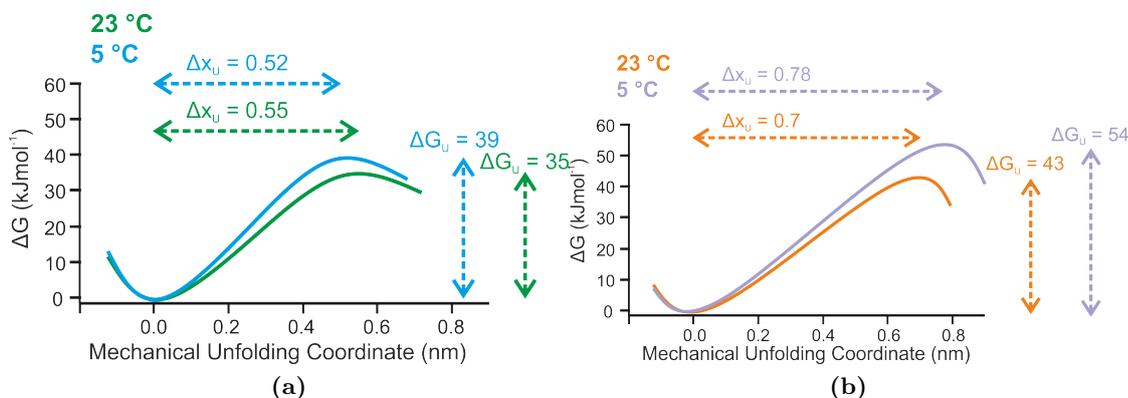


Figure 3.16: Underlying energy landscapes calculated from experiments (a) with no ssDNA present at 23 °C (green) and at 5 °C (blue). (b) In the presence of the CT2 ssDNA oligo at 5 °C (purple) and 23 °C (orange)

complex. The temperature appears to have more of an impact on the complex energetics than on the free protein. In the free protein Δx_U decreases with decreasing temperature, which is the expected relation between Δx_U and temperature in proteins^[161]. This behaviour has been measured previously in the hyperthermophile CSP *Tm*-CSP^[7].

The unfolding of the complexed *Bs*-CSP shows the opposite Δx_U relation to temperature, as the value of Δx_U increases at the lower temperature.

The value of ΔG_U increases slightly at lower temperatures for both the free and complexed *Bs*-CSP. This change is slightly larger in the complex, possibly due to the higher affinity between ssDNA and *Bs*-CSP at lower temperatures.

3.7 Discussion

SMFS measurements on the unfolding of the *Bs*-CSP in the absence and presence of binding ssDNA oligonucleotides found a clear mechanical fingerprint for both systems. Molecular dynamics simulations have identified the probable location of the mechanical clamp in the *Bs*-CSP across β -strands 1, 4 and 5, when it is pulled apart by the N and C termini^[17]. This is far from the ssDNA binding site identified in the crystal structure^[62], on the opposite side of the protein surface. The ΔL_C measured here in both the presence and absence of ssDNA was conserved, showing that ssDNA binding does not change the location of this mechanical clamp. My results show that the conformation of the protein in the presence of ssDNA is still mechanically stabilised by the same mechanical clamp region as it is in the free state.

Measurements of the unfolding force of the *Bs*-CSP revealed an increase in the force required to unfold the protein when ssDNA was present. This stabilisation effect increased the median of the *Bs*-CSP unfolding force by up to 20 pN. To determine if this effect was due to macromolecular crowding, a low affinity oligonucleotide was used in control experiments. Only ssDNA that had a high affinity with the *Bs*-CSP caused an increase in unfolding force, indicating this effect as a consequence of ssDNA binding. Previously, the binding of ssDNA has been shown to increase the mechanical unfolding force of the RRM1 domain of the TDP-41 protein^[112], as discussed in section 1.4.3.6. In this system the ssDNA acted as a mechanical staple, drastically increasing the unfolding force by binding across the mechanical clamp and protecting the domain from unfolding. In my results, the conservation of the ΔL_C suggests that it is not ssDNA

binding across the mechanical clamp causing the increase in F_U . Instead the binding of ssDNA stabilises a protein conformation with a stronger mechanical clamp region.

By completing a pulling dependence study of the mechanical stability of the protein in the absence and presence of ssDNA, it was found that the increase in stability upon ssDNA binding was greater at lower pulling speeds. The effect caused upto a 50 % increase in the unfolding force. This could be due to the ssDNA forming new bonds as the protein is unfolding. The Csp is a small protein, that can bind an oligonucleotide chain nonspecifically, so as the protein is stretched the changing shape of the binding site simply causes bonds to slide along the ssDNA. With the oligonucleotides stabilising each intermediate unfolding structure, as observed in FC measurements on the *Tm*-CSP unfolding^[84]. At fast speeds there is less time for these bonds to be formed so the stabilisation is not measured. At slower speeds the flexible ssDNA could help in holding the stretched structure together, so rather than acting like a mechanical “staple” in this complex, the ssDNA acts like an elastic band or a velcro strap. To test for this, shorter oligonucleotides could be used, however the binding affinity rapidly reduces as the oligonucleotide size is taken below the 6-7 nucleotide bases that cover the binding site^[59], so this would need to be accounted for.

To explore this effect further, the experiments were repeated at a reduced temperature of 5 °C. This reduction in temperature also increases the affinity between the *Bs*-CSP and ssDNA oligonucleotides. In these conditions the increase in unfolding force of the *Bs*-CSP on ssDNA binding was observed to increase at lower pulling speeds.

The temperature dependent function of the *Bs*-CSP makes low temperature behaviour a physiologically relevant condition. The CSP is induced and has to function at cold temperatures. To continue to function as a chaperone in RNA unfolding, the protein should remain flexible in cold environments. While bound to nucleic acids, the protein still requires flexibility to aid in the unfolding of RNA structures^[20]. This

provides a functional explanation as to why the protein does not become more rigid when bound to the ssDNA ligand.

To compare the underlying unfolding energy landscapes, MC simulations were performed. These were fit with the pulling speed dependence of the unfolding force to determine parameters describing the unfolding transition state. At both temperatures the binding of ssDNA increased the ΔG_{TS} of unfolding, and in both the free and complexed *Bs*-CSP the reduction of temperature increased this energy barrier as well. This cold induced increase in ΔG_{TS} was larger when the ssDNA was present.

The ssDNA binding increased the distance between the folded and transition state, Δx_U , of the protein at both temperatures. This effect was stronger at the lower temperature, where ssDNA binding with *Bs*-CSP is higher affinity. The temperature dependent change in Δx_U was found to decrease with decreasing temperature for the ligand free CSP, demonstrating the expected softening of the protein^[7]. When high affinity ssDNA was present, an increase in Δx_U with decreasing temperature was observed. This effective increase in malleability with binding of ssDNA is the opposite effect to previously recorded measurements on the influence of ligand binding on protein mechanics^[96,101,105,106,112].

4

Measuring the binding of ssDNA to the Cold Shock Protein

In this chapter, the affinity of the Cold Shock Proteins (CSPs) from the organisms *Bacillus Subtilis* (*Bs*-CSP) and *Psychrobacter 6* (*PB6*-CSP) binding to single stranded DNA (ssDNA) oligonucleotides is measured using a range of techniques and under a range of conditions. The temperature dependence of this affinity is explored using both tryptophan fluorescence titration and Isothermal Titration Calorimetry (ITC). MicroScale Thermophoresis (MST) is applied to explore potential stoichiometry effects and verify effects seen in the fluorescence quenching titrations. Finally, the potential application of Single Molecule Force Spectroscopy (SMFS) to measure the unbinding force of a CSP-ssDNA complex is explored.

The ability to bind nucleic acids is essential to Cold Shock Protein (CSP) function. As a chaperone for RNA unfolding, the affinity must be strong enough to bind and influence the RNA structure. The binding strength must also be low enough that it does not bind too tightly and render the RNA useless. Monitoring the chaperone activity of this system is difficult, as it requires a system that reports on the folded state of RNA molecules. Recently this was achieved using NMR to monitor RNA oligonucleotides with ^{13}C and

^{15}N labeled U and G nucleotides that produced a measurable signal change between the RNA hairpin state and the hetero-duplex state^[63]. The function of the Cold Shock Protein (CSP) as an RNA chaperone requires the non-specific binding of nucleic acids. To aid RNA unfolding this affinity should be stronger for the unfolded RNA state, with an energetically favourable complex of unfolded RNA. By measuring the dissociation constant (k_D) of the proteins with single stranded DNA (ssDNA) oligonucleotides, an approximation of function can be compared for different proteins and different conditions.

4.0.1 Limits of the Equilibrium Titrations

All binding interaction measurements performed were saturation methods, and the limits to these measurements are described briefly here, more detail can be found elsewhere^[162,163]. The equilibrium between free protein, P, and ligand, L, and the bound complex, PL, is described in equation 4.1.



The change in complex concentration [PL] over time is dependent on the on and off rates, k_{on} and k_{off} , and the free protein and ligand concentrations, [P] and [L].

$$\frac{d[PL]}{dt} = k_{\text{on}}[P][L] - k_{\text{off}}[PL] \quad (4.2)$$

If the system has reached equilibrium we can assume [PL] has reached the steady state

$$\frac{d[PL]}{dt} = 0 \quad (4.3)$$

$$k_{\text{on}}[P][L] = k_{\text{off}}[PL] \quad (4.4)$$

As the free protein concentration [P] equals the total protein concentration $[P_T]$ minus the complex concentration [PL], and the free ligand concentration is the total ligand

concentration $[L_T]$ minus the complex concentration.

$$\frac{d[PL]}{dt} = 0 \quad (4.5)$$

$$k_{on}([P_T] - [PL])([L_T] - [PL]) = k_{off}[PL] \quad (4.6)$$

$$([P_T] - [PL])([L_T] - [PL]) - \frac{k_{off}}{k_{on}}[PL] = 0 \quad (4.7)$$

Where the dissociation constant is defined as $K_D = \frac{k_{off}}{k_{on}}$.

$$[PL]^2 - [PL]([L_T + P_T + K_D]) + [L_T][P_T] = 0 \quad (4.8)$$

Equation 4.8 is a quadratic equation so solving this for $[PL]$ gives the final quadratic velocity equation.

$$[PL] = \frac{([P_T] + [L_T] + K_D) - \sqrt{([P_T] + [L_T] + K_D)^2 - 4[P_T][L_T]}}{2} \quad (4.9)$$

Any equilibrium measurement with a signal dependent on the complex concentration $[PL]$ can be fit with equation 4.9 and the known variables of $[P_T]$ and $[L_T]$ to find a value for K_D . If there is more than one identical binding site on either the protein or ligand, this can be incorporated by including the factor n , for n number of sites, with the protein or ligand concentration, $[P_T]$ or $[L_T]$ depending on where the extra sites are. So if there are n sites on the protein, $[P_T]$ can be replaced with $n[P_T]$.

The binding isotherm is a measure of the amount of ligands adsorbed as a function of the concentration or partial pressure of the ligand at a fixed temperature^[163]. Any method used to accurately determine the K_D requires a measurable signal from a protein concentration of the same magnitude as the dissociation constant. If the protein concentration is too high then saturation is reached very rapidly when the ligand concentration approaches the same concentration as $[P]$. This will lead to high errors in the determination of the K_D as any value lower than about 10 % of the protein concentration produce binding isotherms that are indistinguishable.

4.1 Tryptophan Fluorescence Quenching

Fluorescence spectroscopy can be used to monitor ligand binding if there is a measurable difference between the free and bound protein fluorescence signal. All of the CSPs have a surface tryptophan which produces a fluorescence signal. This signal is very sensitive to environmental changes, making tryptophan a very useful amino acid for monitoring protein folding and binding. When nucleic acids bind to the CSPs, the fluorescent signal of this surface residue becomes quenched^[61,130,164].

The tryptophan fluorescence provides a signal that is directly related to the free protein concentration $[P]$. The quenching of this signal is proportional to the fraction of protein sites bound, which can be stated as the concentration of complex over the total protein concentration $\frac{[PL]}{[P_T]}$. The tryptophan fluorescence spectra for the *Bs*-CSP

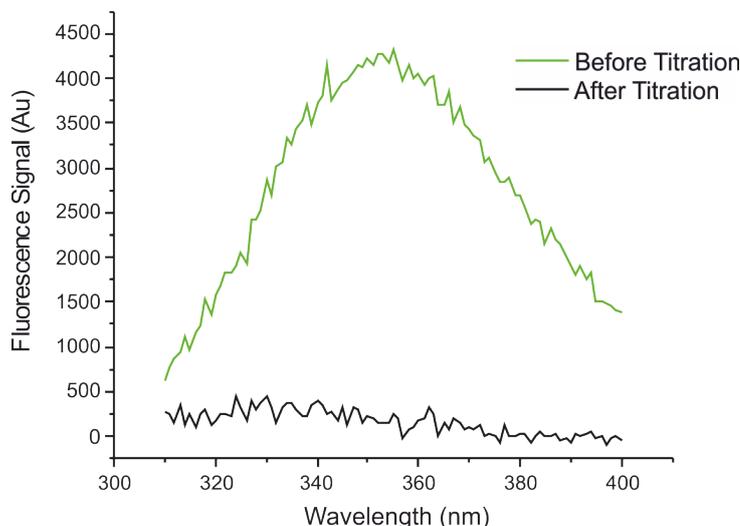


Figure 4.1: Fluorescent emission signal of the *Bs*-CSP, excited at 285 nm and at a concentration of 30 nM, before (green) and after titration (up to 1 μ M) with the dT7 oligonucleotide (black).

at 30 nM (green) and the *PB6*-CSP at 90 nM (blue) concentration, measured at 20 °C are shown in figures 4.1 and 4.2 before and after (black) the quenching titration. The protein concentration for each titration was 30 nM, 90 nM or 300 nM, depending on the corresponding K_D value to improve the precision of the fit with the tight binding

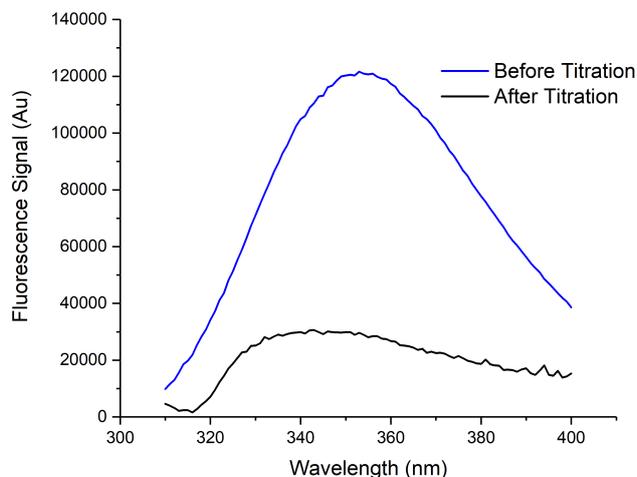


Figure 4.2: Fluorescent signal of the *PB6*-CSP, excited at 285 nm and at a concentration of 90 nM, before (blue) and after titration (up to 1 μ M) with the dT7 oligonucleotide (black).

equation. Due to the tight binding of *Bs*-CSP to dT7 in these conditions, a low protein concentration is required so the signal is lower than that for the *PB6*-CSP. This also results in the final, fully quenched signal (black) being lower for the *Bs*-CSP than for the *PB6*-CSP as there is a higher concentration of *PB6*-CSP and there is still some fluorescence signal of the protein when the titration has reached saturation. The fluorescence emission peak of the tryptophan is quenched by the presence of ssDNA, and by tracking the peak height at 345 nm the concentration of the free protein [P] can be monitored.

4.1.1 *Bs*-CSP

Titration of the *Bs*-CSP were performed with the dT7 oligonucleotide at different temperatures (10, 20 and 30 $^{\circ}$ C) and the resulting isotherms are shown in figure 4.3. The quenching after each titration step, i , was calculated from the fluorescence emission signal by $Q_i = \frac{(F_0 - F_i)}{F_0}$ where F_0 is the initial, maximum fluorescence signal. K_D values from the fits with the tight binding equation are shown in table 4.1.

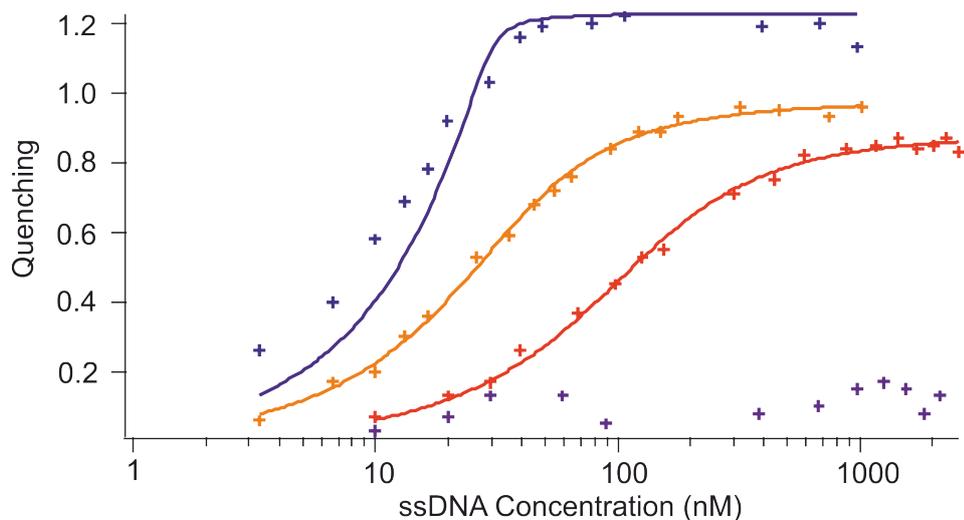


Figure 4.3: Quenching of tryptophan fluorescence in the *Bs*-CSP as dT7 is titrated into the sample. Experimental data shown by markers for experiments recorded at 10 °C (blue), 20 °C (orange) and 30 °C (red). dA7 was also titrated into the *Bs*-CSP at 10 °C (purple) but did not produce a significant quenching effect. Lines show the fits to the tight binding equation, with resulting K_D values shown in table 4.1.

Temperature (°C)	ssDNA	Dissociation Constant, K_D (nM)
10	dT7	0.3 ± 1.2
20	dT7	10.2 ± 1.5
30	dT7	48 ± 7

Table 4.1: K_D values for the *Bs*-CSP binding to ssDNA oligonucleotides determined by fits to the tight binding equation as shown in figure 4.3

Results measured in this work are comparable with those obtained by the same method at 15 °C in 50 mM cacodylate/HCl, 100 mM KCl (pH 7.0), supplemented with 40 μ M N-acetyl-tryptophanamide by Max *et al.*^[62]. The dissociation constant for dT7 with *Bs*-CSP was calculated using these results to be 1.8 ± 0.4 nM. The temperature dependence is also comparable to that measured by Zeeb *et al.* in the same buffer conditions, which measured the K_D to be 1.8 ± 0.4 nM at 15 °C, 60 ± 4 nM at 25 °C and 464 ± 31 nM at 35 °C^[59]. As the same temperature trend and magnitude of K_D is seen in these measurements and the previously published results, these titrations were used as a basis to compare the temperature dependent binding of nucleic acids between the mesophilic *Bs*-CSP and the psychrotropic *PB6*-CSP.

4.1.2 *PB6*-CSP

The titrations of *PB6*-CSP with dT7 were performed at 10, 20 and 30 °C and the resulting isotherms and fits are shown in figure 4.4. At high ssDNA concentrations the quenching was seen to be reduced. This is most likely due to an effect of stoichiometry (section 4.3). Fits to the tight binding equation were performed over the region of data where binding was 1:1, so the final points in these titrations were excluded. The resulting values of K_D are shown in table 4.2.

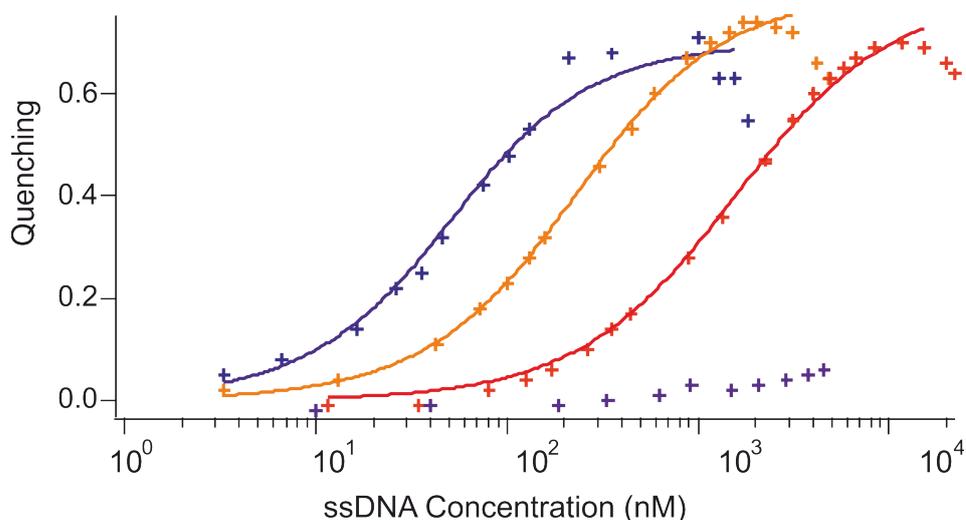


Figure 4.4: Quenching of tryptophan fluorescence in the *PB6*-CSP as dT7 is titrated into the sample. Experimental data shown by markers for experiments recorded at 10 °C (blue), 20 °C (orange) and 30 °C (red). dA7 was also titrated into the *PB6*-CSP at 10 °C (purple) but did not produce a significant quenching effect. The affinity with this oligo is too low to be fit with this method. Lines show the fits to the tight binding equation, fits shown excluded the final two points of each titration, with resulting K_D values shown in table 4.2.

Temperature (°C)	Oligo	Dissociation Constant, K_D (nM)
10	dT7	36 ± 6
20	dT7	182 ± 9
30	dT7	1400 ± 90

Table 4.2: K_D values for the *PB6*-CSP binding to ssDNA oligonucleotides determined by fits to the tight binding equation as shown in figure 4.4

Overall the *PB6*-CSP has a lower affinity for the dT7 oligonucleotide. Both the *Bs*-CSP and *PB6*-CSP display the expected, according to equation 4.13, exponential

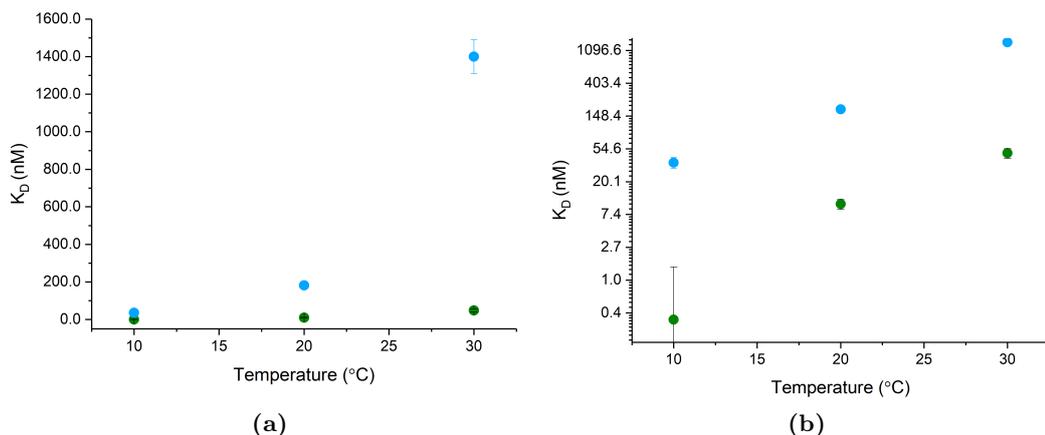


Figure 4.5: Plots showing the temperature dependence of the dissociation constants, K_D , for the proteins *Bs*-CSP (green) and *PB6*-CSP (blue) with the oligonucleotide dT7 as measured by the quenching of the tryptophan fluorescence. Shown with the (a) linear scale and (b) logarithmic scale used for the x-axis.

dependence of K_D on temperature (figure 4.5b). Taking the optimum temperature for *Bacillus Subtilis* to be 37 °C^[49] and that of the *Psychrobacter* 6 to be around 20 °C^[33], then 7-10 °C below that temperature is considered “cold-shock” conditions, under which CSPs are up regulated^[33,44,165]. Comparing the dissociation constants measured at these temperatures, both CSPs have a binding affinity in the tens of nanomolar range. This suggests the activity of the proteins as chaperones is optimised relative to the optimum growth temperatures of the organisms.

4.2 Isothermal Titration Calorimetry

ITC measures the emission or absorption of thermal energy upon binding. Using a thermally isolated cell containing the protein, the change in heat upon binding is measured as the ligand is injected into the reaction cell. Equation 4.10 describes the thermal energy from ligand binding in the cell.

$$Q = \Delta H V_0 [PL] \quad (4.10)$$

Q is the total heat content of the solution, ΔH is the molar heat of ligand binding, V_0 is the total volume in the cell and $[PL]$ is the concentration of the protein ligand complex. Assuming a single set of identical sites, $[PL]$ can be replaced with equation 4.9 with an

additional term n , to account for the number of identical binding sites on the protein.

$$Q = \frac{\Delta H V_0}{2} \left[n[P_T] + [L_T] + K_D - \sqrt{(n[P_T] + [L_T] + K_D)^2 - 4n[L_T][P_T]} \right] \quad (4.11)$$

Equation 4.11 now describes the heat content of the cell with the total protein concentration $[P_T]$, the total ligand concentration $[L_T]$ and the dissociation constant K_D . During the experiment, it is the power required to return the cell to equilibrium temperature while the ligand is injected that is measured. The total cell thermal energy for each injection (i) can be calculated using equation 4.11, and combined with the thermal energy from the previous injection ($i-1$) to calculate the change in thermal energy, ΔQ . The calculation for change in thermal energy for each injection (i) is shown in equation 4.12.

$$\Delta Q(i) = Q(i) + \frac{dV_i}{V_0} \left[\frac{Q(i) + Q(i-1)}{2} \right] - Q(i-1) \quad (4.12)$$

V_i is the added volume of the injection and V_0 is the original volume. The injected volume is displacing an equal volume of the original V_0 , and this will account for about half the thermal energy contribution as the original volume.

The energy change measured in the cell during the injection of the ligand is fit to equation 4.12 to calculate n , K_D and ΔH .

$$\Delta G_{bind}^0 = \Delta H_{bind} - T \Delta S_{bind} = RT \ln(K_D) \quad (4.13)$$

ΔG_{bind}^0 is the Gibbs free energy of binding, ΔS_{bind} is the entropy of binding, ΔH_{bind} is the enthalpy of binding, T is the absolute temperature, R is the gas constant, and K_D is the dissociation constant. As the temperature of the ITC experiment is known equation 4.13 can be used to calculate a value for ΔS_{bind} .

4.2.1 *Bs*-CSP

Bs-CSP has been shown previously to have a very high affinity to ssDNA oligonucleotides with a high thymine content^[62]. To investigate the thermodynamic properties (i.e. enthalpy or entropy driven) of this interaction, the reaction of *Bs*-CSP binding to the

oligonucleotide CT2 (CTTTTTC) was measured using ITC. This oligonucleotide was chosen due to its lower K_D (34 ± 4 nM at 15 °C) relative to dT7. The titration of

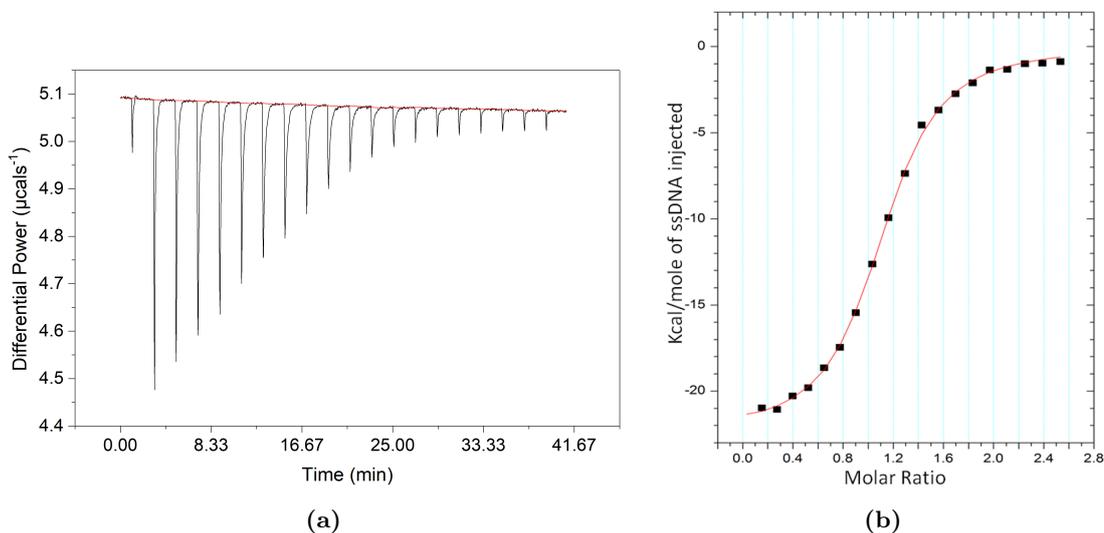


Figure 4.6: Titration of *Bs*-CSP with CT2 ssDNA at 25 °C (a) the power put into the cell to keep it equilibrated during 20 injections, including the initial dud injection to clear the needle. Red line shows the baseline used to adjust for overall temperature drift (b) Energy per mole of injectant plotted against the molar ratio of ssDNA to protein. Each marker (black) is the result of integrating the area of each peak using the baseline. This is fit with the 1 to 1 binding model (red).

Bs-CSP with CT2 ssDNA at 25 °C is shown in figure 4.6a. Each peak corresponds to the energy put into the thermally isolated cell after one injection of ssDNA, to return it to the equilibrium temperature. The peaks in figure 4.6a are clear and distinct from one another, and the resulting isotherm (figure 4.6b) has clear baselines both for the free protein and bound complex. Fitting this binding isotherm with equation 4.11 produces values for the thermodynamic parameters of binding (equation 4.13). The value of enthalpy determined from this fit was $\Delta H = -93.6 \pm 0.7$ kJmol⁻¹. This is within the range expected from previously measured *Bs*-CSP titrations with ssDNA strands that are 23 nucleotides in length^[130], where the value of dC23 (CCCCCCCCCCCCCCCCCCCCCCC) was -25 kJmol⁻¹ and the value for dT23 (TTTTTTTTTTTTTTTTTTTTTTTTTTT) was -110 kJmol⁻¹.

The dissociation constant calculated from the 1 to 1 binding model was (529 ± 25) nM and the entropy value was calculated as -0.192 kJmol⁻¹K⁻¹.

4.2.2 *PB6*-CSP

The *PB6*-CSP has a lower affinity for ssDNA, so the dT7 ssDNA oligonucleotide was used in the titration. The results of this titration at 25 °C are shown in figure 4.7. Parameters derived from the fit to the 1 to 1 binding model in equation 4.11 (figure

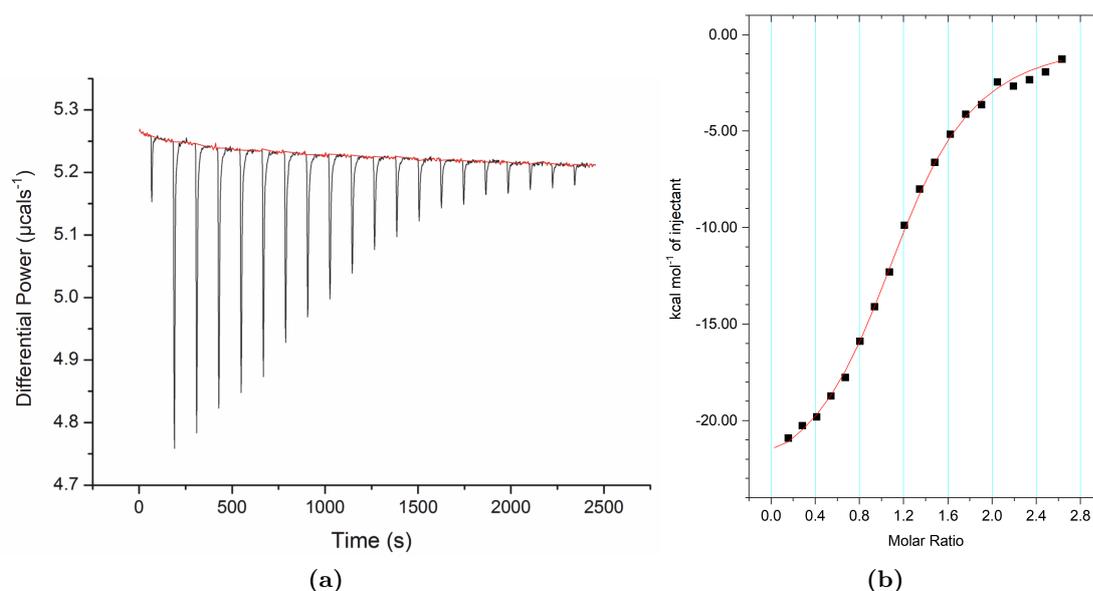


Figure 4.7: Titration of *PB6*-CSP with dT7 ssDNA at 25 °C (a) the power put into the cell to keep it equilibrated during 20 injections, including the initial small injection to clear the needle. Red line shows the baseline used to adjust for overall temperature drift (b) Energy per mole of injectant plotted against the molar ratio of ssDNA to protein. Each marker (black) is the area of a peak integrated using the baseline. This is fit with the 1 to 1 binding model (red).

4.7b) were $\Delta H = (-10.0 \pm 0.1) \text{ kJmol}^{-1}$ and $K_D = (970 \pm 50 \text{ nM})$. The entropy, ΔS was calculated as $-0.219 \text{ kJmol}^{-1}\text{K}^{-1}$. Both the dissociation constant and entropy values are larger for *PB6*-CSP than for the *Bs*-CSP.

The titration was repeated at 10 °C to explore the temperature dependence of the binding thermodynamics. The differential power spectrum seen in figure 4.8a appears to show two peaks associated with each injection. Initially the main peak and then roughly 10 seconds later a smaller peak, with a height that remains more constant than the main peak. The second peak was still seen in a repeat after thorough cleaning of the injection needle. One cause of this feature could be two processes occurring at different rates. As the second peak only appears at lower temperatures, it could be due to a

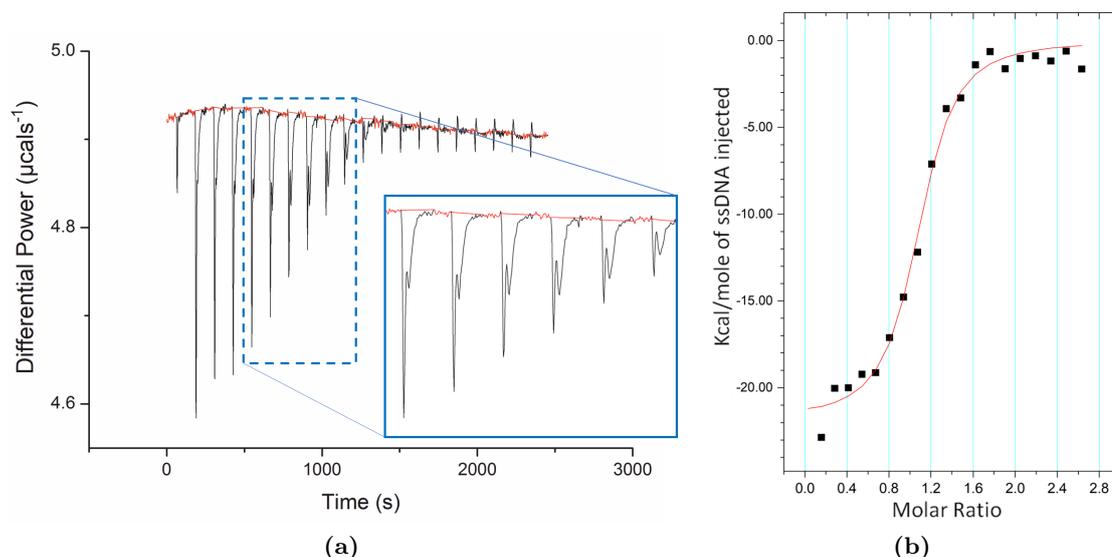


Figure 4.8: Titration of *PB6*-CSP with dT7 ssDNA at 10 °C (a) the power put into the cell to keep it equilibrated during 20 injections, including the initial small injection to clear the needle. Red line shows the baseline used to adjust for overall temperature drift and insert in the blue box shows a larger scale plot of the central region where the double peaks are clear (b) Energy per mole of injectant plotted against the molar ratio of ssDNA to protein. Each marker (black) is the area of a peak integrated using the baseline. This is fit with the 1 to 1 binding model (red).

conformational change in the protein that is present at higher temperatures but too fast to be observable as distinct from the main binding process. At 10 °C the *PB6*-CSP titration with dT7 had a calculated ΔH of -9.1 ± 0.2 kJmol⁻¹ and a K_D value of 230 ± 40 nM. These values show the expected increase in affinity at lower temperatures, but with a slightly decreased enthalpic contribution to the binding. Table 4.3 includes

Temp. (°C)	ssDNA	K_D (nM)	ΔG (kJmol ⁻¹)	ΔH (kJmol ⁻¹)	ΔS (kJmol ⁻¹ K ⁻¹)	$-T\Delta S$ (kJmol ⁻¹)
<i>Bs</i> -CSP 25	CT2	529±25	-35.8 ± 0.3	-93.6 ± 0.7	-0.194	57.8
<i>PB6</i> -CSP 10	dT7	230±40	-35.8 ± 0.8	-91 ± 2	-0.195	55.2
<i>PB6</i> -CSP 25	dT7	970± 3	-34.7 ± 0.3	-100 ± 1	-0.219	65.3

Table 4.3: Parameters obtained from the 1 to 1 binding model fit to ITC measurements of *PB6*-CSP and *Bs*-CSP binding to ssDNA oligonucleotides. From fits shown in figures 4.6b, 4.7b and 4.8b

the ITC parameters calculated for all three systems measured. At 10 °C, the *PB6*-CSP

demonstrates very similar measured thermodynamics the *Bs*-CSP at 25 °C, although with a slightly different oligonucleotide. The enthalpic and entropic energies from these experiments are shown graphically in figure 4.9a. Near the cold shock temperatures for each organism, the contribution to Gibbs free energy from the enthalpic and entropic contributions are similar. Using equation 4.13, the enthalpy, ΔH and entropy, ΔS can be used to predict the dissociation constant over a range of temperatures. This assumes that all these parameters remain constant at the different temperatures, which would assume the conformations of the binding partners remains the same. A plot of this predicted temperature dependence is shown in figure 4.9b. The parameters derived from

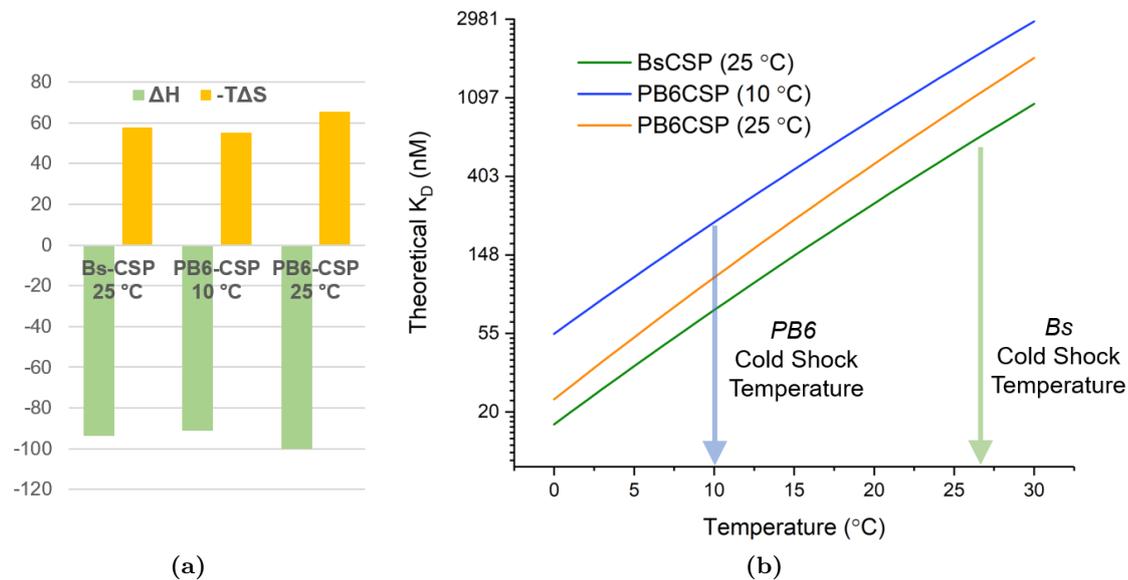


Figure 4.9: (a) The enthalpic (green) and entropic (orange) energy components of ssDNA binding and (b) the temperature dependence of Dissociation constants, K_D , calculated from Gibbs free energy, table 4.3, measured by ITC. The protein and experimental temperature used to derive each set of parameters are the *Bs*-CSP at 25 °C (green), the *PB6*-CSP at 25 °C (orange) and the *PB6*-CSP at 10 °C (blue).

experiments on the *PB6*-CSP at 10 and 25 °C do not describe the same temperature dependence of K_D . If the behaviour seen at 25 °C continued to lower temperatures, the K_D value would be much lower, approaching that of the *Bs*-CSP at low temperatures. Instead what is seen at 10 °C is an affinity of the same magnitude and thermal parameters very similar to those derived for the *Bs*-CSP at 25 °C. This is important as 10 °C is close to the *PB6* cold shock temperature and 25 °C is closer to the *Bs* cold shock temperature

as shown in figure 4.9b. This suggests that a component of the binding is changing to reduce the affinity of the *PB6*-CSP at lower temperatures, possibly a conformational change in the protein. This would correlate with low temperature behaviour measured on this protein with NMR dynamics in chapter 5.

4.3 MicroScale Thermophoresis (MST) to explore Stoichiometry with labelled ssDNA

MST provides a method of tracking a change of signal caused by the free and complexed ssDNA rather than the protein. CT2 (CTTTTTC) ssDNA oligonucleotides labelled with a fluorescent tag at a concentration of 25 nM were used to track the binding of the *Bs*-CSP. The binding of the protein affects how the ssDNA diffuses in response to a temperature gradient. This can be described as thermodiffusion or the Soret effect^[166]. As the concentration of the complex increases, a change in fluorescence signal at a spot heated by an InfraRed (IR) laser is observed, providing a signal to track the complex concentration. The fluorescence signal from the time that the IR laser was switched on to it being switched off is shown in figure B.3. By tracking a signal based on the ssDNA and complex concentration, this experiment provides a different perspective on the binding interaction, tracking the effect of binding on the ssDNA rather than the protein. Figure 4.10 shows a binding isotherm that peaks at a protein concentration of 1000 nM and then drops to less than half of the maximum value. The fluorescence signal at low protein concentrations drops when the IR laser is switched on, indicating the expected behaviour of the labelled ssDNA moving away from the heated spot. This behaviour is typical of biomolecules at these temperatures^[167]. As the protein concentration increases beyond 100 nM (4× the ssDNA concentration) the fluorescence signal starts to increase. This means that as the protein binds to the fluorescently labelled ssDNA, the complex moves towards the heated spot. In the experiment shown here, this effect is maximised at the peak around a protein concentration of 600 nM. This experiment was repeated at 3 different IR laser powers, shown in figure B.4. For each laser strength the peak appears at a different protein concentration, increasing

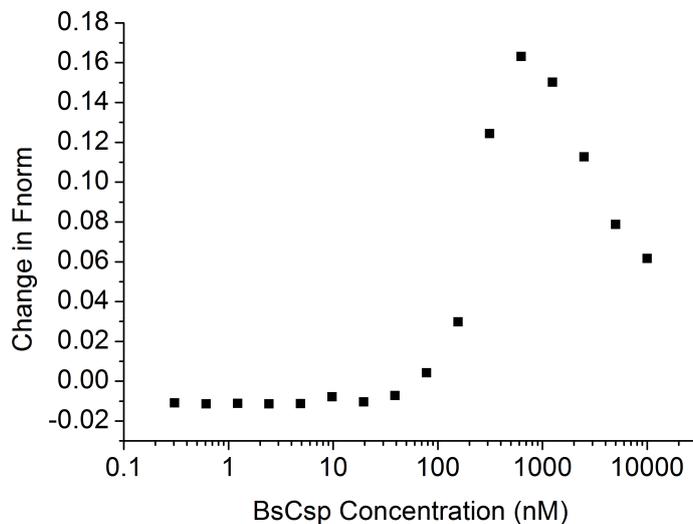


Figure 4.10: Fractional change in fluorescence when the IR laser is switched on plotted against concentration of *Bs*-CSP . The change in fluorescence corresponds to the concentration change in the spot that is heated by the IR laser, at 20 % power. This is decreased when the fluorescent marker moves away from the heated spot and increased when the marker moves towards the heated spot.

with IR power, making it difficult to extract any reliable binding parameters from this experiment.

One cause of this peak could be that the ssDNA oligonucleotides bind at a ratio of 2:1 dT7 to *Bs*-CSP at high ssDNA concentrations, relative to protein concentration. Then when the concentration of protein is increased, binding 1:1. This 2:1 binding would not be observed in tryptophan quenching titrations (section 4.1) or ITC measurements (section 4.2) as those would not differentiate between 2:1 and 1:1 binding until the system is saturated with ssDNA.

The interactions between the *Bs*-CSP and various ssDNA oligonucleotides has been studied previously using a variety of techniques^[44]. The size of the *Bs*-CSP binding site has been determined to be between 6 and 7 nucleotides of ssDNA, based on stoichiometry measurements of longer (23 nucleotides) ssDNA strands^[130]. It was also shown to have a higher affinity for thymine residues^[62]. Max et al. analysed the crystal structure of the complex of *Bs*-CSP with ssDNA oligonucleotide dT6 (TTTTTT), and designed the oligonucleotide TTCTTTT to have the highest affinity with *Bs*-CSP, as shown in the schematic 1.16. The K_D of this sequence with *Bs*-CSP was measured as 0.2 nM at 15

°C. The crystal structure also reveals that it is possible for the dT6 strand of ssDNA to bridge two *Bs*-CSP and vice versa. This implies that both 2:1 and 1:2 protein:ssDNA binding is possible, despite the strand of ssDNA being designed to cover the binding site of one protein.

Based on this ability to bind two ssDNA molecules and a preference for TTCTTTT over dT7, a model to explain this MST result can be conceived. If the protein binds to one half of the dT7 strand, as it does in the crystal structure, the free half of the protein can either bind to the rest of the dT7 strand, or bind to half of another dT7 molecule. Based on the preference for TTCTTTT, a second dT7 molecule might provide a more suitable binding partner as it can bind without needing to fit across the part of the binding surface where a cytosine base is preferable.

4.4 Single Molecule Force Spectroscopy of the ssDNA-*Bs*-CSP Complex Unbinding

The force needed to unbind protein complexes is very small relative to SMFS experiments regularly measured and therefore gets lost in 'noise'. A refined analysis procedure is needed with a distinguishing feature that will identify the unbinding force as separate from non-specific interactions. This feature can be a precise length across binding partners that can be controlled in the experimental design, or a control method that can block this force and cause a measurable difference. If the interaction force can be identified then SMFS can be applied to further understand the system at the scale of individual molecules. This can extract features of the interaction that are otherwise lost in bulk measurements, which rely on population averages. SMFS has been successfully applied to a range of protein-DNA interactions in this way^[93,94]

To measure the force between the *Bs*-CSP and the DNA oligonucleotide dT7, a protein variant (*Cys-Bs*-CSP) was prepared with a cysteine residue positioned to link the protein to one surface with a PEG linker. A dT7 oligonucleotide with a thiol label at the N-terminus was attached to the other surface with a PEG linker. These were used to

prepare both a functionalised surface and a functionalised AFM tip. Performing force extension experiments with this system, the force of unbinding a single molecule of ssDNA from single Cys-*Bs*-CSP was studied.

4.4.1 Cysteine Mutated *Bs*-CSP binding capability

The site chosen for the cysteine mutation in the *Bs*-CSP (Ser 35) was far from the nucleic acid binding site identified in the crystal structure of the complex (figure 2.2)^[62]. This is both to avoid interfering with the active site and to control the pulling direction of the SMFS experiments. To ensure that this substitution does not impede the binding of the *Bs*-CSP with ssDNA a tryptophan fluorescence quenching titration was performed with the ssDNA oligonucleotide dT7 (figure 4.11).

Addition of the dT7 oligonucleotide was found to quench the tryptophan fluorescence

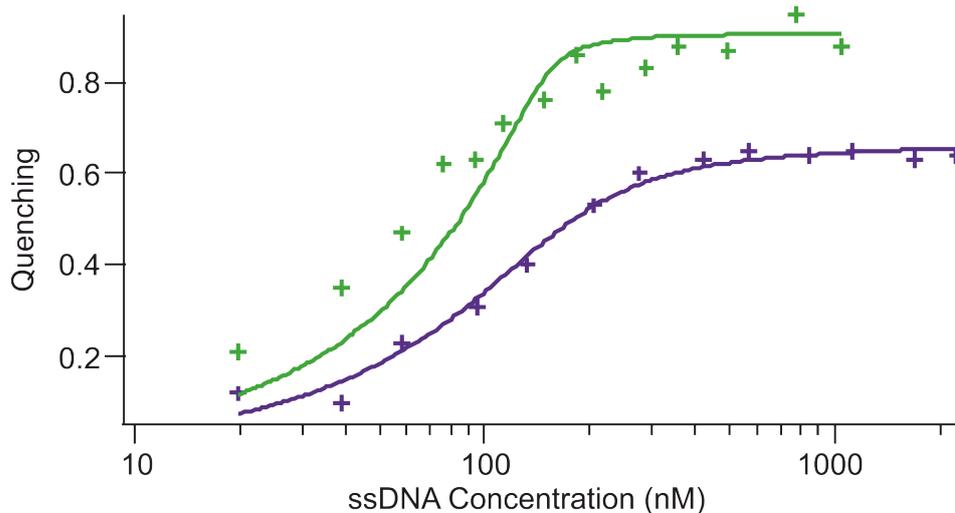


Figure 4.11: Quenching of tryptophan fluorescence of the Cys-*Bs*-CSP as dT7 is titrated into the sample. Experimental data shown by markers for experiments recorded at 15 °C (green) and 25 °C (purple). Lines show the fits to the tight binding equation, with resulting K_D values of 3 ± 16 nM at 15 °C and 42 ± 22 nM at 25 °C, assuming a binding ratio of 2 ssDNA oligonucleotides per *Bs*-CSP at low ssDNA to *Bs*-CSP ratios.

at 15 and 25 °C. The fit to the tight binding equation calculated a stoichiometry of 2:1, dT7:Cys-*Bs*-CSP in these conditions. This titration was performed with a *Bs*-CSP concentration of 300 nM as a less sensitive fluorescence spectrometer was used, meaning the protein was in excess during the initial injections of the titration. As dT7 contains 7

identical nucleotides it can bind in either the 2:1, 1:1 or 1:2 form of dT7:*Bs*-CSP. Unlike the ITC measurements where the signal is dependent on the whole binding site on the protein, so a 1 to 1 ratio would still be recorded for dT7 oligonucleotides which match the size of the binding site, this measurement is based on the interactions of a single amino acid side chain. As seen in the crystal structure, a single ssDNA strand can bind at the tryptophan site of two *Bs*-CSPs. This does not match results recorded of the wild type *Bs*-CSP (section 4.1) but this titration was performed on an instrument with lower sensitivity, therefore requiring a larger protein concentration for a fluorescence emission signal to be recorded. With this fit the K_D values were calculated as 3 ± 16 nM at 15 °C and 42 ± 22 nM at 25 °C. This shows that the *Cys*-*Bs*-CSP has maintained binding affinity to dT7 and can be used to investigate this binding interaction by SMFS.

4.4.2 Traces Measured of Unbinding Force at Controllable Length

SMFS experiments were performed using surfaces functionalised with the *Cys*-*Bs*-CSP and tips functionalised with the ssDNA strand dT7. Results were analysed using software written by Dr Yun Chen, which fits all potential peaks with the Worm Like Chain (WLC) model. The WLC is used to describe polypeptide stretching so peaks that did not fit with this model were excluded as non-specific interactions such as those between the tip and the surface. Traces with more than one peak were also excluded as this implies multiple interactions which can obscure the unbinding force measurement. Example traces included in analysis from these experiments are shown in figure 4.12.

The combined length of the NHS-PEG₂₄-maleimide linkers, *Cys*-*Bs*-CSP and the ssDNA strand thiol-GCTTTTTTTT was estimated to be 25 nm. Traces measured to have around this contour length by the fit with the WLC model are shown in figure 4.12, with the WLC fit shown in black. At this length it is difficult to distinguish between traces that are due to the CSP-ssDNA unbinding, and those that are caused by other surface interactions.

The length of the ssDNA was increased by an additional 50 nucleotides, using the sequence shown in table 2.20. These longer strands were designed to contain a long region

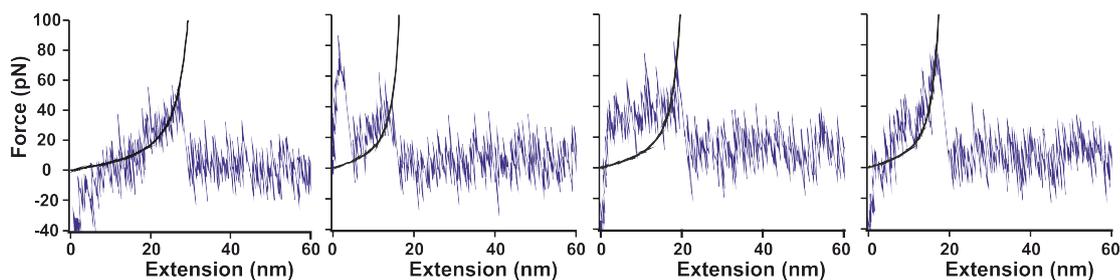


Figure 4.12: Example traces from SMFS experiments of the unbinding of the *Bs*-CSPCSP-dT7 complex carried out at 600nm s^{-1} , with the WLC fit shown in black. Traces with contour lengths (from left to right) of 31, 22, 25, and 20 nm from experiments using an ssDNA strand with the overall expected contour length of 25 nm

with a low association for *Bs*-CSP and low probability predicted of forming secondary structures, predicted using the online tool Mfold^[168]. These regions were added between the thiol linker and the dT7 sequence with the aim of increasing the expected contour length by around 17 nm to a total contour length of 42 nm. Unfortunately the optimisation of this set-up was not successful and no reliable measurements of unfolding force were recorded.

4.4.3 Statistical Analysis of Unbinding Forces

Traces from each experiment were analysed and used to create a probability density map of the peaks in terms of rupture force and contour length. The two density maps shown in figure 4.13 a and b show this probability density from experiments performed with and without the presence of ssDNA dA7 (AAAAAAA), for the shorter predicted contour length (25 nm). The presence of dA7 was used as a control experiment as it is the complementary strand to the dT7 on the functionalised tip. As the *Bs*-CSP does not bind double stranded DNA, the presence of ssDNA should block any measured interactions of the unbinding force.

Histograms plotted of the contour lengths and rupture forces obtained from the WLC fit are shown in figure 4.14. These were produced from the data analysis of one experiment, performed at 600 nm s^{-1} both with and without the presence of dA7, as a control. The histograms were fit with a Gaussian distribution, except for the contour length in the presence of dA7 as this was centred at too short a length. The centre of the Gaussian fit

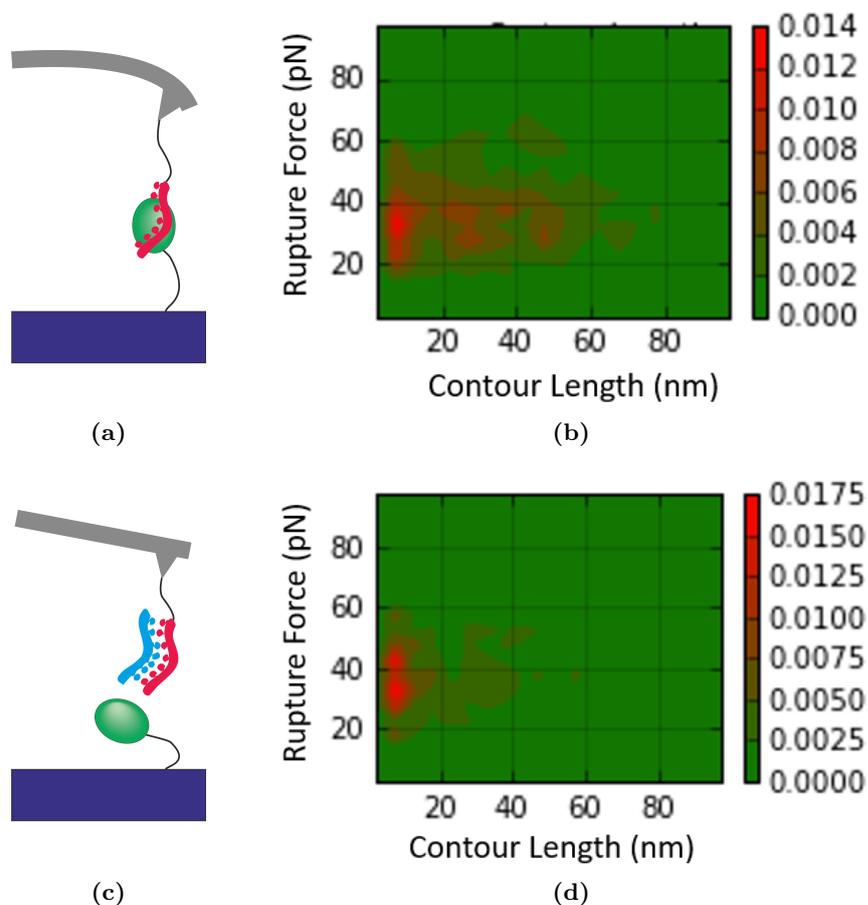


Figure 4.13: SMFS experiments of the unbinding of the *BsCSP*-dT7 complex carried out at 600nm s^{-1} . (a) Diagram of the unbinding force measurement (b) Peaks identified in FX traces represented as a probability density of unbinding events by force measured and contour length obtained from the WLC fit. The red-green scale indicates the probability density measured from 4000 traces. The predicted contour length of the structure is 25 nm and some rupture peaks are recorded at this length with forces between 20 and 50 pN that appear to be less likely in the control experiment (c) Diagram of the control experiment, where the complementary strand of ssDNA (dA7) was added to the solution to block the interaction (d) Probability density measured in the control experiment in the presence of dA7.

of the contour length was 21 ± 1 nm, close to the expected value of 25 nm. The forces calculated were 38 ± 1 without dA7 and 38 ± 1 with dA7.

In both the experiment and control, the highest proportion of measured contour length is below 10 nm. This is most likely due to the surface interaction. A number of peaks of the expected contour length, around 25 nm, were measured in the non control experiment (figure 4.13 a). This number was reduced after the dA7 was added (4.13 b)

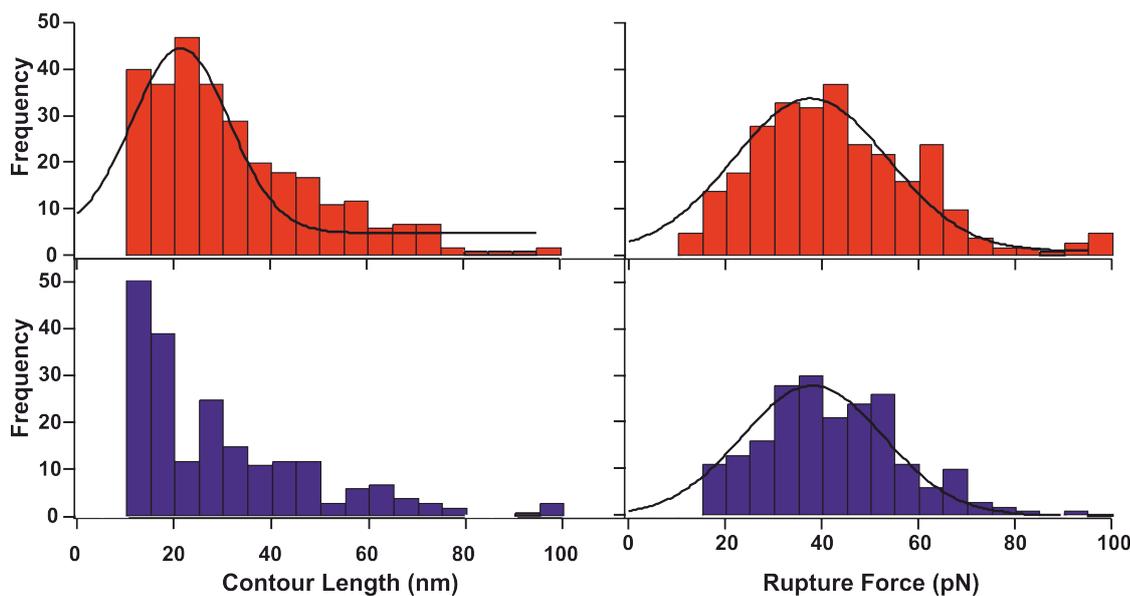


Figure 4.14: Histograms of the WLC fit contour lengths (left) and unbinding forces (right) measured in one SMFS experiment of the unbinding of the *Bs*CSP-dT7 complex carried out at 600nm s^{-1} (red) and then repeated in the presence of the ssDNA strand dA7 (blue) as a control. Contour lengths below 10 nm were excluded and a Gaussian model (black) was used to fit the data. The centre of the Gaussian fit of the contour length was 21 ± 1 nm but the the data collected in the presence of dA7 did not fit the Gaussian model. The forces calculated were 38 ± 1 without dA7 and 38 ± 1 with dA7.

which may be an indication that the unbinding of dT7 ssDNA from Cys-*Bs*-CSP is being measured. However, it is difficult to differentiate on an individual level which peaks are unbinding peaks and which are other, non-specific interactions. The data measured in the presence of dA7 demonstrates the presence of non-specific interactions in the measurements of force, and as the force calculated in the control (figure 4.14) matches the force measured in the main experiment it is clear that the non-specific interactions will obscure any complex interaction data.

There is potential for this system to be improved by finding an appropriate surface concentration of protein and ssDNA. If the concentrations are too high then single interactions are obscured, but too low and not enough data will be recorded. It could also be improved by the use of optical tweezers in place of the AFM. These can accurately determine forces of a smaller scale than the AFM, which is likely the prime cause of difficulty in these measurements^[169].

4.5 Discussion

The function of the CSP is known to include the role of an RNA chaperone when systems experience “cold shock”. To perform this role, the CSP must bind nucleic acids with an affinity strong enough to bind and influence the RNA structure without hindering the RNA function.

4.5.1 Affinities Adapted to Organism Native Temperature

Tryptophan fluorescence quenching titrations with dT7 ssDNA oligonucleotides performed on the mesophilic *Bs*-CSP and psychrotropic *PB6*-CSP, showed *PB6*-CSP had a lower affinity for this sequence. The affinity to dT7 ssDNA was highly temperature dependent in both proteins. At temperatures considered “cold shock” (8-10 °C below the organism optimum temperature) the measured affinity of both proteins to dT7 ssDNA was in the tens of nanomolar range. This is evidence of the protein function being adapted to function relative to the organism native temperature^[170].

Exploring the driving forces behind this temperature dependent affinity ITC was performed with both proteins. At 25 °C the *Bs*-CSP had a stronger affinity to the ssDNA oligonucleotide, at (529 ± 25) nM, than the *PB6*-CSP, at (970 ± 53) nM. The thermodynamic parameters measured for the *Bs*-CSP matched values predicted for the binding of CT2 from results measured previously on dC23 and dT23 oligonucleotides^[130]. *PB6*-CSP showed higher enthalpic and entropic energy differences on dT7 binding. This difference may also have been due to the difference in oligonucleotides used for these experiments. At the lower temperature measurement of the *PB6*-CSP these thermodynamic parameters were much closer to the *Bs*-CSP values at 25 °C, as shown in table 4.3.

The binding affinity at 10 °C of the *PB6*-CSP binding dT7 does not follow the trend predicted from thermodynamic parameters measured at 25 °C by ITC. The *PB6*-CSP

appears to dampen the low temperature binding affinity by adjusting the thermodynamic properties of binding between 25 and 10 °C. This could be related to the extra peak seen in the injection peaks of the ITC titration. This peak seems to show a slow change occurring in the presence of ssDNA. The slow process could be a conformational change that is too fast to observe at higher temperatures, but aids low temperature function. This possibility is explored in the NMR measurements on *PB6*-CSP low temperature dynamics in chapter 5. One consequence of this result is that low temperature behaviour of proteins cannot be predicted purely from measurements taken at higher temperatures.

Dissociation constants measured by ITC are orders of magnitude larger than those measured using tryptophan fluorescence quenching. This is partially due to the high protein concentration required to produce a measurable signal in the ITC experiment. As the protein concentration used was 10 μM , K_D values below 10 % of this value become less distinct. As all of the K_D values measured by tryptophan fluorescence quenching below 25 °C, are below 1 μM , these will not be precisely determined by ITC.

4.5.2 Stoichiometry of ssDNA Oligonucleotides binding to the CSPs

A concentration dependent stoichiometry was observed by MST, which tracked the change in ssDNA behaviour in response to increasing protein concentrations. This showed that at high protein:ssDNA ratios, the stoichiometry appeared to shift to a lower number of ssDNA oligonucleotides per protein. This suggests that initially two ssDNA oligonucleotides were bound per protein. This could explain the effects seen in tryptophan fluorescence quenching titrations performed with protein concentrations much higher than the K_D , such as the titration shown in figure 4.11. Titrations performed with these concentrations showed evidence of 2:1 binding at low ssDNA concentrations, while still showing 1:1 binding properties when concentrations of the protein and DNA were more equal.

Based on the complex structures measured in Max 2006^[62] and shown in figure 1.16,

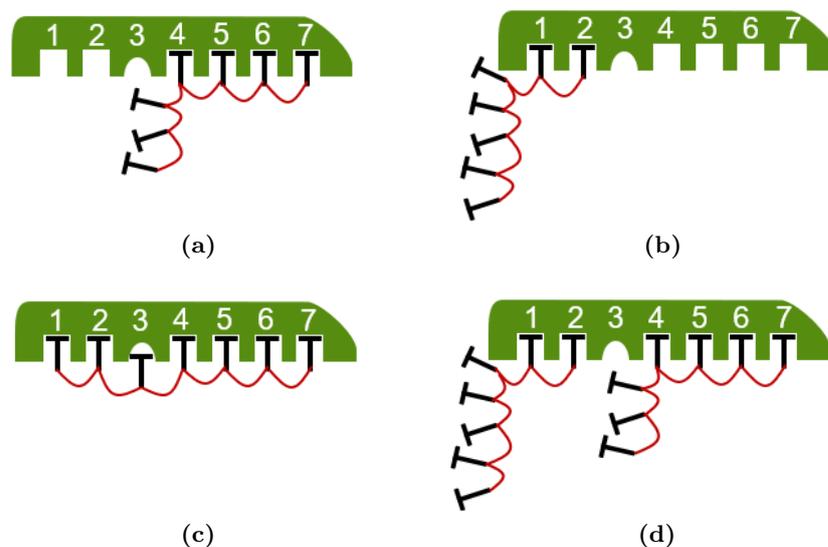


Figure 4.15: Binding Models based on the crystal and solution structures identified in Max 2006^[62]. (a) and (b) show partial binding of dT7 with the *Bs*-CSP with the two halves bound as seen in the crystal structure. Site 3 of the protein corresponds to the position that preferentially binds to a cytosine. (c) and (d) show possible binding positions that saturate the *Bs*-CSP binding surface, with (d) demonstrating 2:1 binding of dT7:*Bs*-CSP.

four possible binding conformations are shown in figure 4.15. Of these the complex measured in solution, figure 4.15c, should be the most stable as it has the maximum number of bonds between the two molecules. However, due to the identical bases of the dT7 strand partial binding of the *Bs*-CSP surface is also possible. From the complex models identified in figures 4.15a and 4.15b, the preference for a cytosine at site 3 might make the affinity for a second ssDNA molecule higher than the affinity for the already bound molecule, as shown in figure 4.15d. If there is a high enough concentration of ssDNA in the solution, the binding model shown in figure 4.15d might be more likely than that shown in figure 4.15c and this would explain the behaviour seen in the MST experiment.

A concentration dependent stoichiometry would account for these results, and could also be the cause of the reduction in quenching seen at high ssDNA concentrations in the tryptophan fluorescence quenching of the *PB6*-CSP seen in figure 4.4. If the stoichiometry is changing at these high ssDNA concentrations, the fluorescence signal of the complex could be affected. Potentially, a complex of 2:1 ssDNA to protein could expose the tryptophan residue, or bind in a manner that does not quench the

fluorescence as efficiently.

4.5.3 SMFS Measurements of ssDNA Unbinding from the *Bs*-CSP

Initial experiments were performed to measure the unbinding force of a single molecule interaction between *Cys-Bs*-CSP and dT7 ssDNA. Evidence was recorded of an interaction that is disrupted by the addition of dA7, shown in figure 4.13, the complementary strand to dT7 that should block the protein-ssDNA binding. The high flexibility of the ssDNA in this interaction is one obstacle in obtaining a clear, disruption force profile. One potential improvement in the experimental design would be the use of DNA bound to the surface at both termini with the binding site positioned halfway along the length. The tethering of both termini would position the binding site of the DNA in a more uniform orientation when the complex is stretched before unbinding. This could help counter the flexibility of ssDNA, which results in a wide range of orientations when the complex is pulled apart, and the potential of protein “sliding” along the ssDNA^[171] before unbinding, leading to weaker measured forces. The experiment could also be moved to an optical trap system, as this has been shown to be capable of accurately measuring the low forces of protein DNA interactions^[73].

5

Protein Dynamics measured by NMR

Folded proteins will have dynamic properties determined by the characteristics of the amino acid side chains, the sequence of amino-acids and the way they interact with each other. These properties will not be uniform across the whole protein structure and while many experimental techniques measure average properties of the overall protein, protein NMR is useful because it can measure structural and dynamical characteristics of individual residues.

In this chapter the NMR signals from Cold shock protein from the psychrotropic *Psychrobacter 6* (*PB6*-CSP) and from *Bacillus Subtilis* (*Bs*-CSP) are investigated at different temperatures. As *Bs*-CSP has been studied previously by NMR^[58,59], it is used here as a mesophilic control for *PB6*-CSP so that the temperature dependence of the dynamics of each protein can be compared. The structure of *Bs*-CSP was first measured by X-ray crystallography^[53] and this is shown in figure 5.1a. The side chains associated with nucleic acid binding are shown in pink, these were identified from the structure of the protein bound to a strand of single stranded ssDNA, with sequence TTTTTT (dT6)^[62].

There is no published structure for the *PB6*-CSP, so a 3D structure predicted from the amino acid sequence was obtained using the I-Tasser server^[172–174]. This takes the sequence of the protein and compares it with sequences of protein structures in the Pro-

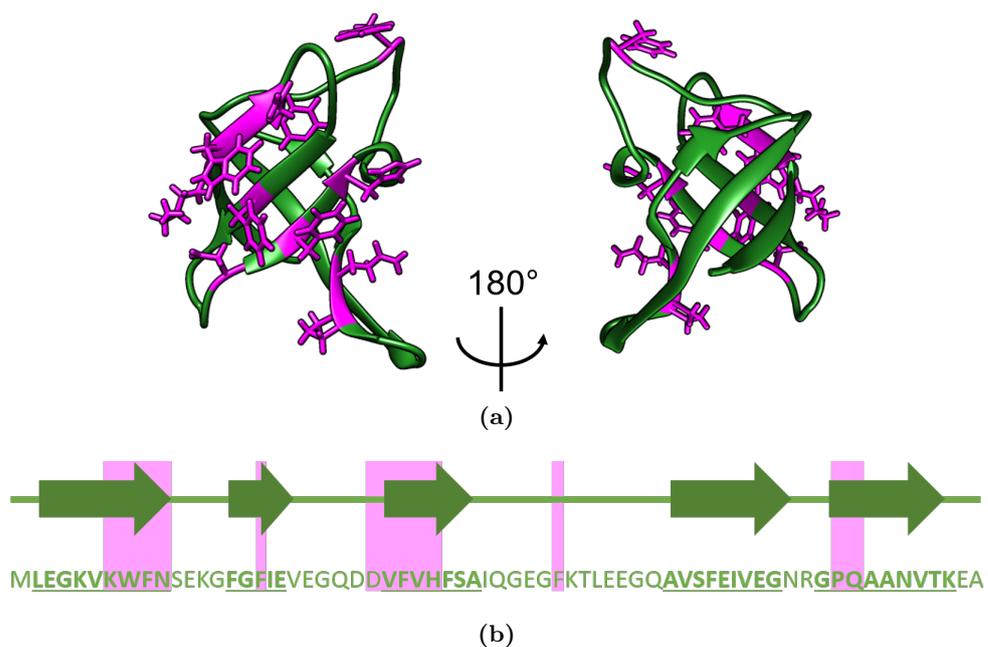


Figure 5.1: (a) Crystal structure of the *Bs*-CSP, pdb code 1CSP^[53], showing side chains of residues that interact directly with dT6 ssDNA in the complex structure (pink), pdb code 2ES2^[62], (b) Sequence of the *Bs*-CSP with structure and ssDNA binding residues identified (pink)

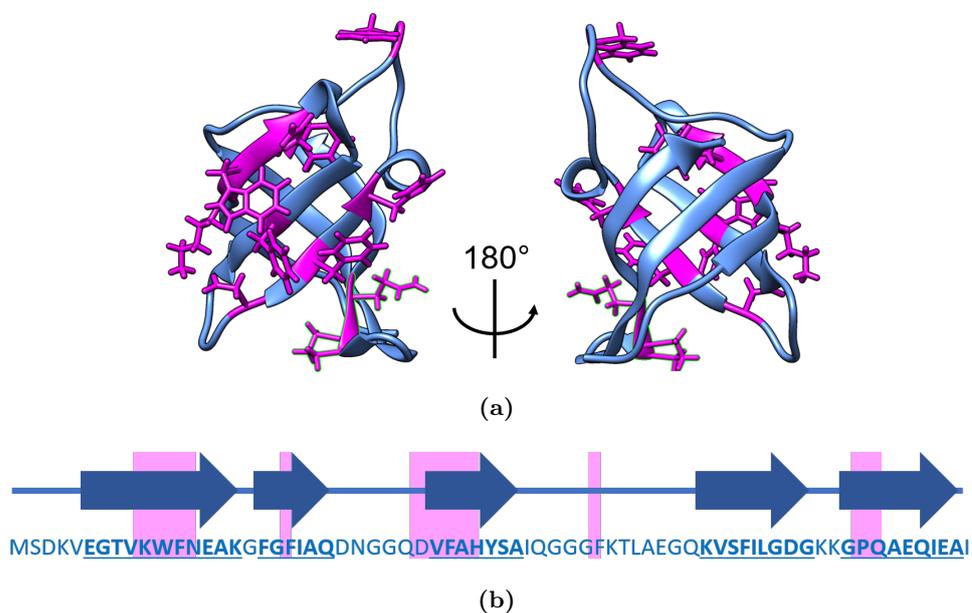


Figure 5.2: (a) Predicted structure of the *PB6*-CSP using the I-Tasser server^[172], showing side chains of residues conserved from the *Bs*-CSP and associated with ssDNA binding highlighted in pink (b) Sequence of the *PB6*-CSP with structure and ssDNA binding residues identified (pink)

tein Data Bank (PDB)^[10] to generate a structure based on similarity. The structure predicted using this method is shown in figure 5.2a, showing the recognisable CSP β -barrel backbone structure. The nucleic acid binding residues of the *Bs*-CSP are easily identifiable in the *PB6*-CSP sequence, and are consistent with the nucleic acid binding surface conserved across cold shock domains^[53,57]. The position of these side chains are shown in figure 5.2a, however these should be considered uncertain as predicting the side chain positions from the protein sequence is generally unreliable.

These structures (figures 5.1 and 5.2), mean that localised dynamics can be understood in relation to the functional regions of the protein. To adapt protein function to the cold shock temperature of each organism, around 29 °C for *Bacillus Subtilis* and around 10 °C for *Psychrobacter* 6, it is supposed that the active regions must maintain flexibility. By mapping the localised dynamics at the functional temperature of each protein, this theory can be tested. Differences between the two proteins can then be explored to identify how the temperature dependent function has been adapted.

PB6-CSP has previously been measured by NMR using a N-terminal histidine tagged variant (Michael Wilson, PhD thesis, University of Leeds). As it is possible that a histidine tag, comprising 6 positively charged histidine residues, will influence the dynamics of this small protein, a histidine tag free *PB6*-CSP construct was made. Design of this construct is described in methods section 2.2.5. For this work the growth and purification process of a histidine-tag free protein was then optimised. The polydispersity of this purified sample was measured using SEC-MALS, ensuring that the protein was monomeric in the NMR samples. The thermostability was then assessed by Circular Dichroism spectroscopy to ensure the removal of the histidine tag did not dramatically alter the properties of the protein. ¹⁵N and ¹³C labelled samples were then used to assign the backbone residues. These assignments were used to analyse the backbone dynamics recorded at different temperatures. Low temperature dynamics of the *PB6*-CSP were explored further using dispersion relaxation measurements, which were then used to calculate properties of the underlying chemical exchange process.

5.1 Sample Characterisation

^{15}N labelled samples of *Bs*-CSP and *PB6*-CSP, and ^{15}N and ^{13}C labelled samples of *PB6*-CSP were prepared as described in section 2.3.

5.1.1 Monitoring Polydispersity of *PB6*-CSP by SEC–MALS

The potential for aggregation at the high protein concentrations required for NMR was investigated using SEC–MALS. Two 50 μL sample of 5 mgml^{-1} or 675 μM *PB6*-CSP (NMR concentration is 1.5 mgml^{-1} or 200 μM) were injected onto the size exclusion column equilibrated to conditions matching those used in NMR measurements at room temperature with 20 mM NaP, 150 mM at pH 7.4 buffer. The results are shown in figure 5.3, where both runs are overlaid. As the light scattering intensity is directly proportional to the ratio of molecular mass to concentration, and the concentration can be measured by UV absorbance, the molecular mass can be calculated over time. Taking a time averaged value, the calculated average molecular weight over both runs was 7 ± 2 kDa, error given is standard deviation. This fits with the expected value of 7.4 kDa. The protein was too small to get a precise measure of the radius of gyration, but the results give a strong indication that the protein is monomeric under these conditions.

From the results shown in figures 2.18a and 5.3, the *PB6*-CSP samples prepared for NMR measurements were pure and monomeric.

5.1.2 *PB6*-CSP Thermo-Stability measurements using CD

Circular Dichroism was used to measure the thermal stability of histidine-tag free *PB6*-CSP in similar buffer conditions as those used in the NMR experiments. As HEPES gives a strong signal in the far UV range it could not be used as a buffer in these experiments, so 20 mM NaPh, 150 mM at pH 7.4 was used as a substitute. The results of the temperature ramp from 7 to 80 $^{\circ}\text{C}$ are shown in figure 5.4.

Figure 5.4a shows the far-UV ellipticity readings during a thermal ramp at six commonly

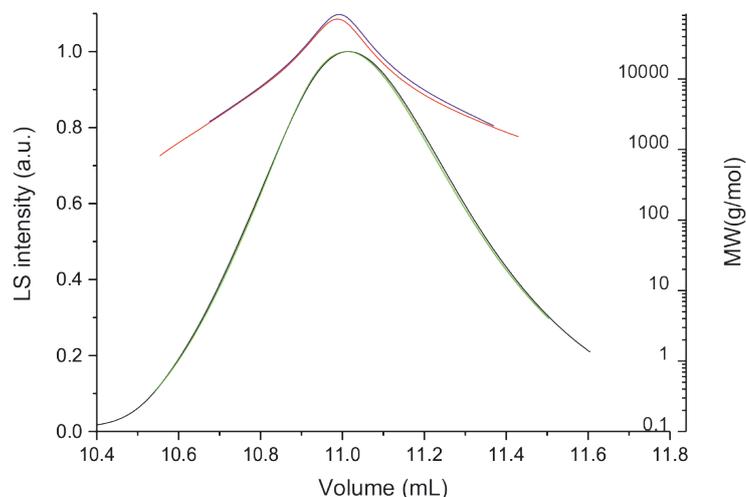


Figure 5.3: SEC–MALS results: Light scattering intensity (green and black) and calculated molecular mass (blue and red) for two runs of *PB6*-CSP are overlaid. Molecular mass is calculated from the light scattering intensity and the protein concentration measured by the UV absorbance.

used wavelengths for CD of proteins, with fits to equation 2.2 shown. From these fits, the baselines of pre-transition and post transition were used to normalise the measured signal to fraction of folded protein, shown in figure 5.4b. The average melting temperature (T_m) from these fits was $(41.6 \pm 0.4) ^\circ\text{C}$, which is only slightly lower than previous measurements performed on histidine tagged *PB6*-CSP by Michael Wilson, $(43.2 \pm 0.6) ^\circ\text{C}$.

5.2 NMR Spectra

5.2.1 HSQC

For most proteins, the 1D NMR spectrum of ^1H is too crowded to resolve for individual amino acids. Heteronuclear single quantum coherence (HSQC) spectroscopy is a technique that results in a 2D spectrum with ^1H chemical shift as one axis and the chemical shift of a covalently attached heteronucleus as the second axis^[175], as described in section 1.5.2.3. As each residue in the protein backbone (except proline) contains

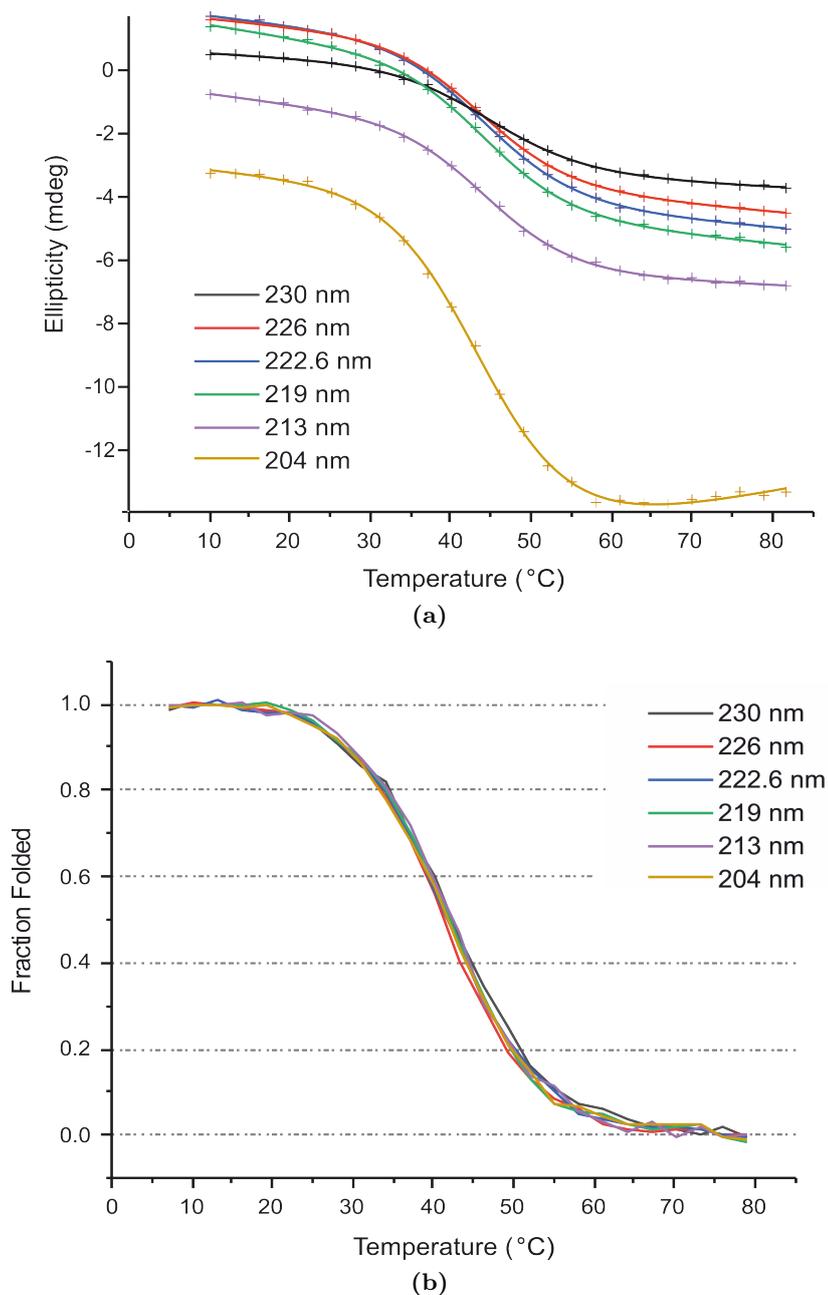


Figure 5.4: Temperature ramp of the *PB6*-CSP measured by far UV-CD spectroscopy at six wavelengths. (a) CD values with fits to equilibrium unfolding equation, (b) Fraction unfolded for each wavelength calculated from the fit.

one nitrogen covalently bonded to hydrogen, ^{15}N labelled protein samples provide a convenient extra dimension for protein spectroscopy. In the 2D ^{15}N - ^1H HSQC, each residue is represented by an individual peak reflecting the chemical environment of the ^1H - ^{15}N nuclei, identified in the protein backbone in figure 1.26a. The spectra generated

with this technique will have a single peak for each N-H bond if the protein structure is fairly rigid. The chemical environment of each of these nuclei will determine where in the spectrum these peaks appear, and the dynamics of the N-H bond will influence the shape of the signal. This is how HSQC spectra provide qualitative information about the protein state.

5.2.2 Temperature Dependent HSQC spectra

The HSQC spectrum of *PB6*-CSP was recorded at 5, 10, 20 and 37 °C as shown in figure 5.5. This temperature variation causes a change in chemical shift for most of the peaks, this is typical behaviour due to the change thermal motion and hydrogen bond strength^[176,177]. Interestingly other changes are observed: some changes in peak intensity, with most residues generating the largest signal at 20 °C, also the appearance of peaks only detectable at low temperatures.

The protein has 70 amino acid residues so 68 peaks, excluding the one proline and the N-terminal residue, should be visible if they do not overlap with one another. This excludes side-chain peaks which appear in the region of the spectrum between 110-115 ppm in the ¹⁵N dimension and 6-8 ppm in the ¹H dimension. At 20 °C, around 75 peaks were counted. As the temperature is lowered, more peaks appear around the centre of the spectrum, as shown in figure 5.6. Extra peaks also appear at 37 °C but this is quite close to the measured melting temperature of 42 °C. The extra peaks at low temperatures could be due to cold denaturation, but the protein structure, as measured by CD and shown in figure 5.4, is calculated to be over 99 % folded at 10 °C according to the two state model.

The appearance of additional peaks could arise for several reasons. Firstly it could be due to the the population of unfolded protein states that are more solvent exposed, so that the ¹H is exchanging with the deuterium in the solvent. As this effect is reduced at lower temperatures, the peaks become visible. Secondly, it could be due to the protein being in slow exchange with states corresponding to these peaks at low temperatures. This would correspond to a more frustrated protein structure where

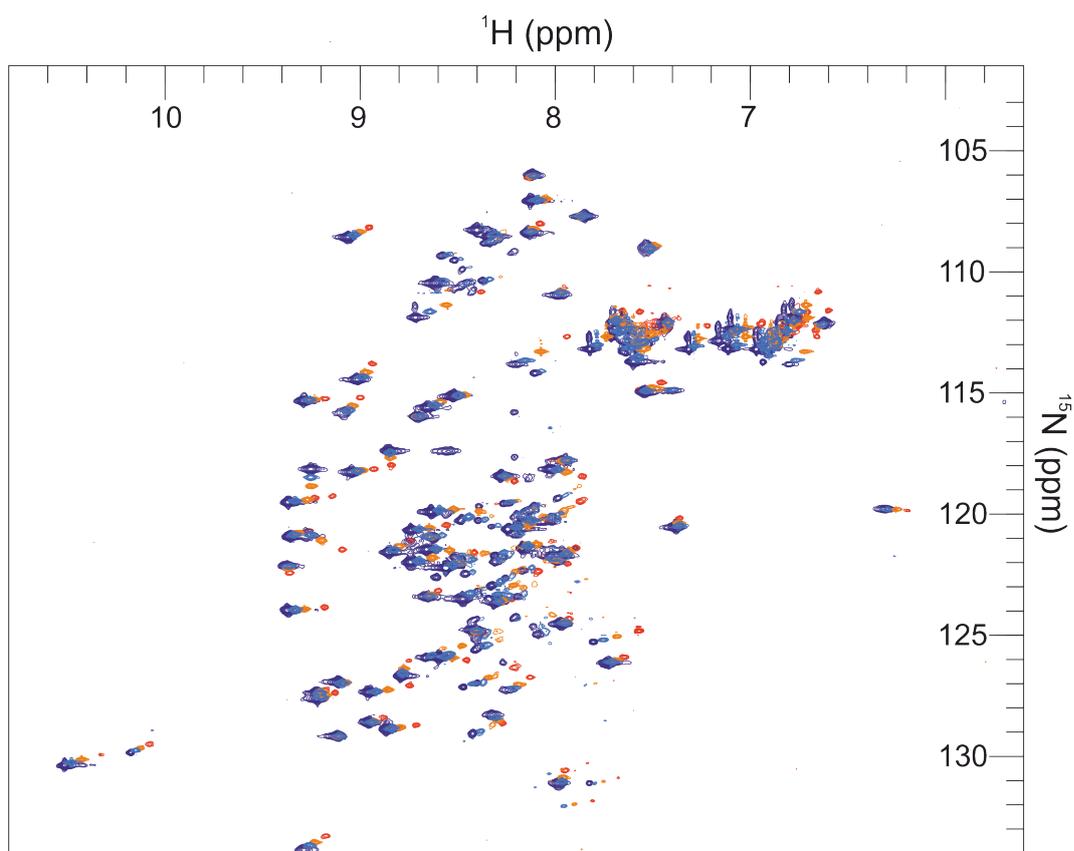


Figure 5.5: HSQC spectra of the *PB6*-CSP at 5 °C (navy), 10 °C (blue), 20 °C (orange) and 37 °C (red)

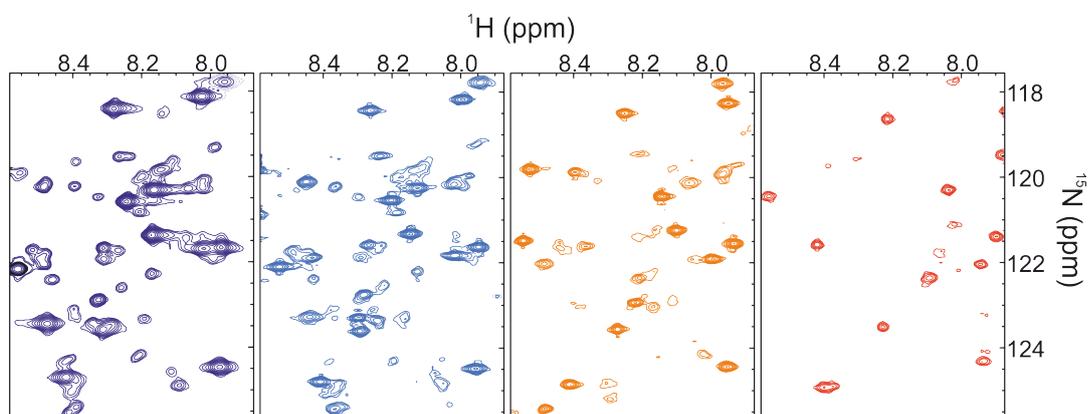


Figure 5.6: Central region of the HSQC spectra of the *PB6*-CSP at 5 °C (navy) 10 °C (blue), 20 °C (orange) and 37 °C (red)

low temperatures lead to slow exchange between shallow, minimum energy wells. As 10 °C is the temperature at which this protein should be functional *in vivo*, this extra state could be useful for maintaining protein function. To explore this possibility,

identification of the source of these extra peaks is essential.

The HSQC spectra of the *Bs*-CSP are shown in figure 5.7. Here little difference is seen

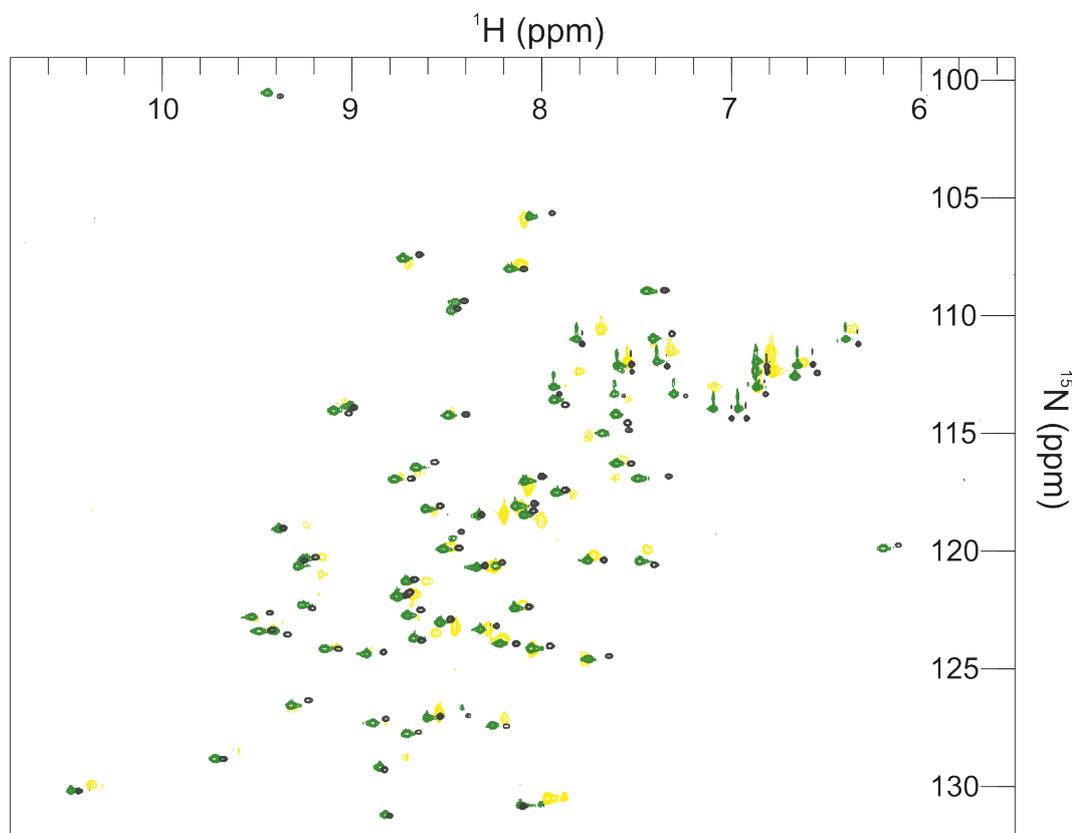


Figure 5.7: HSQC spectra of the *Bs*-CSP at 10 °C (black), 20 °C (green) and 37 °C (yellow)

between 10 and 20 °C. At 37 °C more peaks are no longer present and some peaks are broadened, indicating an increase in μs -ms dynamics at this temperature. This is shown more clearly in figure 5.8 where the central region of the spectrum is shown for the *Bs*-CSP spectra. The spectra at 10 and 20 °C have almost the same number of peaks, with little evidence of any alternative states.

The difference between the low temperature (10 °C) HSQC spectrum of the *PB6*-CSP and *Bs*-CSP suggests that the protein from the cold adapted organism, *Psychrobacter 6*, is more dynamic in this environment. The optimum temperatures of *Psychrobacter 6* and *Bacillus Subtilis* are around 20 °C and 37 °C respectively. If the cold shock temperature for these organisms is approximated as ~ 8 °C below the optimum temperature then the

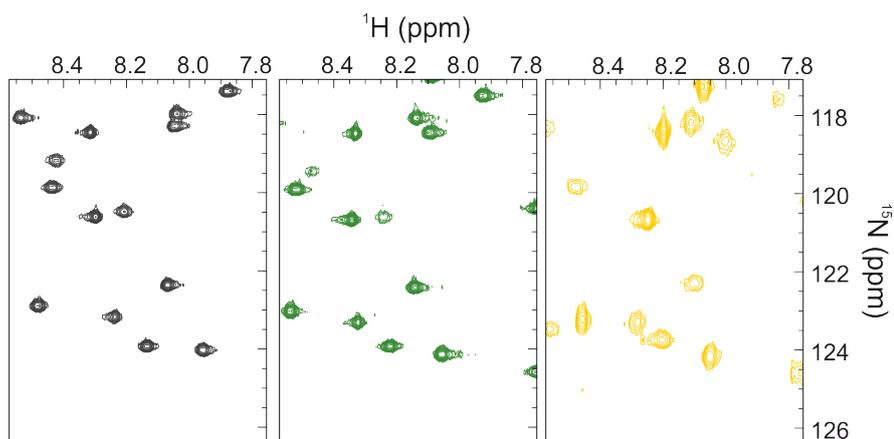


Figure 5.8: Central region of the HSQC spectra of the *Bs*-CSP at 10 °C (black), 20 °C (green) and 37 °C (yellow)

active temperature for the *PB6*-CSP is around 12 °C and for the *Bs*-CSP it is around 29 °C. The extra peaks seen in the *PB6*-CSP spectrum around this temperature are evidence of low temperature dynamics that may help this protein function in cold environments. To explore this possibility further NMR experiments were performed to clarify what characteristic is causing these extra peaks, which parts of the protein are involved, what other dynamic effects can be measured and potentially how this could influence the RNA chaperone activity at low temperatures.

5.3 Protein Backbone Assignment using 3D Triple-Resonance Experiments

While some idea of what residue corresponds to which peak can be gleaned from an HSQC spectrum based on typical amino acid characteristics, such as the lower chemical shifts of glycine in the nitrogen dimension, more information is required to label all the peaks^[123]. For this, triple resonance experiments on a sample labelled with ¹⁵N and ¹³C, is required.

The *Bs*-CSP has been assigned in solution at 25 °C in the presence and absence of bound ssDNA^[59]. This spectrum was overlaid onto the spectrum previously measured at 20 °C^[59] for the *Bs*-CSP which matched well for most of the residue assignments, as shown

in figure 5.9.

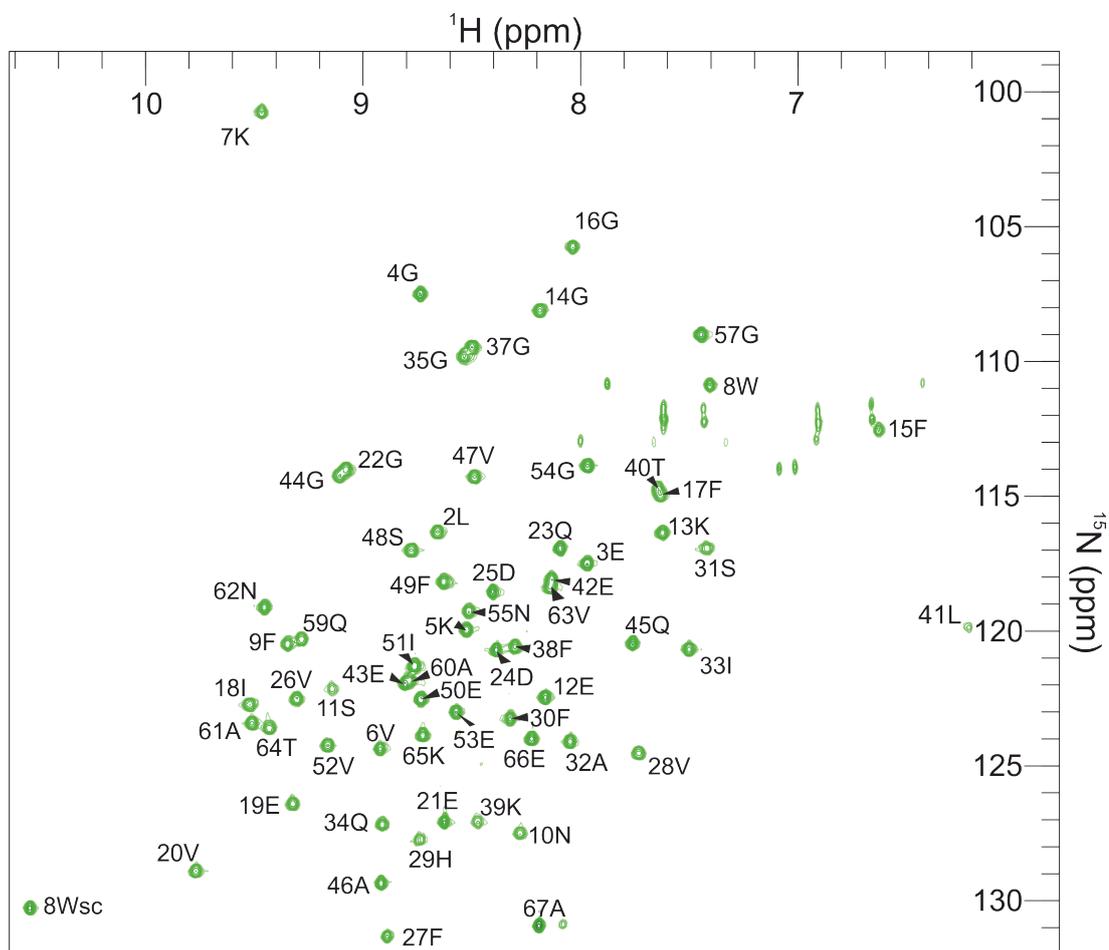


Figure 5.9: HSQC of the *Bs*-CSP with assignments, measured at 20 °C

5.3.1 Assignment of *PB6*-CSP

PB6-CSP has not been assigned previously, so signals from the backbone residues needed to be identified by the methods described in section 2.6.2. As an example of this process, figure 5.10 shows the sequential identification of residues 35-38 using HNCA, HN(CO)CA and HN(CA)CB spectra, as described in section 1.5.2.4.

63 out of the 70 residues of the *PB6*-CSP were assigned and these assignments are shown in figure 5.11. One of these unassigned residues is a proline, residue 62, which does not

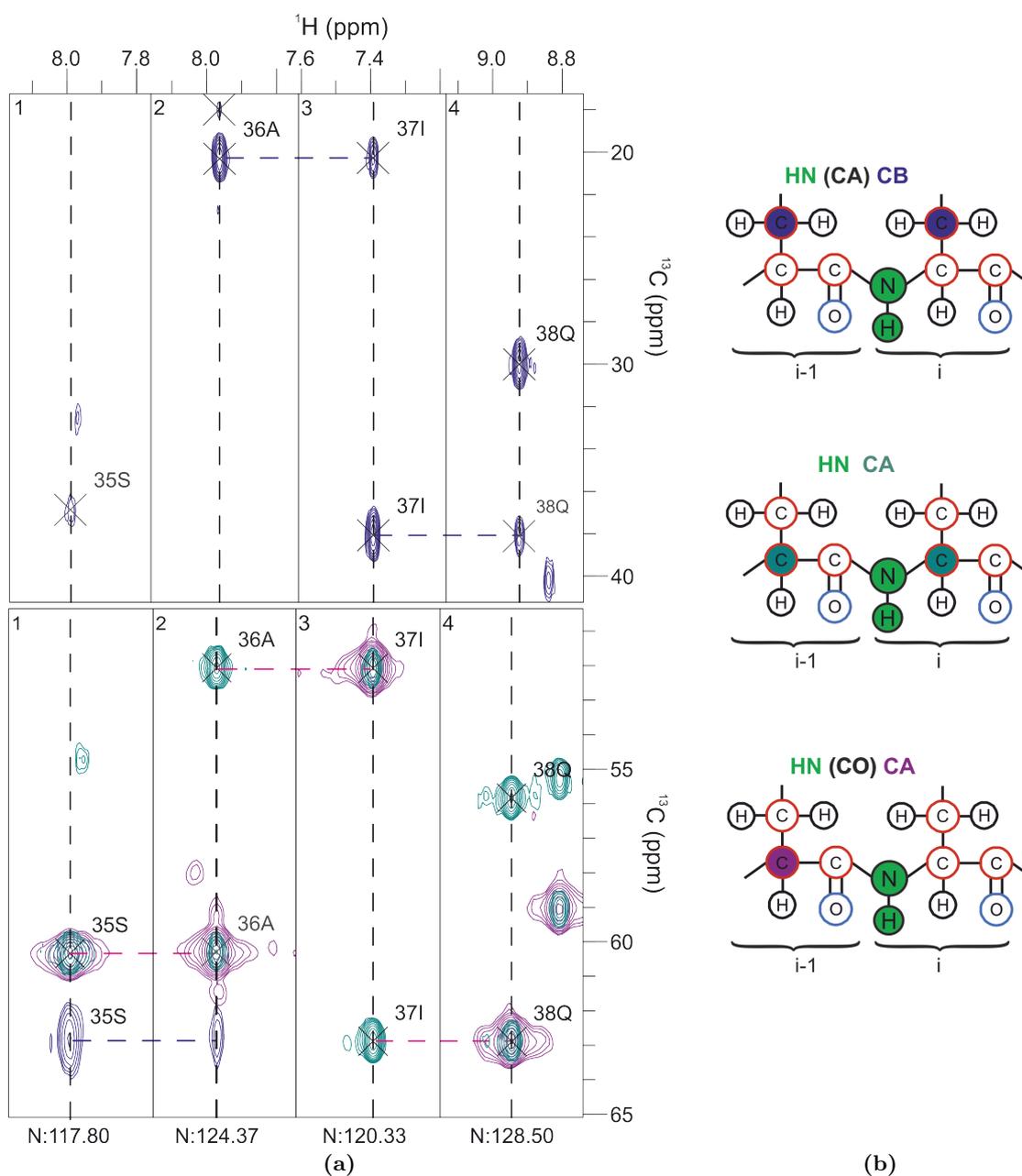


Figure 5.10: (a) HNCA (teal), HN(CO)CA (purple) and HN(CA)CB (dark blue) spectra corresponding to 4 neighbouring residues. In panel 1 the location of the HN(CA)CB signal identifies this residue as a serine. (b) Peptide diagram with colours matching the spectra produced to the corresponding atoms with the initial H-N residue (i) and previous residue (i-1) identified. Initial amide bond highlighted in green and magnetisation is passed from there through HNCA to the $\text{C}\alpha$ (teal), HN(CO)CA to the $\text{C}\alpha$ of the i-1 residue (purple) and HN(CA)CB to the $\text{C}\beta$ atom of the side chain (dark blue)

have a signal in ^1H - ^{15}N HSQC measurements. The N-terminus residue does not appear in these measurements due to the free amine group. The remaining 5 unassigned residues

(Ser2, Gly41, Gly42, Lys59 and Lys 60) make up the N and C terminus of the protein and a glycine rich region. This positioning makes these regions more likely to be flexible and hence more prone to dynamic behaviour. Intermediate time-scale dynamics in these regions would obscure the signal from these residues, as described in figure 1.32.

Unidentified peaks are visible in the centre of the HSQC spectrum (figure 5.11). These

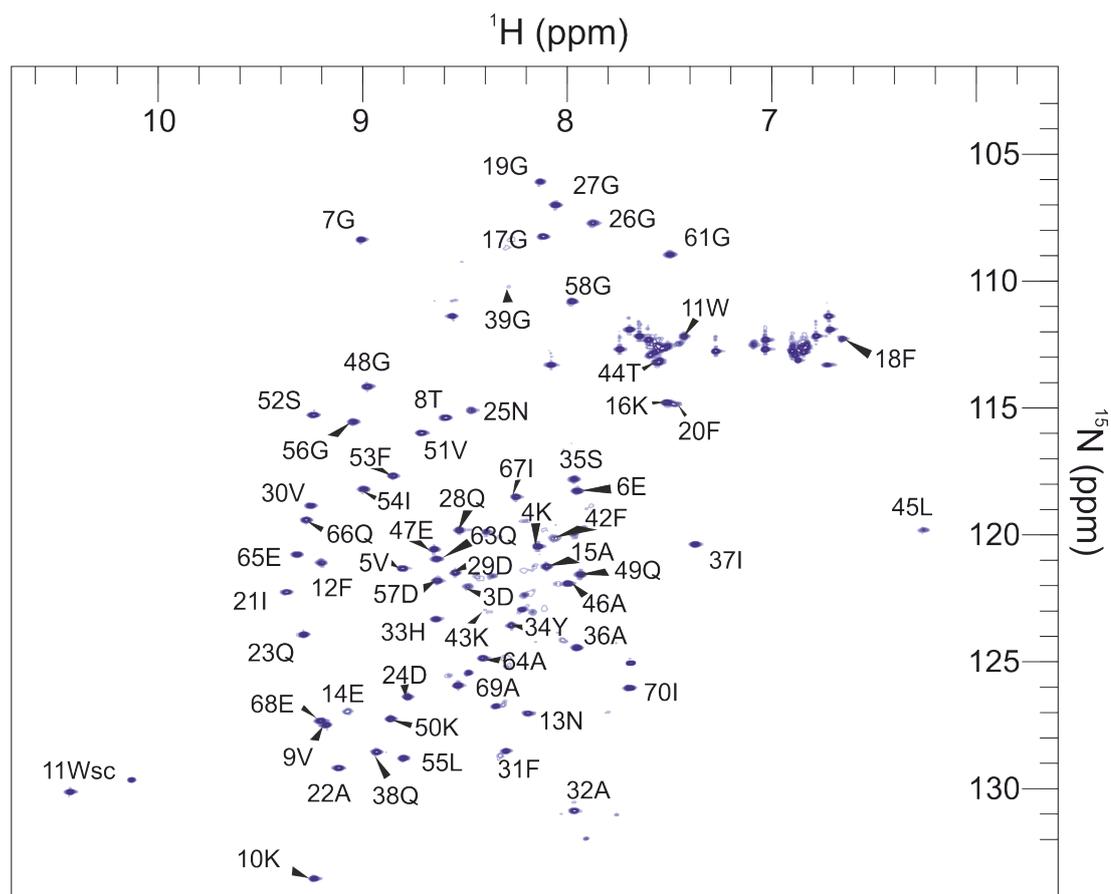


Figure 5.11: HSQC of the *PB6*-CSP with assignments, measured at 20 °C

did not match the missing residues described, so it is likely they are produced by an alternative protein states. While intermediate time-scale chemical exchange will obscure the peaks in the spectrum so that dynamic regions cannot be assigned, slow exchange will produce an extra NMR signal for each state. The lesser populated “excited” state would produce these extra peaks which have lower signal intensity than the peaks produced by the more highly populated “ground” state.

Some of these peaks did have signals in the *C-α* and *C-β* spectra, which connected them

as a group of 4 neighbouring residues as shown in figure 5.12. Identified here as residues A, B, C and D, the chemical shifts of these unidentified signals does match the theoretical chemical shift values of the 4 residues (Ile 67, Glu 68, Ala 69 and Ile 70) that make up the C-terminus of the *PB6*-CSP. It is possible that these extra peaks are from a second protein state in which these residues are in a different chemical environment than they are in the native, ground state.

5.3.2 Investigating the Additional Peaks

It is possible that the excited state residues may be due to unfolded or denatured proteins given their position in the centre of the ^1H spectrum. However, this seems unlikely given the thermo-stability measured for the *PB6*-CSP in figure 5.4b. Figure 5.13 shows a comparison of the centre of the spectrum where most of the extra peaks appear at low temperature (blue), with a spectrum recorded at the same temperature, 5 °C, after the sample had been fully heat denatured (red). The protein was held for a 36 hour period at 37 °C, very near the melting temperature ($T_m = 41.6 \pm 0.4$ °C). This led to a loss of signal of the assigned peaks, without loss of protein concentration by aggregation, suggesting that the signal is generated from the majority of the protein population trapped in the denatured state. While the information gained from this comparison is not ideal as the structure of the denatured state is unknown, it could indicate that the protein is in exchange with a fully or partially unfolded state.

The extra, unassigned peaks at 5 °C seen in figure 5.13 match well with the peaks of the denatured protein. This is an indication that the state causing these extra peaks is a more disordered one, possibly with residues more exposed to the solvent.

The simplest example to explore is the peak from the tryptophan side-chain, shown in figure 5.14. As there is only one tryptophan residue in the *PB6*-CSP sequence, the two peaks seen in blue in figure 5.14a are a clear signal of an extra protein state. The minor peak matches the signal from the heat denatured protein state. This minor state peak also matches the protein state when bound to ssDNA, shown in figure 5.14b which is taken from previous NMR measurements of this protein (Michael Wilson, PhD Thesis) with and without dT7 ssDNA bound. This previous work was performed on the *PB6*-

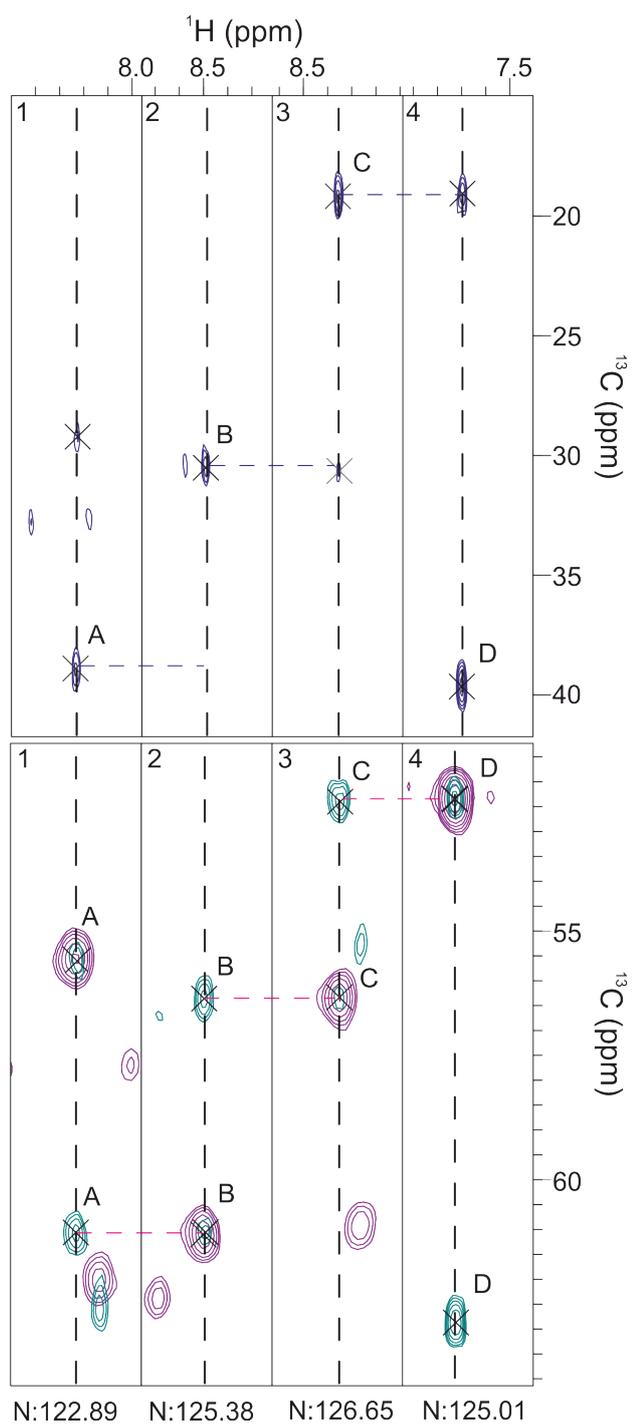


Figure 5.12: HNCA (teal), HNCOCA (purple) and HNCACB (dark blue) spectra corresponding to 4 low signal unassigned peaks.)

CSP with a histidine tag and showed that the presence of ssDNA caused a change in the chemical shift of most residues in the protein. The tryptophan side chain was the only

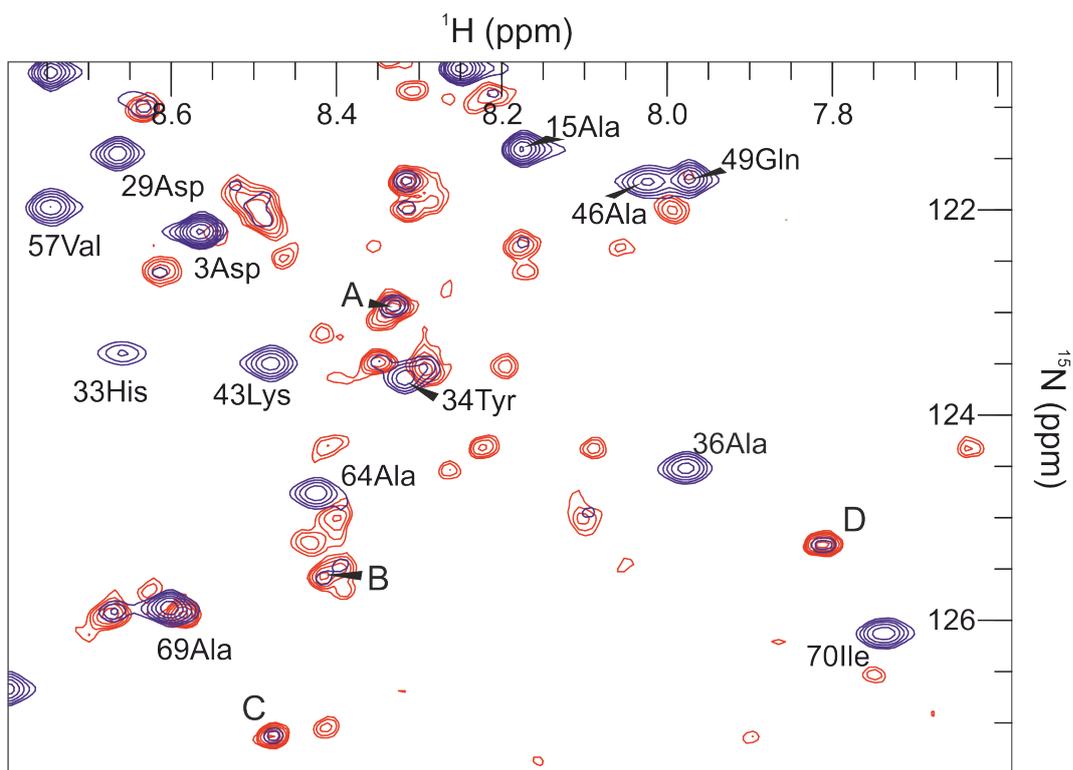


Figure 5.13: HSQC spectra of the *PB6-CSP* at 5 °C (blue), and after heat denaturation (red) when the protein was held near the melting temperature for 36 hours and seemed to become kinetically trapped in a heat denatured state.

clear signal that could be identified.

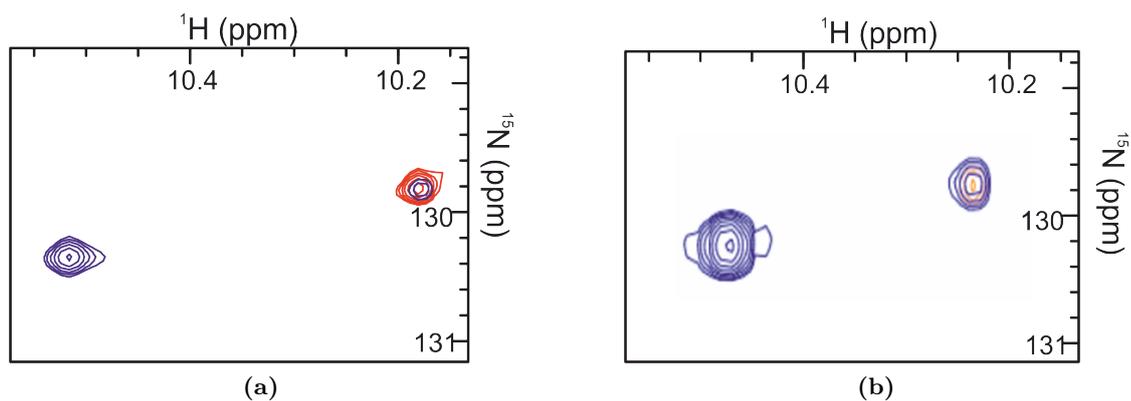


Figure 5.14: HSQC spectra of the Tryptophan side chain of *PB6-CSP*, (a) at 5 °C (blue), and after heat denaturation (red) (b) In free state (blue) ssDNA and with ssDNA bound (orange)

5.4 Relaxation Measurements

5.4.1 ZZ-Exchange

Signals in slow exchange (s) can be identified using the zz-exchange experiment. This will show cross-peaks between peaks that are the result of a particular residue in undergoing slow exchange. These measurements were performed on the *PB6*-CSP by Dr. Arnout Kaldera at 5 and 10 °C to examine if residues in slow exchange could be identified and connected to the additional peaks. Under these conditions a few crosspeaks were seen but it was difficult to connect them with specific residues due to the crowded nature of the spectrum in these conditions. The residues that were identified as being in exchange with one of the additional peaks are shown in figure 5.15. Here the presence of peaks in the zz-exchange spectrum that are not present in the HSQC spectrum, indicate that there is exchange occurring between the the signals aligned with these cross-peaks. These

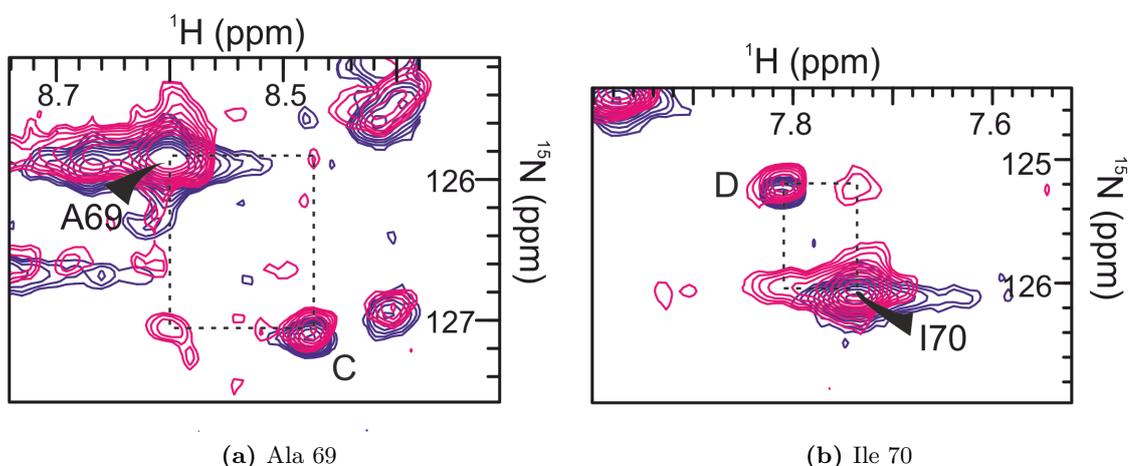


Figure 5.15: ZZ-exchange spectra (Pink) with HSQC spectrum (Blue) at 5 °C for (a) Alanine 69 and (b) Isoleucine 70

results agree with other evidence gathered that sections of the C-terminus β -strand is in slow exchange with another state. Peaks identified as C and D from the triple resonance assignment experiments (figure 5.12) are in exchange with residues Ala 69 and Ile 70 respectively. This is more evidence that these peaks are caused by a different chemical state of the C-terminus, which is in slow exchange with the main “ground” state.

Another exchange signal was identified in an uncrowded region of the spectrum for the

side chain of the tryptophan 11 residue as shown in figure 5.16a. This indicates that the

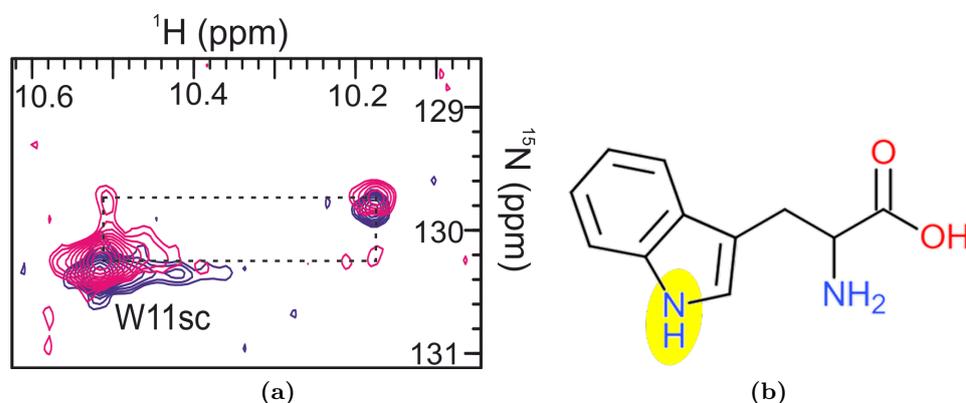


Figure 5.16: (a) ZZ-exchange spectra (Pink) with HSQC spectrum (Blue) at 5 °C for the side chain of tryptophan 11 (b) the tryptophan residue with the side chain amide bond highlighted in yellow

tryptophan side chain residue is in slow exchange with the second state corresponding to a more disordered state, involved in ssDNA binding as shown in figure 5.14.

The location of these residues on the predicted protein structure is shown in figure 5.17. The tryptophan side chain is a part of the hydrophobic surface associated with nucleic acid binding and the C-terminal residues (Ala 69 and Ile 70) are on the opposite side of the β barrel. Between them lies the loop region between β sheets 3 and 4. While the hydrophobic surface area is highly conserved between CSPs^[58], the residues associated with the C-terminus in *PB6*-CSP are some of the least conserved between this protein and those CSPs adapted to warmer environments. Differences in these loop and termini regions could weaken interactions in this region and expose the hydrophobic core to the solvent. The comparable motion of the C-termini residues and the tryptophan side-chain is potentially evidence of allostery being used in the adaptation to the cold^[42].

From the triple resonance experiments connecting peaks A, B, C and D (figure 5.12) and these zz-exchange measurements identifying C and D as alternative states of residues 69 and 70, it follows that peaks A and B correspond to alternative states of residues Ile 67 and Glu 68. As these peaks are also clear in the spectrum of heat denatured *PB6*-CSP (figure 5.13) this alternative state is potentially a more disordered one in which β sheet 5 has peeled away from the rest of the protein structure. This exposing of the hydrophobic core could then lead to the W8 side chain taking a conformation that matches that of

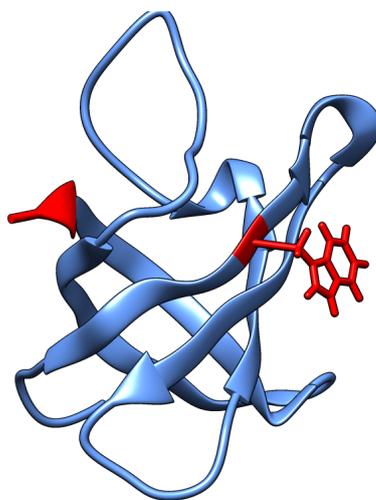


Figure 5.17: Location on the predicted *PB6*-CSP structure of the residues identified as undergoing slow exchange by *zz*-exchange experiments

the denatured and ssDNA bound state seen in figure 5.14. The crowded spectrum and relatively low signal intensity of other cross peaks made it difficult to identify other residues using the *zz*-exchange method, but the presence of these extra peaks indicates that these are not the only residues affected by this slow exchange. Methods that can measure dynamics at faster time-scales were then employed to investigate the sources of different dynamic behaviour between this cold adapted *PB6*-CSP and the mesophilic *Bs*-CSP.

5.4.2 Relaxation Measurements and Time-scales

Proteins in solution are not static, as Brownian motion causes both global movement of the protein and more localised motion depending on protein flexibility. One of the advantages of NMR is that it can be performed in solution so that proteins are in near native conditions. The signal is also sensitive to these motions and so can be used to measure them on a nucleus by nucleus basis. The application of the static B_0 field induces a preferred lower energy orientation of the nucleus, but the motions of the particle change the orientation of B_0 with respect to the nucleus. This will affect the signal obtained and can be used to investigate the motions of each nuclei in the molecule.

The molecular motions of biomolecules occur on a range of times-scales and the relation

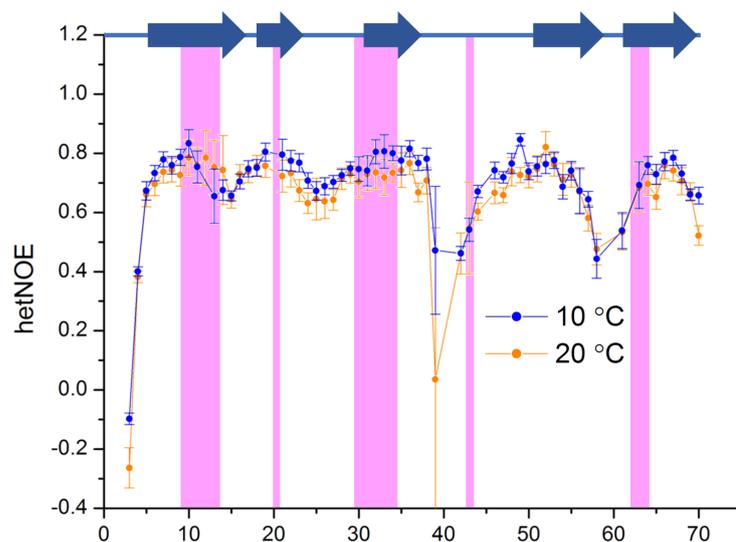
of these motions with biological function is fundamental^{[178], [37]}. A variety of NMR techniques have been developed to explore these motions with each technique providing data specific for a particular time-scale (figure 1.29).

At one end of the time spectrum, fast motions are those that occur faster than the measurement time of the experiment (figure 1.32). This includes bond vibrations and side-chain orientation. On the other end of the spectrum there is real time NMR, for processes slower than the time of the measurement (usually 10s of seconds). If the process results in a change of signal then this can be tracked in real time. In between these two extremes of time-scale more refined experimentation is required if the protein exhibits dynamics. ¹⁵N-¹H-NOE relaxation measurements are influenced mainly by the fast, ps - ns motions that cause amide bond vibration. Transverse relaxation rates are also influenced by fast motions, but if there is slower exchange with another state this will also have an effect that needs to be separated for analysis. Relaxation dispersion measurement measure these slower, μ s - ms rates if they are present.

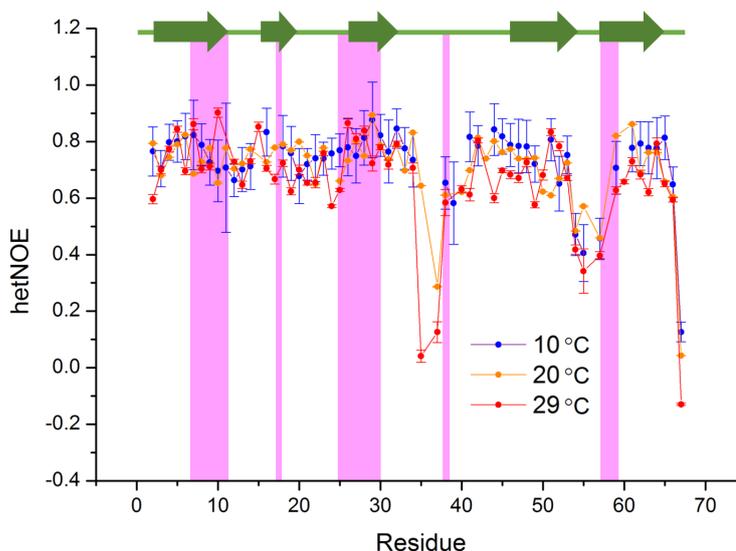
5.4.3 ¹⁵N-¹H-NOE Relaxation

¹⁵N-¹H-NOE relaxation measurement gives an indication of fast dynamics of the hydrogen and nitrogen nuclei. The resulting ratio of saturated intensity against unsaturated intensity can range from 1 (less dynamic) to between -3 and -4 (more dynamic) depending on the field strength of the NMR static field, B_0 .

The ¹⁵N-¹H-NOE of the *PB6*-CSP at 10 and 20 °C is shown in figure 5.18a. The five β sheets expected from the predicted structure are clearly identifiable in the graph. The N terminus appears to be more flexible than the C terminus, with low values for residues 2 and 3 at both temperatures. A similarly dynamic region is seen between residues 38 and 43, corresponding to the predicted loop region between β strands 3 and 4. Values for residues in this region, and around residue 60, are missing and these correspond to residues that were not identified during assignment. As these are present in highly flexible regions, it can be surmised that these peaks are not clear in the spectrum due to broadening of the signal by decreased relaxation rates.



(a) *PB6-CSP*



(b) *Bs-CSP*

Figure 5.18: The ^{15}N - ^1H -NOE ratios for (a) the *PB6-CSP* measured at 10 and 20 °C (b) the *Bs-CSP* measured at 10, 20 and 29 °C. Protein structure and binding regions (pink) according to structures shown in figures 5.1 and 5.2 are displayed.

The ^{15}N - ^1H -NOE of the *Bs-CSP* is shown in figure 5.18b. Compared to the *PB6-CSP* profile, the loop regions are less distinct except for the region between β -sheets 3 and 4, over residues 35-40. Here the C-terminus appears more flexible than the N-terminus. Temperature does not seem to affect the β -sheet parts of the structure in this temperature range, and linearly increases the dynamics of the loop region and

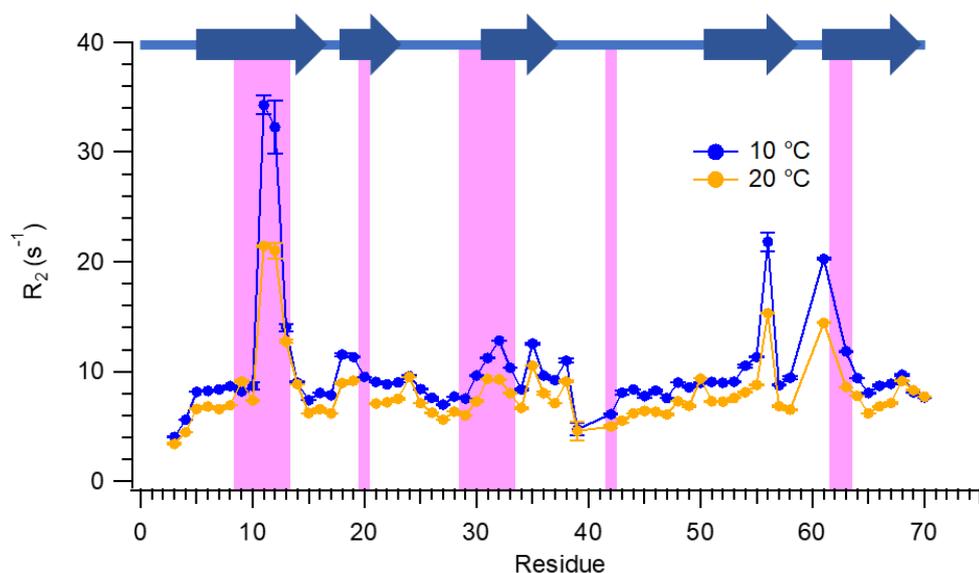
C-terminus.

Comparing the dynamics of figures 5.18a and 5.18b, there is little difference in the β -sheet regions. The *PB6*-CSP has regions with a lower ^{15}N - ^1H -NOE ratio, indicating more ps-ns dynamics, but this does not vary much with changing temperature. The loop regions for both *PB6*-CSP and *Bs*-CSP are clear, have similar values in both proteins and show increasing flexibility (decreasing ^{15}N - ^1H -NOE ratio) with increasing temperature. The N terminus, at the start of the protein sequence, shows more evidence of fast dynamics in the *PB6*-CSP than in the *Bs*-CSP, and appears to be the most dynamic region of the *PB6*-CSP at this time-scale. The more dynamic region of the *Bs*-CSP is at the C terminus, which appears less dynamic in the *PB6*-CSP. The N and C termini are the regions showing the clearest difference between *PB6*-CSP and *Bs*-CSP at this time-scale, otherwise they appear to undergo similar dynamics at this time-scale.

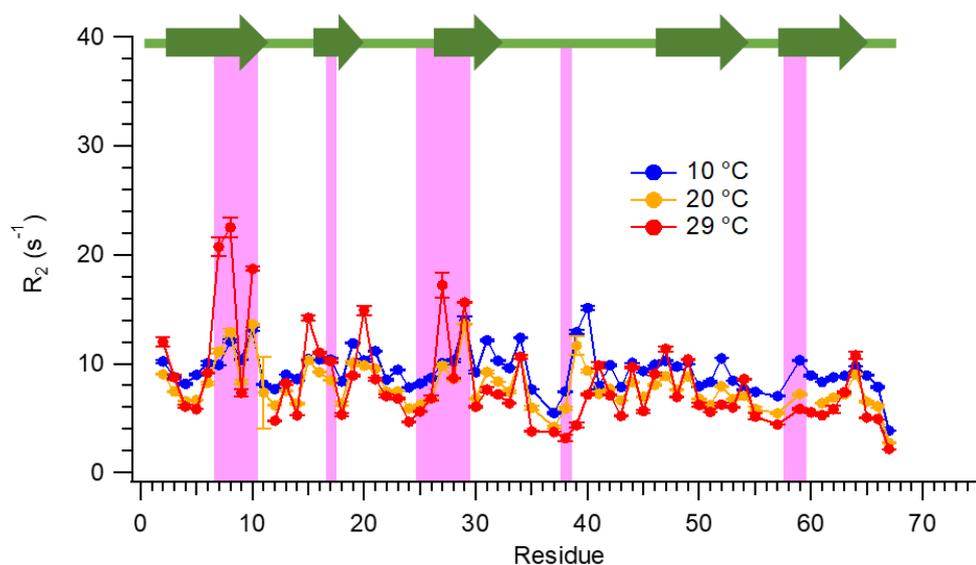
5.4.4 Transverse Relaxation

The transverse relaxation rate, R_2 , is the rate at which spin coherence is restored. The measured value of this is influenced by both fast dynamics and any exchange between different chemical states. While this is affected by the overall tumbling of the molecule, differences across the protein can indicate more flexible and dynamic regions.

Figure 5.19a shows the transverse relaxation rate measured across the *PB6*-CSP protein. Experiments at 29 °C were attempted but the protein was not stable enough in these conditions for the length of time needed for the experiments. As this is a temperature at which the protein starts to denature according to the CD melt, holding this temperature for an extended length of time seemed to cause an irreversible denaturing of the molecules. While the loop region that was shown to be most dynamic by ^{15}N - ^1H -NOE relaxation measurements also shows low R_2 relaxation, other regions show a higher R_2 rate that increases with decreasing temperature. In particular the residues 11 and 12, tryptophan and phenylalanine, show a large increase in R_2 on decrease in temperature. To identify which residues were most affected by the change in temper-



(a) *PB6-CSP*



(b) *Bs-CSP*

Figure 5.19: R_2 rates for (a) the *PB6-CSP* measured at 10 and 20 °C (b) the *Bs-CSP* measured at 10, 20 and 29 °C. Protein structure and binding regions (pink) according to structures shown in figures 5.1 and 5.2 are displayed.

ature the R_2 rates measured at 20 °C are plotted against the rates measured at 10 °C in figure 5.20. Outliers 39G, 70I, 69A, 9V, 14E, 50K, 24D, 68E and 13N have R_2 rates that increase with decreasing temperature to a lesser extent than most other residues in the structure. Residues 43K, 58G, 30V, 63Q, 32A, 61G, 56G, 12F and 11W have R_2 rates that increase at 10 °C more than the rest of the structure. This difference in

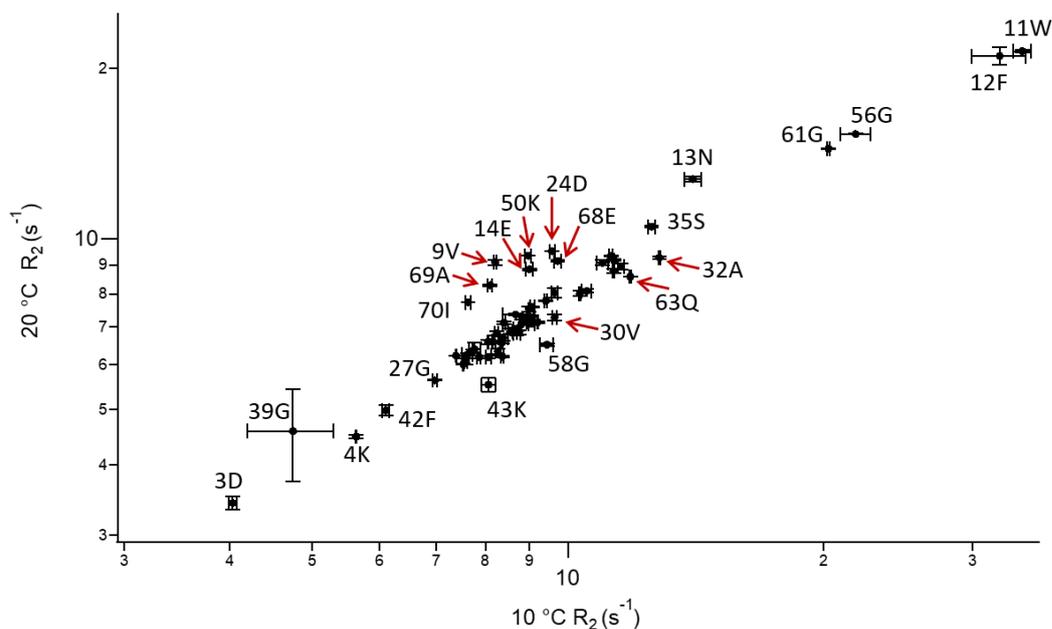


Figure 5.20: R_2 rates for the *PB6*-CSP measured at 20 °C plotted against R_2 rates measured at 10 °C. Residues that appear as outliers are identified.

temperature dependence suggests localised dynamics that have differing responses to a change in temperature, potentially due to exchange with at least one alternative state. These extra states represent alternative protein conformations with different chemical shift signals to the native state. The rate of exchange, R_{ex} will change with temperature in a manner dependent on the relative energies of the two states and that of the transition state separating them. This feature is explored more by the relaxation dispersion measurements in section 5.4.5.

The *Bs*-CSP was recorded at 10, 20 and 29 °C, as shown in figure 5.19b. Between 10 and 20 °C this appears to follow a linear relationship, with no region standing out as particularly affected by the change in temperature. This can be seen in figure 5.21, where like figure 5.20 the R_2 rates recorded at 20 °C are plotted against the R_2 rates recorded at 10 °C. Here the outliers appear to be those residues that have R_2 rates that increase at 20 °C more than the rest of the structure. These include residues 54G, 41L, 11S, 7K, 8W, 10N and 29H, with only 59Q and 40T appearing to increase more at the lower temperature 10 °C.

When the temperature is raised to 29 °C, the temperature classed as "cold-shock" for

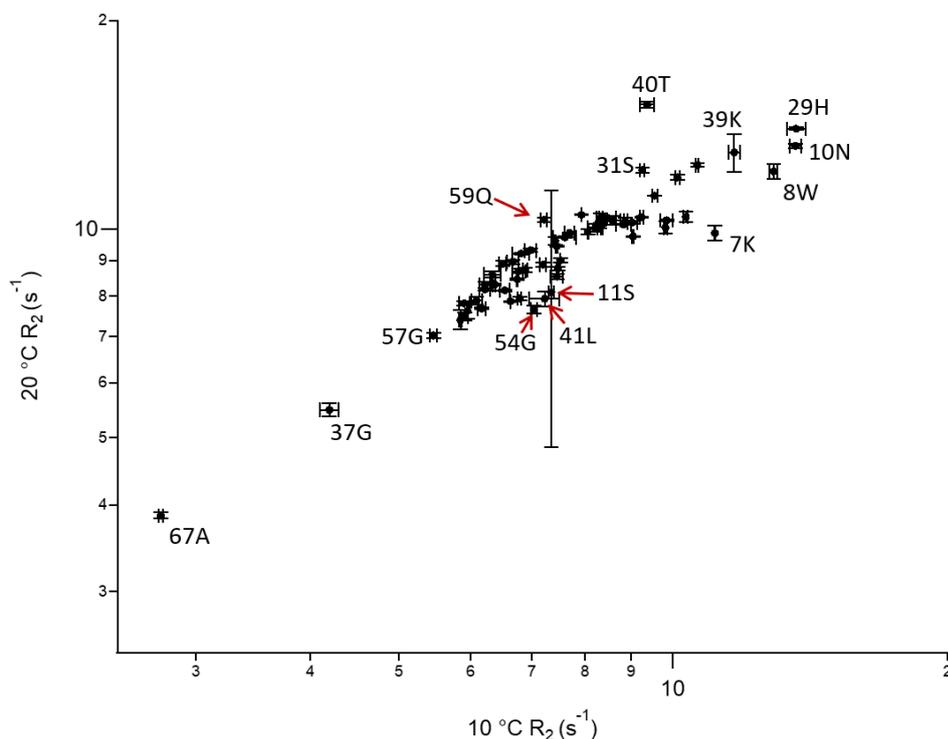


Figure 5.21: R_2 rates for the *Bs*-CSP measured at 20 °C plotted against R_2 rates measured at 10 °C. Residues that appear as outliers are identified.

the organism *Bacillus Subtilis*, then some regions of the backbone show a significant jump. This includes the tryptophan residue, which is at position 8 in this protein as opposed to 11 in the *PB6*-CSP . Residue 10 disappears above 10 °C, indicating intermediate exchange at these temperatures obscuring the signal. Figure 5.22 shows the values of R_2 measured at 29 °C plotted against those recorded at 20 °C, so the effect of increased temperature on each residue can be seen more clearly. Many residues show an increase in R_2 at the higher temperature in this protein. The residues 36A, 38Q and 39G from the loop region between β strands 4 and 5 show a decrease in R_2 at 29 °C, indicating an increase in ps-ns dynamics.

Comparing the change in R_2 dynamics with temperature of the *PB6*-CSP and the *Bs*-CSP, it appears that the regions associated with ssDNA binding (shown in pink in figure 5.19) show the most temperature dependence. For the *PB6*-CSP this is through higher R_2 values that are increased with decreasing temperature. For the *Bs*-CSP these regions are not distinct from the rest of the protein at lower temperatures, but when the

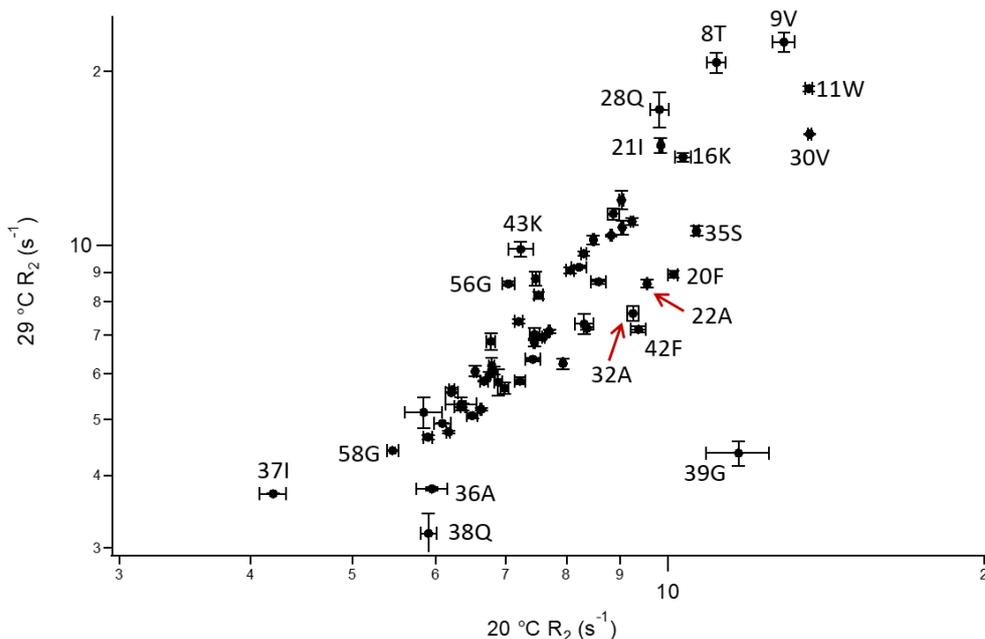


Figure 5.22: R_2 rates for the *Bs*-CSP measured at 29 °C, close to the cold-shock temperature for this organism, plotted against R_2 rates measured at 20 °C. Residues that appear as outliers are identified.

cold-shock temperature (29 °C) is reached these regions show a considerable increase. For this CSP variant, localised dynamics increase with increasing temperature. As this change only occurs in some parts of the protein it is likely due to a change in localised dynamics. It is also not seen on in the ^{15}N - ^1H -NOE ratios (figure 5.18), so is likely due to slower dynamics, which account for the extra R_{ex} term in equation 1.18.

5.4.5 Relaxation Dispersion measured with CPMG

μ -ms dynamics between states that have distinct chemical shift signals (figure 1.32) in the HSQC spectrum were measured using relaxation dispersion NMR spectroscopy^[179]. These μ s - ms dynamics account for the extra term R_{ex} in equation 1.18, because during the measurement of R_2 rates any residue that spends time in a second alternative state and returns to the state being measured will change the resulting $R_{2,measured}$. Carl-Purcell-Meiboom-Gill experiments use increasing numbers of refocusing pulses to amplify the effect of the second state, R_{ex} , as described in section 1.5.5. This allows the separation

of R_{ex} from the measured $R_{2,measured}$ value, to obtain the true $R_{2,0}$ reading. These results for CPMG measurements on the *PB6*-CSP recorded at 10 °C are shown in figure 5.23. Some regions of the protein display R_{ex} values indicating the presence of localised dynamics at this μs - ms time-scale.

R_{ex} values have been recorded at 25 °C for the *Bs*-CSP both with and without dT7

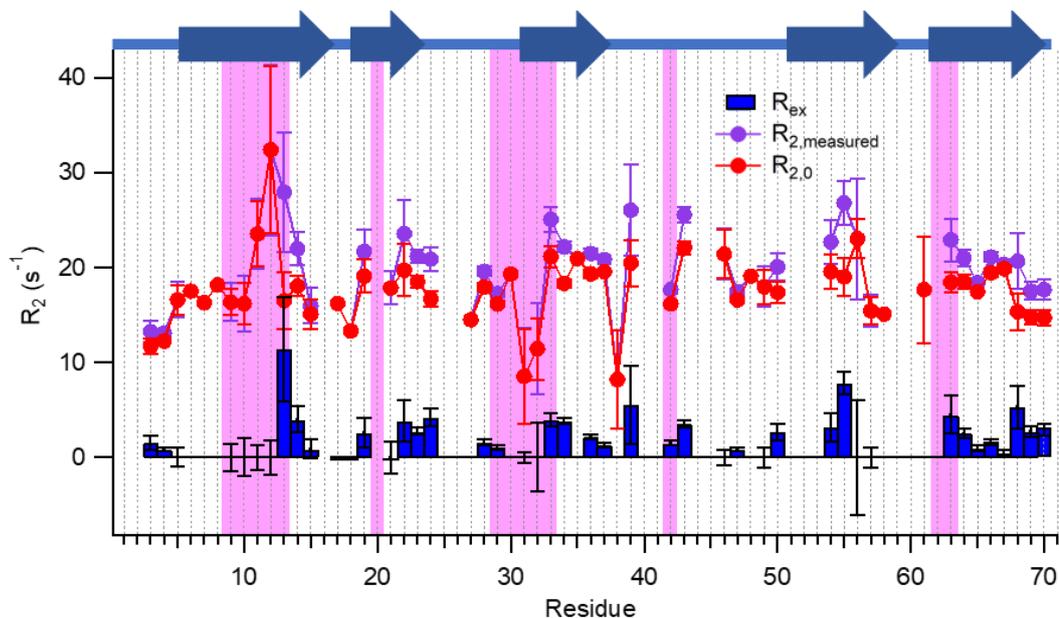


Figure 5.23: The calculated R_2 values from the fits to CPMG dispersion curves. $R_{2,0}$ (red) is the transverse relaxation rate when R_{ex} (blue bars) is excluded. $R_{measured}$ (Purple) is the sum of the other two, so what would be measured without accounting for R_{ex} . Protein structure and binding regions (pink) according to structures shown in figures 5.1 and 5.2 are displayed.

ssDNA present^[59]. Values up to 10 s^{-1} were recorded in the free state and was associated with global unfolding at a rate of 12 s^{-1} of the protein with a refolding rate of 1070 s^{-1} . Most of these dynamics disappeared with the addition of the dT7 oligonucleotide, indicating that ssDNA binding stabilised the native state.

Example relaxation curves are shown in figure 5.24, with fits that showed fast and slow exchange for different residues. Relaxation curves with R_{ex} fits for all of the residues measured can be found in figure B.5. R_{ex} is roughly the difference between R_2^{eff} recorded at $\nu_{CPMG} = 0$ and $\nu_{CPMG} 1000 \text{ Hz}$. The relaxation curve of R_2^{eff} plotted against ν_{CPMG} can be analysed to extract information about the two states A and B. This includes the fractional population of state A (P_A), the rate of exchange between the two states (k_{ex})

and the difference in chemical shift between the two states ($|\Delta\omega|$). For example, for W11, results of the fit of the relaxation dispersion curve to the Carver-Richards-Jones all-timescales dispersion equation^[155] in figure 5.24a were $k_{ex} = 2000 \pm 1600 \text{ s}^{-1}$ and $P_a = 99 \pm 9 \%$, errors calculated by Monte-Carlo fit. This has a high uncertainty due to noise in the data and because the fit was performed with all parameters of the Carver-Richards-Jones all-timescales dispersion equation^[155] being left free. Using a similar approach for Q28, the fit in figure 5.24b gave values $k_{ex} = 3 \pm 3 \text{ s}^{-1}$ and $P_a = 79 \pm 3 \%$. This demonstrates that at 10 °C, the *PB6*-CSP displays a range of dynamics spanning different time-scales across different sections of the protein.

Recording these experiments at two different B_0 fields allows the separation of the

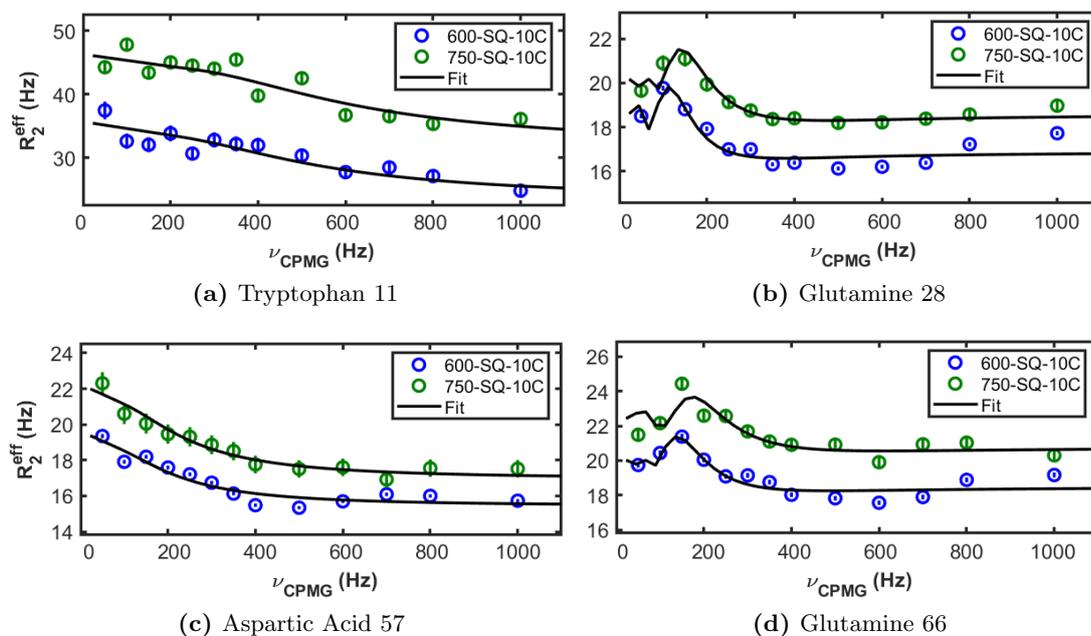
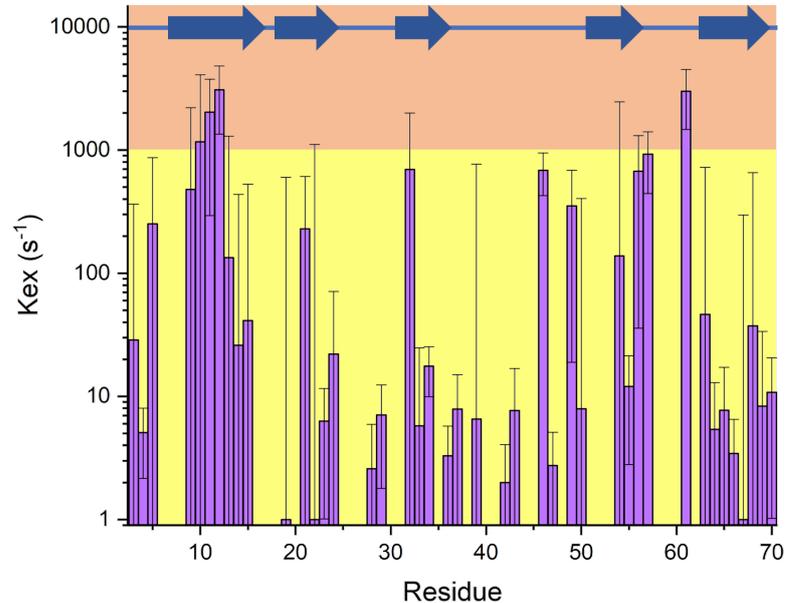
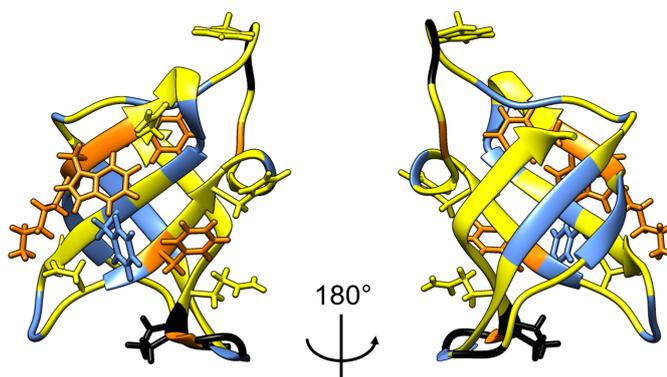


Figure 5.24: Example relaxations with fits at two B_0 fields. (a) Residue 11 is a Tryptophan with $k_{ex} = (2,000 \pm 1,700) \text{ s}^{-1}$ (b) Residue 28 is Glutamine with $k_{ex} = (2.6 \pm 3.3) \text{ s}^{-1}$ (c) Residue 57 is Aspartic acid with $k_{ex} = (930 \pm 480) \text{ s}^{-1}$ (d) Residue 66 is Glutamine with $k_{ex} = (3.4 \pm 3.1) \text{ s}^{-1}$. Fits for all of the residues can be found in figure B.5

chemical shift difference $|\Delta\omega|$, as this will change with B_0 while all other variables should remain constant. To explain the extra peaks appearing in the spectrum, the residues exchanging with these states need to demonstrate slow exchange with a low enough P_a that would generate a state visible in these measurements.



(a) k_{ex} of the *PB6-CSP*



(b) k_{ex} mapped onto protein predicted structure

Figure 5.25: Values measured at 10 °C of (a) k_{ex} for the CPMG measurements on the *PB6-CSP* as calculated using the GUARDD program. Uncertainties calculated by Monte-Carlo. (b) These values plotted onto the protein predicted structure. μ s dynamics shown in orange, ms in yellow and no measured exchange shown in blue. Black indicates residues that were not assigned. Side chains associated with nucleic acid binding are shown in colours matching the backbone amide motion. The relaxation curves and R_{ex} fits used to generate this data are shown in figure B.5

Figure 5.25a shows the k_{ex} values calculated using the GUARDD software for the *PB6-CSP*. This shows a variety of dynamics with some residues showing no evidence of exchange, while other regions show very fast, over 1000 s^{-1} , exchange. Errors were calculated using the Monte-Carlo method and are very high for some residues, indicat-

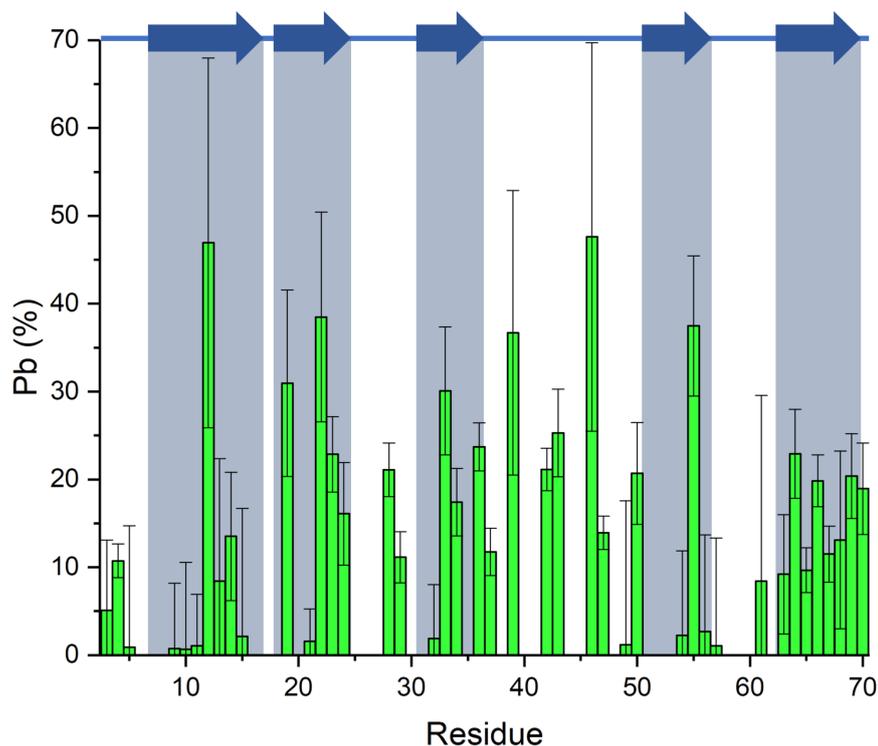


Figure 5.26: Values of P_b for the CPMG measurements on the *PB6*-CSP as calculated using the GUARDDD program. Uncertainties calculated by Monte-Carlo.

ing a noisy dispersion curve. The region between residues 9 - 15 shows a consistent dynamic curve, with slower dynamics surrounding the faster dynamics of residues 11 and 12. As residue 11 is the tryptophan thought to be key to binding nucleic acids, this extra low temperature flexibility could be key to the low temperature function of this protein. Another region showing fast dynamics is between residues 55 and 61. Residues 59 and 60 were not assigned, probably because the signal in the backbone measurements was obscured by intermediate dynamics. This can be seen in figure 5.25a as they are surrounded by residues displaying R_{ex} values between the μ s-ms range.

Parts of the protein also show dynamic behaviour at the slower, millisecond time-scale. The first region to stand out as being part of a slow exchange is the 5th β -sheet connected to the C-terminus, residues 60 - 70. This is also a region that shows a sizeable fraction to be in the excited B state, as shown in figure 5.26. Residues 62 - 70 all show a B state population above 10 %, which could contribute to the slower dynamics observed in the *zz*-exchange experiments in section 5.4.1 and the extra peaks seen in the HSQC

spectrum, figure 5.5, at 10 °C.

5.4.6 Region Separation in Fit of Relaxation Dispersion

By separating the protein into regions and fitting the dispersion curves in groups, areas that move together can be identified. This links the parameters Pa and k_{ex} , and fits the dispersion curves for a group of residues assuming these values will be consistent. The quality of the resulting fits can then be assessed and the grouping judged. This can give a better idea of the Pa and k_{ex} parameters as it brings more restrictions into the non-linear fitting of the curves.

As the residues adjacent to the C-terminus of the *PB6*-CSP in these conditions seem to show some consistently slow dynamics they were grouped in various arrangements. Some of the groupings that gave good fits included residues 66 - 70. The fit of this region gave a k_{ex} value of $6 \pm 3 \text{ s}^{-1}$ and a Pa value of $77 \pm 3 \%$. Next to these the group of 3 residues 63 - 65 was fit with the k_{ex} value of $700 \pm 500 \text{ s}^{-1}$ and a Pa of $99 \pm 3 \%$. Another region that still generated a good fit of each dispersion curve when grouped together was residues 10 - 13. This group fit with a Pa value of $50 \pm 20 \%$ and a k_{ex} of $7000 \pm 2000 \text{ s}^{-1}$.

5.5 Discussion

HSQC spectra of *Bs*-CSP and *PB6*-CSP both showed temperature dependent changes, with evidence of fast (ps-ns) dynamics increasing at higher temperatures for the *Bs*-CSP. This is represented in figure 5.27, with the higher temperatures resulting in fast exchange between the native state A and a more excited state B. This is one explanation for the line broadening observed in figure 5.7 and the increase in R_2 relaxation rates at 29 °C seen in figure 5.19b.

The *PB6*-CSP HSQC spectra shows an increase in the number of peaks at low temperatures, indicating slow exchange with a second conformational state. Relaxation measurements of the *PB6*-CSP show an increase in R_2 relaxation at 10 °C, indicating

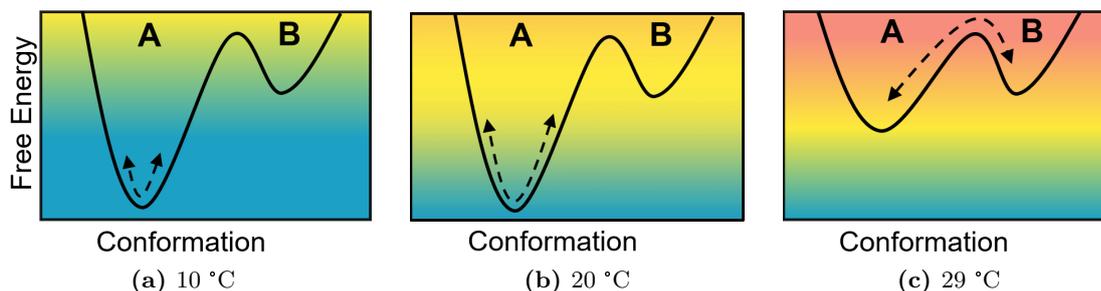


Figure 5.27: Model of *Bs*-CSP dynamics at (a) 10, (b) 20 and (c) 29 °C represented by energy landscape diagrams. The protein native state A is fairly stable at 10 and 20 °C, but at roughly the "cold shock temperature" of 29 °C dynamics increase as the thermal energy increases the rate of exchange between the more excited, higher energy state B.

the increase in intermediate (μ s-ms) exchange in some regions at this lower temperature. This effect could be explained by the presence of a second state B that is a more frustrated state than the native state A, as shown in figure 5.28. The second state B becomes more energetically favourable at low temperatures therefore increasing the number of molecules in this state at equilibrium. As state B is more frustrated, this increase in population increases the μ s-ms dynamics and hence the R_{ex} contribution to the R_2 measurements as well as the slower exchange between states A and B. This

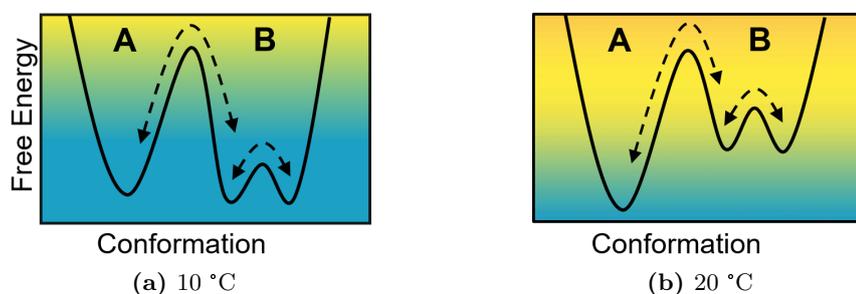


Figure 5.28: Model of *PB6*-CSP dynamics at (a) 10 and (b) 20 °C represented by energy landscape diagrams. Exchange between the native state A, and a more frustrated state B changes with lowering of the temperature, as the more frustrated state becomes more energetically favourable at 10 °C.

faster exchange is occurring in the backbone region around the tryptophan residue known to be involved in ssDNA binding. The side chain of this residue is involved with nucleic acid binding for most CSPs studied previously^[44]. Evidence from a range of measurements show this side chain is involved in slow exchange with a state that has the same chemical shift signal as that in the denatured and ssDNA bound state.

This residue, which is conserved between CSP homologues, has been shown in the hyperthermophile *Tm*-CSP to be a key control for ssDNA binding at the cold shock temperature of the organism^[70]. This could also be the case for low temperature ssDNA binding in this psychrophile homologue, where the excited state seen at lower temperatures allows for better nucleic acid binding.

Slower exchange is seen around the β -strand closest to the C-terminal, connected to more fast exchange in the loop region between β -strands 4 and 5. Though in different regions of the protein, these dynamics could be connected through allosteric interactions. This could be the difference between the native state A, and the more frustrated state B as seen in figure 5.28a. If this C-terminus region is looser in the second state B, this would expose the hydrophobic core to the solvent resulting in a less stable structure, an effect that can be seen during the cold denaturation of some proteins^[18]. The aromatic residues associated with nucleic acid binding are highly conserved between CSPs, so mutations for cold adaptation that increase the flexibility of this region remotely is one possible interpretation. When comparing the conservation of CSP sequences it is the C-terminus of *PB6*-CSP that has the least similarity with other CSPs. A solution for the solvent structure of the *PB6*-CSP would allow these results to be explored in more detail. The high quality HSQC spectrum indicates that this would be possible at 10 or 20 °C. Looking at the arrangement of amino acid side chains in these regions could help elucidate which are the likely cause of these low temperature dynamics not seen in the mesophile *Bs*-CSP dynamic measurements. The C-terminus motion could also be due to a lack of evolutionary pressure for stability in this region in lower temperature environments. CPMG experiments performed at a different temperature, like 20 °C could help elucidate the temperature dependence of these dynamics. It would also be interesting to perform these measurements in the presence of ssDNA, or even to obtain a structure of the protein while in complex. This could be compared with previous structures of the complex *Bs*-CSP with dT6 ssDNA. This could show if the binding surface is truly conserved, and how the rest of the protein is altered by this interaction. To act as an RNA chaperone, flexibility is key at the operating temperature of the protein^[20]. The role of protein disorder in cold environments has been demonstrated to aid freeze tolerance^[40], but is also key for nucleic acid interactions^[180]. The role of

localised protein unfolding in regions far from the active site have even been used to increase low temperature activity in the *E. coli* adenylate kinase^[42]. Mutations made which favoured localised unfolding far from the binding site were used to tune low temperature activity and affinity through dynamic allostery.

CSPs from organisms that are adapted to colder environments, *Listeria monocytogenes* and *Colwellia psychrerythraea* 34H have been studied previously using NMR^[68,69]. These studies measured the dynamics with and without dT7 ssDNA at 25 °C, showing these proteins to be more flexible than mesophilic versions, with dynamics that are reduced in the presence of ssDNA. This adds evidence to the theory that cold adapted proteins need to be more flexible, but at 25 °C it does not compare the proteins at the temperatures at which they function *in vivo*.

Here the low temperature frustration of the *PB6*-CSP has been demonstrated by measured chemical exchange not observed in the mesophilic *Bs*-CSP dynamics. While it has not been shown that this directly affects the low temperature chaperone activity, recent experimental developments have shown that this can be measured with real time NMR^[63]. Without these experiments it cannot be confirmed that the *PB6*-CSP adaptations aid low temperature function, but the relative dissociation constants provide some evidence that function is tailored to organism temperature.

6

Discussion and Future Work

6.1 Summary

In section 1.6 certain questions relating to the cold shock proteins were outlined. Here these questions are answered and discussed on the basis of the research contained here and previous studies.

Can the effect of nucleic acids binding to the *Bs*-CSP be measured in the protein unfolding force?

The mechanical unfolding fingerprint of the *Bs*-CSP while in a complex with ssDNA was recorded in the experiments described in chapter 3. The presence of ssDNA oligos that are known to bind to the *Bs*-CSP resulted in an increase in the unfolding force of the *Bs*-CSP while the presence of oligos that do not bind with high affinity did not. The unfolding of free *Bs*-CSP and ssDNA bound *Bs*-CSP domains resulted in the same recorded increase in contour length. This implies that the mechanical clamp, the region of the protein mainly responsible for determining the unfolding force, is in the same location on the protein whether bound to ssDNA or not. Previous studies on the effect of single stranded nucleic acid binding on protein mechanics have shown that ssDNA binding over the mechanical clamp region can induce a large increase in the force required to unfold

the protein^[112]. In this study the ssDNA is thought to have bound on the opposite side of the protein structure to the mechanical clamp, but still caused an increase in unfolding force. The ssDNA binding could be stabilising a protein conformation that is favourable for the forces holding the mechanical clamp together. This implies a correlation between the mechanical clamp (β -strands 1, 4 and 5) and the nucleic binding site, identified in the crystal structure as mainly crossing β -strands 1, 2 and 3 (figure 5.1).

What are the mechanical properties of the *Bs*-CSP when bound with nucleic acids?

Unfolding the (I27-*Bs*CSP)₃-I27 at a range of pulling speeds in the presence and absence of ssDNA showed an increase in the stabilising effect of ssDNA binding at lower speeds. This effect was maintained when the temperature was reduced to 5 °C. Other SMFS studies on the effect of ligand binding have found that stabilisation usually involves a decrease in protein malleability^[105]. The *Bs*-CSP has already been shown to be resistant to mechanical unfolding^[83]. In this study the addition of flexible ssDNA has been shown to increase the malleability across the *Bs*-CSP. This result is represented in figure 6.1, where the addition of ssDNA across the binding site stabilises the transition state of the protein, allowing the protein to be extended further by the mechanical force before it unfolds. The addition of a flexible molecule to a small, rigid protein could be a tactic used to increase the stability of proteins used in a biotechnological role without reducing the flexibility.

In the nucleic acid bound state, maintained flexibility should be inherent to the protein function as an RNA chaperone^[20]. In particular, the role of the CSP as a response to cold stress implies that this function must be maintained at low temperatures, and a temperature dependent malleability has already been recorded in a hyperthermophilic CSP using SMFS^[7]. This hyperthermophilic protein showed the typical protein temperature softening^[161], with the distance from the folded to the transition state, Δx_U increasing with increasing temperature as can be seen in figure 1.19. This effect was also observed with the unfolding of the free *Bs*-CSP, seen in figure 3.16a, with a very small decrease (0.3 nm) in Δx_U at 5 °C. In contrast, comparing the two unfolding

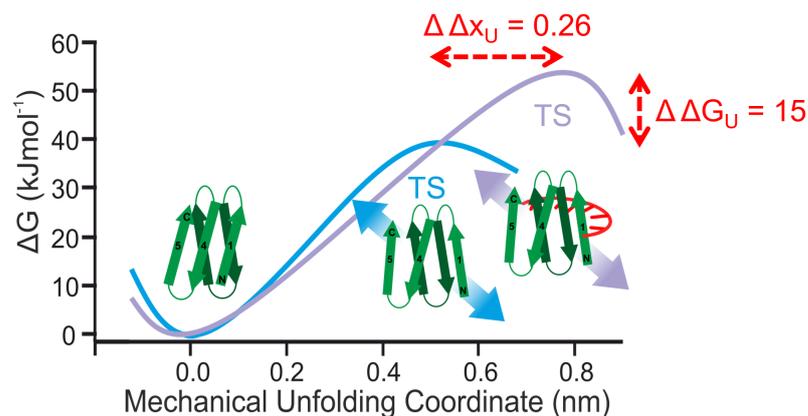


Figure 6.1: Schematic of the effect of ssDNA (red) binding on the transition state (TS) of unfolding when the *Bs*-CSP is under mechanical force across the N and C termini at 5 °C.

schematics for the *Bs*-CSP-ssDNA complex at 23 and 5 °C (figure 3.16b), the opposite effect can be seen. By including the nucleic acid bound state in these studies, more insight into the mechanisms that aid function of this particular protein was attained. In future a more general understanding of how ligand binding affects protein mechanics could unlock the potential for future protein and nucleic acid based biotechnologies.

Is the binding affinity of the CSPs to nucleic acids regulated relative to the optimal growth temperature of the organism?

When cold shock proteins are induced at low temperatures they act as RNA chaperones melting RNA hairpins caused by the low temperatures^[63]. Here, cold shock proteins from the mesophile *Bacillus subtilis* (optimum temperature of 37 °C) and the psychrotroph *Psychrobacter 6* (optimum temperature 20 °C) were used in ssDNA binding studies. In this study an approximation was made by measuring the affinity to short ssDNA oligonucleotides assuming this to be correlated with enzyme activity because the role of enzyme activity involves the binding and releasing of the substrate. The RNA and ssDNA binding surface has been identified as identical in the *Bs*-CSP^[58], so ssDNA was used as it is the response of the protein that is being studied here. The temperature dependence of the protein affinity to ssDNA was measured for *Bs*-CSP and *PB6*-CSP. Both were found to have highly temperature dependent binding affinities to the ssDNA oligonucleotide dT7. At the “operating” temperature of each protein (the cold shock

temperature defined as 7-10 °C below the optimal growth temperature of the organism) both proteins demonstrated a similar affinity to dT7. At 30 °C the *Bs*-CSP affinity to dT7 was measured to have a K_D of 48 ± 7 nM, and the *PB6*-CSP had an affinity to dT7 at 10 °C with a K_D of 36 ± 6 nM. This implies a very controlled affinity so that binding is of the appropriate strength in the conditions that this protein is required to function. Following this logic, proteins from organisms adapted to higher temperatures should have an increased binding affinity. This has been observed in a single stranded ssDNA binding protein from a hyperthermophilic organism^[181]. It has also been observed in the CSP from the thermophilic *Bacillus caldolyticus*, which was measured binding to dT7 with a K_D of 0.9 ± 0.2 nM at 15 °C^[57], half that of the mesophilic *Bs*-CSP .

The thermodynamics of the binding of these CSPs to ssDNA was explored further using ITC in section 4.2. These experiments also provided values for K_D of each interaction, given in table 4.3. These values are much higher than those measured using tryptophan fluorescence quenching, with the K_D of the *PB6*-CSP binding to dT7 at 10 °C increasing by almost 7 fold. This problem has been seen before, with *Bs*-CSP-ssDNA interactions measured in previous studies also showing higher K_D values when ITC was used. The interaction between *Bs*-CSP and dT7 measured at 15 °C by Max *et al.*^[62] had a K_D value of 1.8 ± 0.4 nM when using tryptophan fluorescence quenching while a K_D of 12.1 ± 0.7 nM was measured by Sachs *et al.*^[58] when using ITC. This is also almost a 7 fold increase. This issue could be due to the need for high concentrations in ITC, decreasing the precision for measuring very low K_D values. It could also be due to the effects of partial binding shown in figure 4.15. Tryptophan fluorescence quenching would not differentiate between complexes that only partially cover the protein surface, as the fluorescence is quenched if the single tryptophan residue is bound to a nucleotide. The results in the current study showed that the *Bs*-CSP and *PB6*-CSP bound to CT2 and dT7 with similar thermodynamic parameters at 25 and 10 °C respectively. When the *PB6*-CSP was measured binding to dT7 at 25 °C, the thermodynamic parameters calculated were used to predict the binding affinity at lower temperatures, as shown in figure 4.9b. If these parameters were consistent at the cold-shock temperatures, the *PB6*-CSP

should have a much higher affinity to the ssDNA. Instead at 10 °C, the measurements show that this affinity, and the values of ΔH and ΔS are closer to the *Bs*-CSP values at 25 °C, suggesting that some conditions of this interaction are altered as the temperature is reduced. This could be something in the *PB6*-CSP conformation that reduces the low temperature affinity so that the nucleic acids do not remain bound to the CSP. This would be necessary for the protein to function successfully as an enzyme that increases low temperature activity of nucleic acids, as to bind too strongly would decrease the number of nucleic acids in the free state where they can be used in protein production. The *PB6*-CSP interaction at 10 °C also featured a double peak, not visible at 25 °C, indicating the presence of two processes occurring at different rates after each ssDNA injection. The second peak was smaller and appeared consistently 10 s after the main injection peak, also shrinking in intensity as more ssDNA was added. This could be an indication of a slow conformational change induced by the binding of ssDNA, possibly one of the exchange processes measured by NMR on the *PB6*-CSP at 10 °C. It might even be that this conformational change is what helps reduce the affinity to nucleic acids at low temperatures, to help maintain the enzyme activity.

Can the unbinding of ssDNA from *Bs*-CSP be measured by SMFS?

Some indication that the unbinding force of ssDNA from the *Bs*-CSP had been recorded was seen in section 4.4. This suggests that this is a viable technique, however the unbinding forces could not be differentiated from other non-specific interactions. With optimisation of protein surface concentration and contour length, this system could be employed to measure unbinding forces directly. Techniques that are more sensitive to small changes in force would make this approach more tenable. Investigating this form of extremophile protein adaptation by SMFS could help understand in molecular detail how the binding affinities are tuned to the native temperature for the organism. Applying this to hyperthermophilic proteins interacting with ligands an increase in the forces could be measured. Conversely, if this force could be elucidated for a psychrophilic protein, the force would likely be lower. However if this could be measured with more sensitive methods then evidence of whether disorder in psychrophilic proteins is observed or not at the single molecule level could be measured^[31].

Is there a difference between the *PB6*-CSP and *Bs*-CSP in the way localised dynamics change with temperature?

The temperature dependent dynamics of the *Bs*-CSP and *PB6*-CSP were investigated using protein NMR spectroscopy as reported in chapter 5. It was expected that the protein from the psychrotropic organism would display more dynamic behaviour, as a more flexible structure, matching what other cold adapted CSPs have demonstrated^[68,69]. Dynamics were compared between the mesophilic *Bs*-CSP and the psychrophilic *PB6*-CSP at 10 and 20 °C. A key difference was the effect of this temperature change on the dynamics around the ssDNA binding site. Increasing the temperature appeared to increase μ s-ms dynamics in the *Bs*-CSP with HSQC peaks becoming obscured above 30 °C. The *PB6*-CSP on the other hand, showed the clearest signal at 20 °C, and μ s-ms dynamics that increased as the temperature was lowered. A source of low temperature dynamics to compensate for the lack of thermal energy in psychrophilic organisms can be a frustrated native state, or a rough native energy landscape. This would mean that the low temperature conformation of the protein causes amino acid residues to repel other residues that are brought closer by the compression that comes with lowering the temperature. This then pushes the protein conformation into a range of 3D structures with shallow energy wells. The exchange between these states is a way of compensating for the lack of thermal energy, so the protein can still sample functional conformations at low temperatures, just at a slower rate. The slow dynamics (per second) measured at 10 °C for the *PB6*-CSP seemed to show the C-terminus peeling away from the rest of the protein and exchanging with a more disordered, solvent exposed state. At the same time, fast dynamics were observed around the binding site. It is possible that the the unfolding of β -strand five (the β -strand attached to the C-terminus) from the rest of the protein exposes the hydrophobic protein core. This could then upset the force balances of the rest of the protein structure and generate the frustrated state thought to aid in low temperature activity, as represented in the energy landscape of figure 6.2.

This is only one potential model, and more work would be needed to explore whether or not this is an appropriate description. Experimentally the protein could be grown

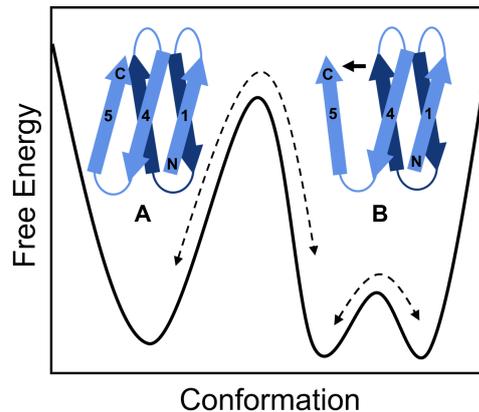


Figure 6.2: Schematic of the effect of reduced temperature on the dynamics of the *PB6*-CSP. Dynamics measured by NMR at 10 °C suggest slow dynamics (ms-s) based around the C-terminal here shown as exchange between state A and B, and faster exchange (μ s-ms) around the binding region, which could be caused by a more frustrated B state.

with sections of the C terminus β -strand removed. NMR measurements could then be performed to see if the protein is now in the second, more disordered state. This technique might be difficult as it is likely to reduce the stability of the *PB6*-CSP. MD simulations could also be used but the slow exchange rate of the C-terminus unfolding means this process is out of the time-scale that most MD simulations can reach. A comparison of the dynamics of the protein when the simulation is started with the fifth β -strand bound and unbound, or even removed completely, might reveal if this is the case. The next goal would be an understanding of whether this process relates to the adaptation of protein activity to low temperatures. This could point towards methods of optimising enzyme activity to low temperatures in proteins used in biotechnology. Alternatively it could just be that these proteins are less stable because in cold environments the protein stability is not selected for.

One conclusion that can be taken from these results is the importance of the temperature at which the protein functions *in vivo*. When comparing proteins from different extremophilic organisms, features can be missed if they are compared at one temperature. The low temperature dynamics of the *PB6*-CSP would have been overlooked here if measurements were only recorded above 20 °C.

6.2 Implications on the molecular adaptation of CSPs to low temperature function

Results in chapter 3 and 5 demonstrate evidence of a correlation between the region of the protein associated with stability^[182] (the mechanical clamp which includes the N and C termini) and the active site^[62]. The binding of ssDNA was shown to affect the stability across the N and C termini, and the *PB6*-CSP low temperature dynamics found in this region underwent slow dynamics, with around 20 % of the 5 residues nearest the C-terminus populating the second state (figure 5.26). These dynamics could be a part of the adaptation of activity to lower energy levels as has been demonstrated in other extremophilic enzymes^[183]. As the nucleic acid binding site is highly conserved between CSP homologues, the adaptation of activity must come from changing the structure and dynamics surrounding this site. If the C-terminus is key for the protein stability, then it could be the easiest source of added flexibility when adapting to cold environments. The chain termini have previously been shown to be crucial for the extra stability of the *Bc*-CSP from *Bacillus caldolyticus* when compared to the *Bs*-CSP^[184]. Both SMFS measurements of the complex unfolding and the low temperature dynamics measured by NMR imply a correlation between this region and the function of ssDNA binding. Figure 6.3 shows two suggestions of how these unfolding dynamics affect the CSP function as an RNA chaperone. The second state could reduce the overall binding



Figure 6.3: Representations of how the low temperature exchange with a second state, possibly by the unfolding of the C-terminus, might influence the *PB6*-CSP (blue) binding to nucleic acids (red) and how it functions as a chaperone. (a) The second state might have a reduced affinity to nucleic acids so that the complex is not stabilised by the decrease in temperature, allowing the nucleic acid to unbind more easily. (b) The second state might aid in the unfolding of nucleic acid hairpins.

affinity of the *PB6*-CSP to nucleic acids as demonstrated in figure 6.3a. This could account for the difference in thermodynamics seen at 10 and 25 °C, measured using ITC (figure 4.9b). Tuning the affinity this way would prevent the ssDNA binding too tightly at low temperatures by slow conformational exchange with a state that does not favour tight binding^[124,185,186]. In figure 6.3b, the slow dynamics aid in the enzyme function by facilitating the unfolding of nucleic acid secondary structures. Disorder has been found to be a common attribute in RNA chaperones^[20,21], and it is possible that this helps loosen kinetically trapped RNA hairpins. It is possible that the reason both the *PB6*-CSP and *Bs*-CSP display increased dynamics at their respective cold-shock temperatures is that they sample an unfolded state so that when they bind RNA hairpins the complex can favour the unfolded conformation of the RNA. This could be explored further by implementing the NMR technique developed recently to track RNA hairpin unfolding^[63], and tracking the abilities of *PB6*-CSP at low temperatures when compared to the *Bs*-CSP.

6.3 SMFS to Confirm Localised Unfolding Identified in NMR Measurements

The possibility of slow unbinding and rebinding of the C-terminal β -sheet identified in these NMR relaxation measurements is something that SMFS is ideally suited to investigate. Mechanical unfolding of the *PB6*-CSP could reveal an alternative unfolding trajectory starting from the protein conformation with the C-terminal unbound from the rest of the *PB6*-CSP. This would cause an unfolding peak with a smaller ΔL_C as the protein is already partially unfolded in this state. It would likely also unfold at a lower force as the C-terminus β -sheet is part of the mechanical clamp that causes the relatively high F_U of CSPs.

If the measured unfolding forces of the partially unfolded *PB6*-CSP are measured to be significantly lower, then the SMFS could be used to differentiate which state the protein is in initially. If the presence of ssDNA stabilises the fully folded structure it

would decrease the number of low force and shorter ΔL_C unfolding peaks measured for the *PB6*-CSP at low temperatures. Alternatively if the partially unfolded state is the conformation that binds the ssDNA with higher affinity at low temperatures, the presence of ssDNA might decrease the measured unfolding force. Combining this technique with NMR experiments on the *PB6*-CSP in the presence of ssDNA would provide a way to compare the single molecule statistical results with those obtained by NMR on the relative populations.

6.4 Application to Protein Based Materials

Proteins are the machines that keep cells healthy, protected and adaptable. All are made up of the same limited set of amino acids. The variety of properties comes from the way these are ordered, according to instructions written into the DNA. If this bottom up approach could be exploited in material development, it could be possible to create a range of useful materials with biomedical applications^[187]. Such applications would require proteins that are mechanically stable enough to maintain their shape under the stress that would be involved in linking the sub-structures together. A material with proteins incorporated into the structure could be engineered with properties that respond to biological environments, such as the presence of ligands. As proteins are known to change conformation upon ligand binding, this material would exploit the wide range of protein functions to create a responsive bio-material. With mechanical properties that can be altered by the addition of a small molecule, the many useful roles that proteins play could be adapted to materials. It could also be applied as a way of carrying small molecules to a target, with controllable release mechanisms. The ability of a small material to store RNA or other drugs has been shown to have useful benefits in suppressing tumours^[188,189]. Nucleic acids could even be used as the linking regions, with highly controllable self assembly properties. The incorporation of proteins into these materials, shown to be possible for folded proteins^[190], would unlock a huge range of biotechnological tools.

Appendix A

Appendix Methods

Primer	Sequence
Forward	CGCCCAGAAAGGTGAAATCA
Reverse	GCTTCGCAACGTTCAAATCC

Table A.1: Primers used for the sequencing of the final pMal c5x plasmid containing the *PB6-CSP* gene

Protein	Amino Acid Sequence
His-tagged <i>Bs-CSP</i>	M H H H H H S S L E G K V K W F N S E K G F G F I E V E G Q D D V F V H F S A I Q G E G F K T L E E G Q A V S F E I V E G N R G P Q A A N V T K E A
<i>Bs-CSP</i> with TEVp cleavage site	M H H H H H H E N L Y F Q G S L E G K V K W F N S E K G F G F I E V E G Q D D V F V H F S A I Q G E G F K T L E E G Q A V S F E I V E G N R G P Q A A N V T K E A
Cys- <i>Bs-CSP</i>	M H H H H H S S L E G K V K W F N S E K G F G F I E V E G Q D D V F V H F C A I Q G E G F K T L E E G Q A V S F E I V E G N R G P Q A A N V T K E A

Table A.2: Amino Acid Sequences of *Bs-CSP* Variants used

Protein	Amino Acid Sequence
MBP- <i>PB6</i> -CSP construct	M H H H H H H K I E E G K L V I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W D A V R Y N G K L I A Y P I A V E A L S L I Y N K D L L P N P P K T W E E I P A L D K E L K A K G K S A L M F N L Q E P Y F T W P L I A A D G G Y A F K Y E N G K Y D I K D V G V D N A G A K A G L T F L V D L I K N K H M N A D T D Y S I A E A A F N K G E T A M T I N G P W A W S N I D T S K V N Y G V T V L P T F K G Q P S K P F V G V L S A G I N A A S P N K E L A K E F L E N Y L L T D E G L E A V N K D K P L G A V A L K S Y E E E L V K D P R I A A T M E N A Q K G E I M P N I P Q M S A F W Y A V R T A V I N A A S G R Q T V D E A L K D A Q T N S S S N N N N N N N N N L G I E G R I S H M E N L Y F Q G S D K V E G T V K W F N E A K G F G F I A Q D N G G Q D V F A H Y S A I Q G G G F K T L A E G Q K V S F I L G D G K K G P Q A E Q I E A I

Table A.3: Amino Acid Sequence of the MBP-*PB6*-CSP construct

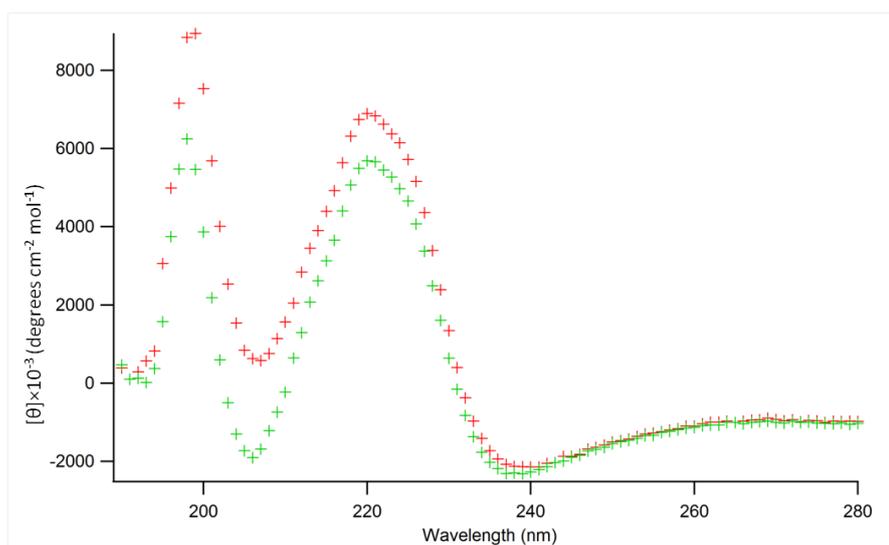


Figure A.1: CD scan of the Cys-*Bs*-CSP at room temperature to confirm folding of the variant.

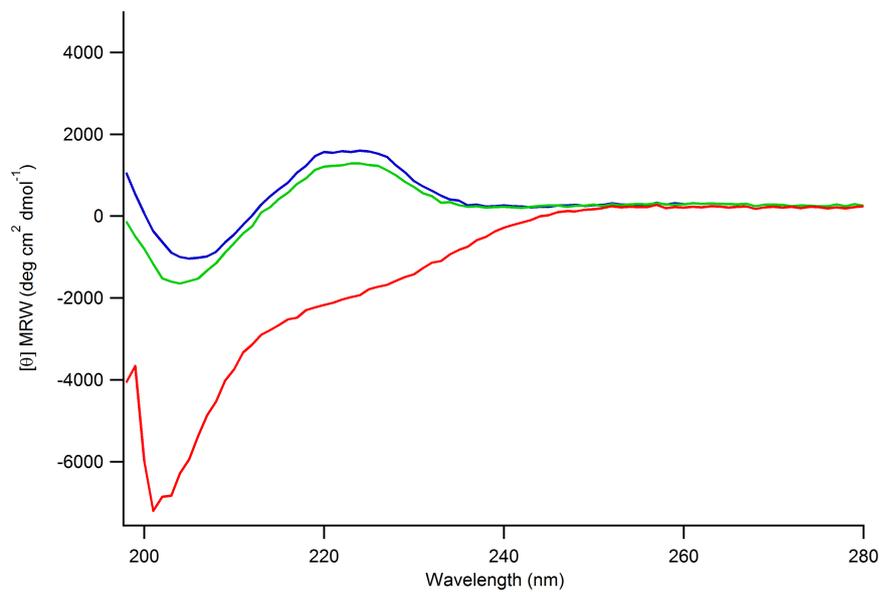


Figure A.2: CD scans of the *PB6*-CSP at start (green) of the temperature ramp, at 80 °C during the temperature ramp (red), and after returning to 7 °C after the temperature ramp (blue)

Appendix B

Appendix Results

Conditions	Linear Fit	<i>Bs-CSP</i>			I27		
		x_U (nm)	k_U s^{-1}	ΔG_U (kJmol $^{-1}$)	x_U (nm)	k_U s^{-1}	ΔG_U (kJmol $^{-1}$)
23 °C no ssDNA	Min	0.62	0.153	36.5	0.31	0.0003	50.9
	Fit	0.55	0.3	34.9	0.21	0.001	47.9
	Max	0.51	0.37	34.4	0.25	0.004	45
23 °C + CT2	Min	0.84	0.002	46.5	0.48	0.0000004	66.3
	Fit	0.7	0.01	42.8	0.39	0.00002	57.5
	Max	0.52	0.09	37.7	0.32	0.0002	51.9
5 °C no ssDNA	Min	0.59	0.016	41.7	0.23	0.0004	50.3
	Fit	0.52	0.05	39.0	0.21	0.0011	47.9
	Max	0.46	0.09	37.7	0.18	0.005	44.4
5 °C + CT2	Min	0.95	0.000004	60.96	0.24	0.0002	51.9
	Fit	0.78	0.0001	53.5	0.19	0.0027	45.8
	Max	0.72	0.0003	50.9	0.17	0.007	43.6

Table B.1: Parameters of the underlying energy landscapes calculated by Monte Carlo Simulations from experiments at 23 °C and 5 °C both with and without ssDNA present. Errors are estimated by the maximum and minimum linear fits that describe the data.

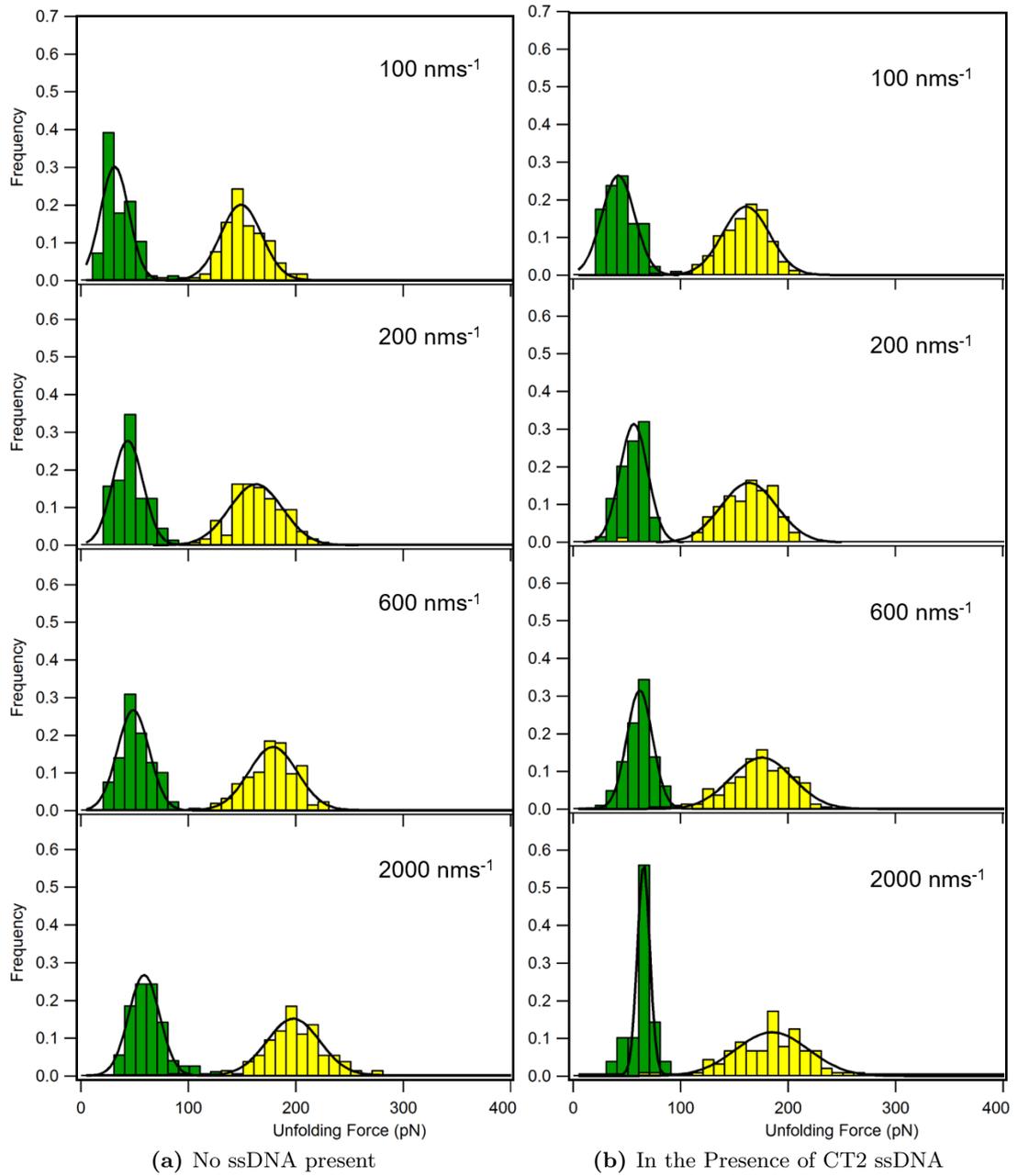


Figure B.1: Histograms of all the unfolding forces recorded at room temperature on the $(I27-BsCSP)_3-I27$ polyprotein.

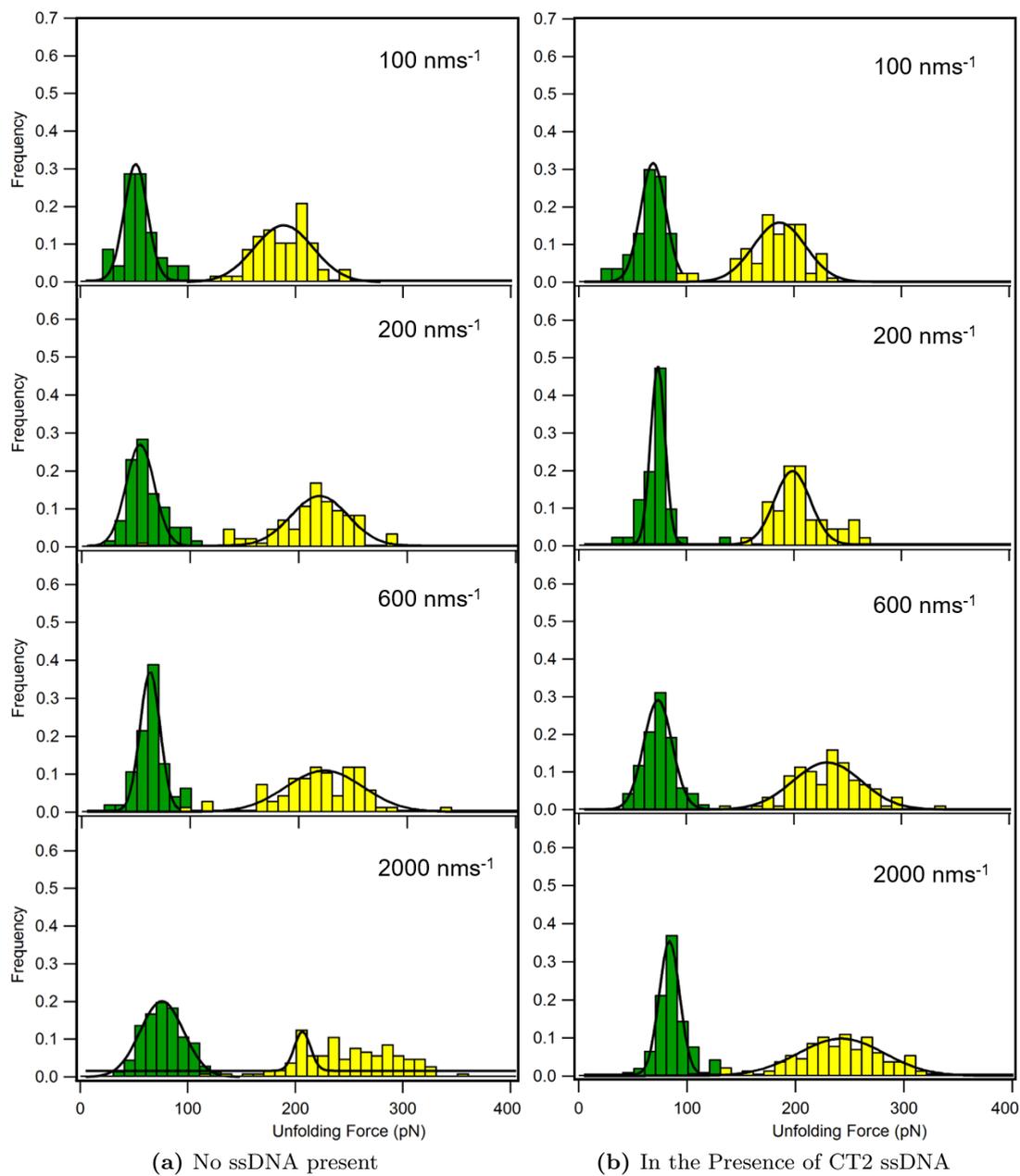


Figure B.2: Histograms of all the unfolding forces recorded at 5 °C on the $(I27-BsCSP)_3-I27$ polyprotein.

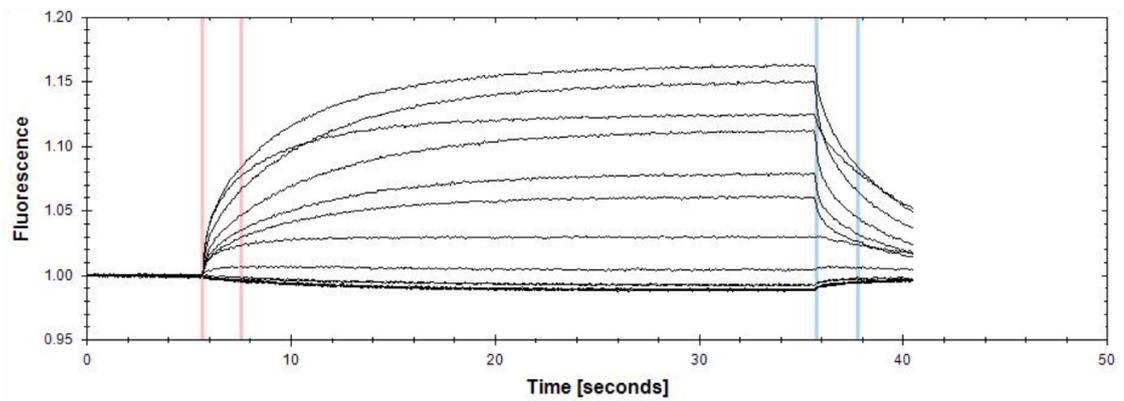


Figure B.3: Raw fluorescence signal over time when the IR laser is switched on (red lines) until shortly after it is switched off (blue lines) across all concentrations of *Bs*-CSP . The change in fluorescence corresponds to the concentration change in the spot that is heated by the IR laser, here at 20 % power.

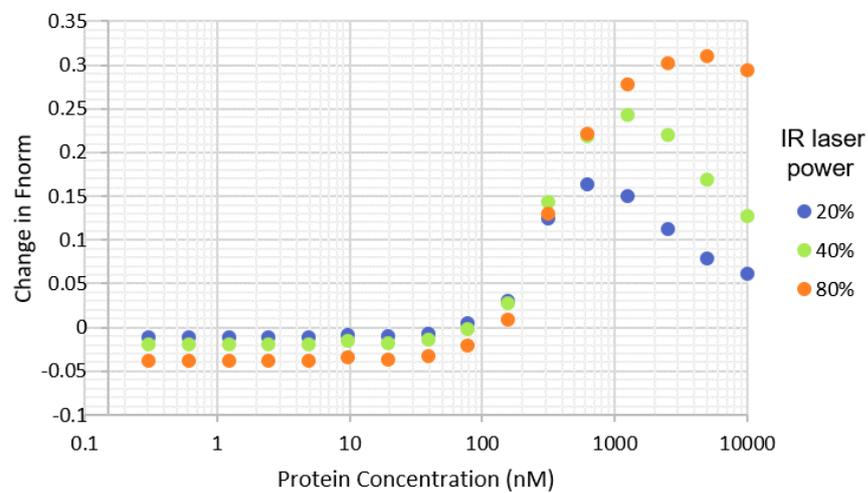


Figure B.4: Fractional change in fluorescence when the IR laser is switched on plotted against concentration of *Bs*-CSP . The change in fluorescence corresponds to the concentration change in the spot that is heated by the IR laser. Includes 3 repeats carried out at various IR laser strengths.

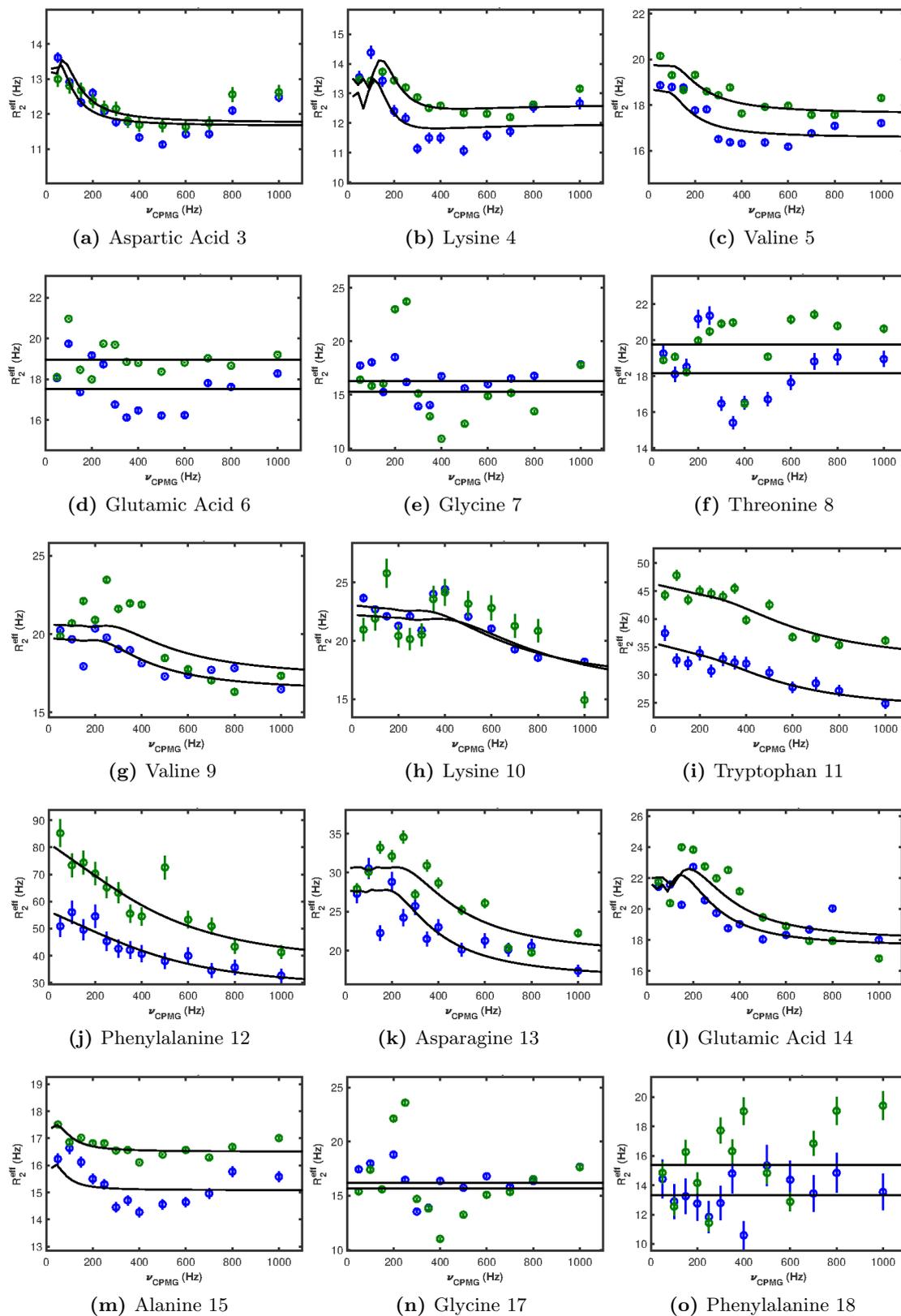


Figure B.5: First 15 relaxation curves

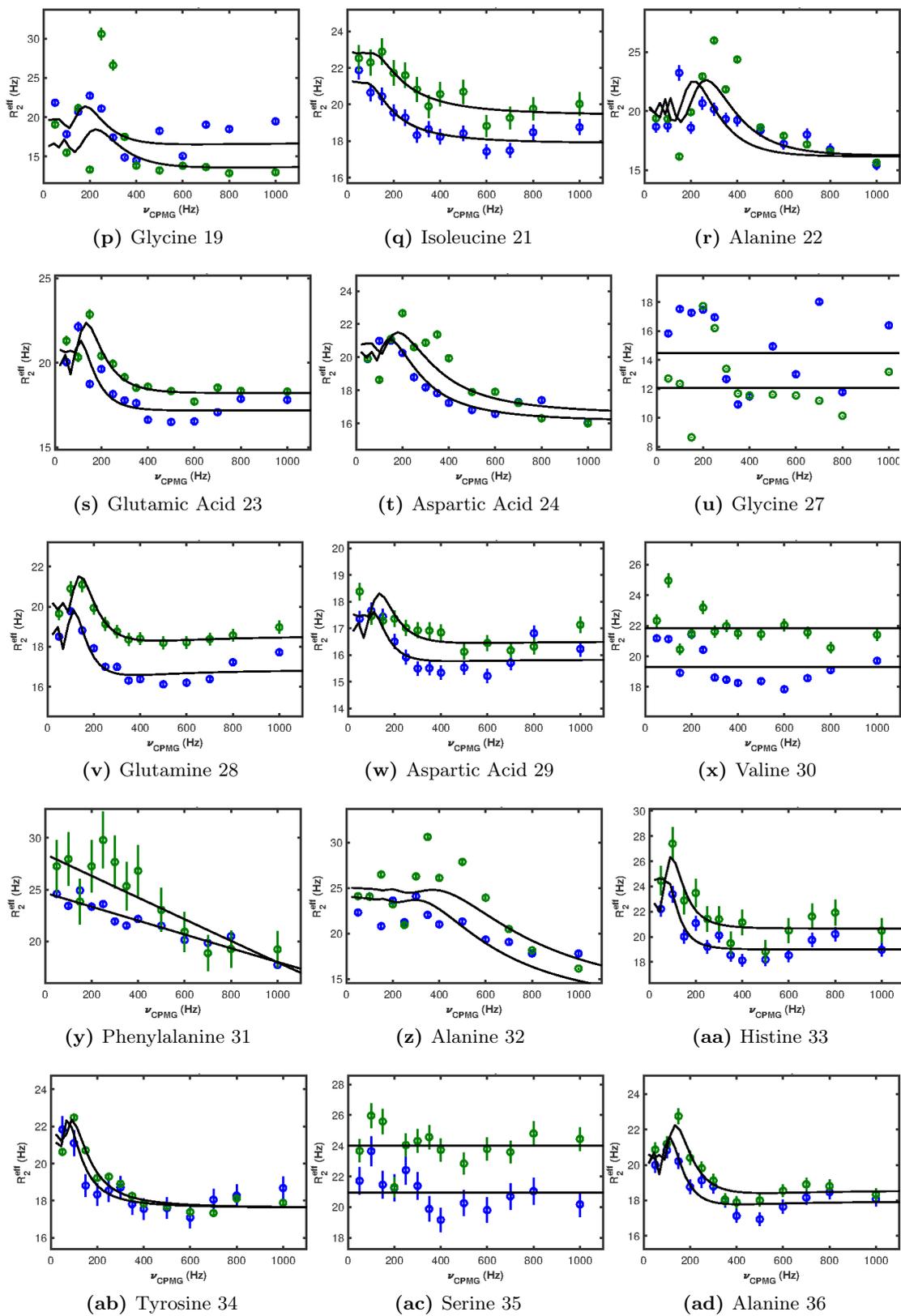


Figure B.5: Second group of 15 relaxation curves

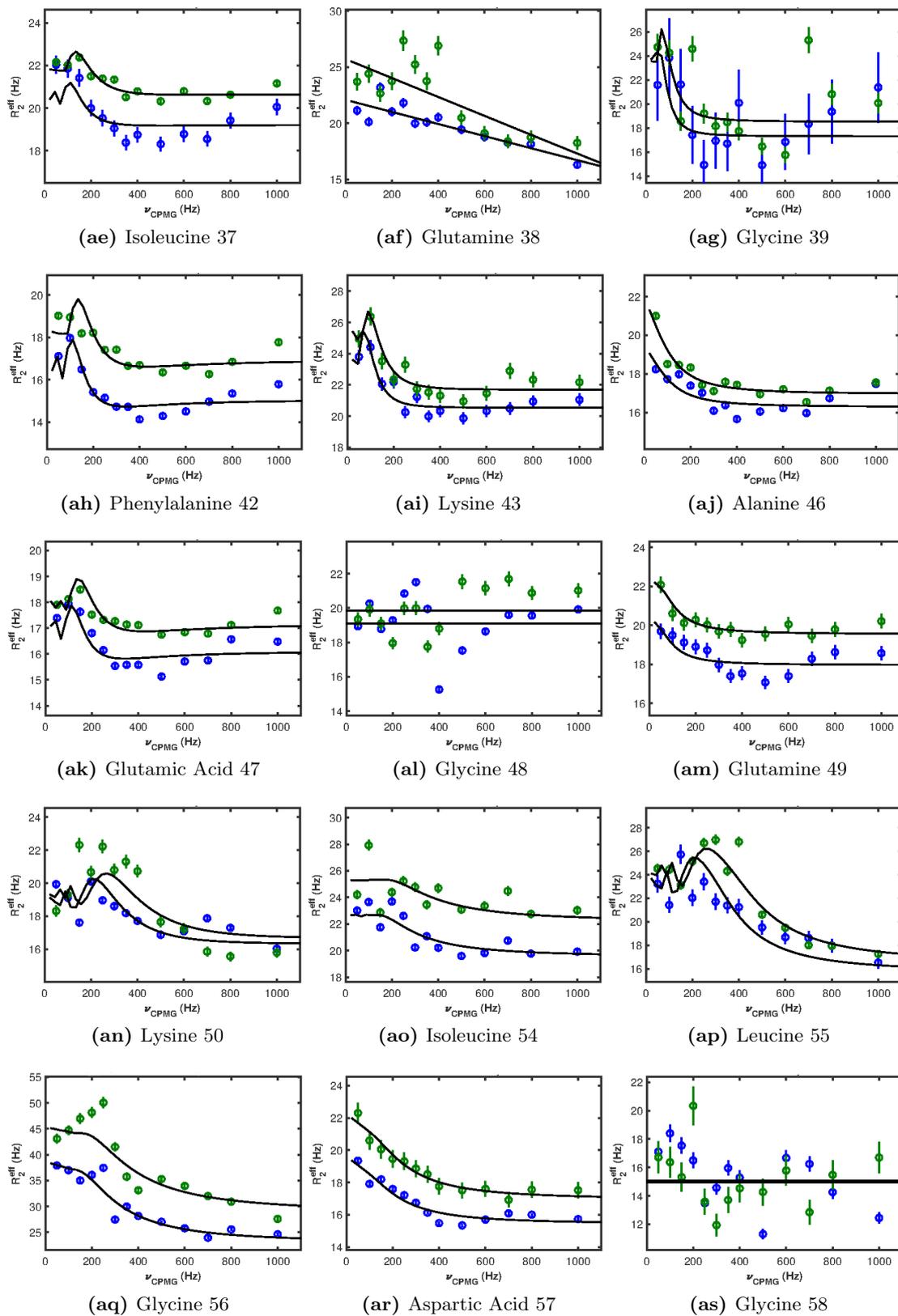


Figure B.5: Third group of 15 relaxation curves

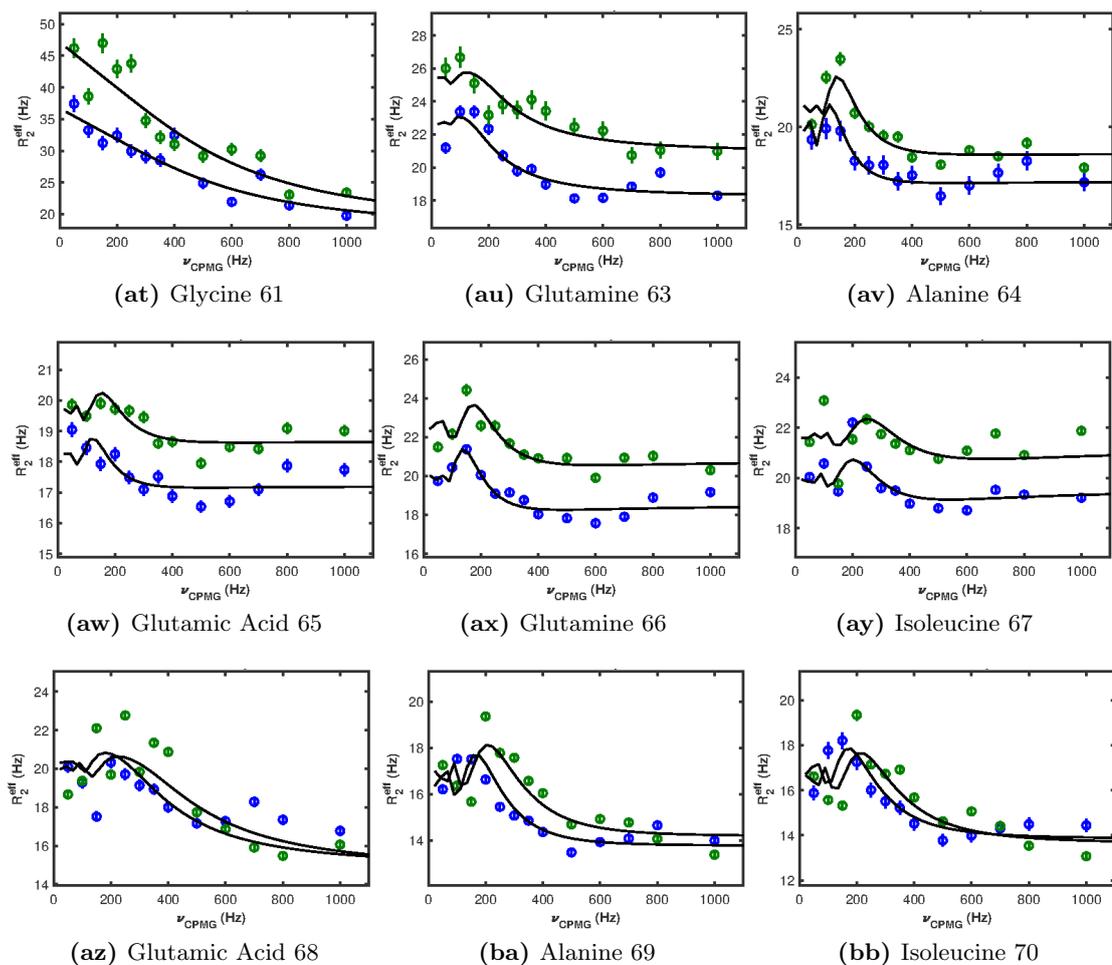


Figure B.5: Relaxation curves and R_{ex} Fits (black) from the CPMG experiment on the *PB6*-CSP at 10 °C measured at 600 (blue) and 750 (green) MHz field strengths.

Bibliography

- [1] Emmanuelle J. Javaux. Extreme life on Earth – past, present and possibly beyond. *Research in Microbiology*, 157(1):37–48, 2006. doi: 10.1016/J.RESMIC.2005.07.008. URL <https://doi.org/10.1016/j.resmic.2005.07.008>.
- [2] Debora Frigi Rodrigues and James M Tiedje. Coping with our cold planet. *Applied and Environmental Microbiology*, 74(6):1677–86, 2008. doi: 10.1128/AEM.02000-07. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2268296/>.
- [3] E M Rivkina, E I Friedmann, C P McKay, and D A Gilichinsky. Metabolic activity of permafrost bacteria below the freezing point. *Applied and environmental microbiology*, 66(8):3230–3, 2000. doi: 10.1128/AEM.66.8.3230-3233.2000. URL <http://www.ncbi.nlm.nih.gov/pubmed/10919774>.
- [4] Michelle J Chua, Richard L Campen, Lindsay Wahl, Joseph J Grzymiski, and Jill A Mikucki. Genomic and physiological characterization and description of *Marinobacter gelidimuriae* sp. nov., a psychrophilic, moderate halophile from Blood Falls, an antarctic subglacial brine. *FEMS Microbiology Ecology*, 94(3), 2018. doi: 10.1093/femsec/fiy021. URL <https://doi.org/10.1093/femsec/fiy021>.
- [5] Brent C. Christner, John C. Priscu, Amanda M. Achberger, Carlo Barbante, Sasha P. Carter, Knut Christianson, Alexander B. Michaud, Jill A. Mikucki, Andrew C. Mitchell, Mark L. Skidmore, Trista J. Vick-Majors, the WISSARD Science Team, W. P. Adkins, S. Anandakrishnan, G. Barcheck, L. Beem, A. Behar, M. Beitch, R. Bolsey, C. Branecky, R. Edwards, A. Fisher, H. A. Fricker, N. Foley, B. Guthrie, T. Hodson, H. Horgan, R. Jacobel, S. Kelley, K. D. Mankoff, E. McBryan, R. Powell, A. Purcell, D. Sampson, R. Scherer, J. Sherve, M. Siegfried, and S. Tulaczyk. A microbial ecosystem beneath the West Antarctic ice sheet. *Nature*, 512(7514):310–313, 2014. doi: 10.1038/nature13667. URL <http://www.nature.com/articles/nature13667>.
- [6] Georges Feller and Charles Gerday. Psychrophilic enzymes: hot topics in cold adaptation. *Nature reviews. Microbiology*, 1(3):200–8, 2003. doi: 10.1038/nrmicro773. URL <http://www.ncbi.nlm.nih.gov/pubmed/15035024>.
- [7] Katarzyna M. Tych, Toni Hoffmann, David J. Brockwell, and Lorna Dougan. Single molecule force spectroscopy reveals the temperature-dependent robustness and malleability of a hyperthermophilic protein. *Soft Matter*, 9(37):9016, 2013. doi: 10.1039/c3sm51439k. URL <http://pubs.rsc.org/en/Content/ArticleHTML/2013/SM/C3SM51439K>.
- [8] John G. Stockner. Observations of Thermophilic Algal Communities in Mount Rainier and Yellowstone National Parks. *Limnology and Oceanography*, 12(1):13–

- 17, 1967. doi: 10.4319/lo.1967.12.1.0013. URL <https://doi.org/10.4319/lo.1967.12.1.0013>.
- [9] Michael P Robertson and Gerald F Joyce. The origins of the RNA world. *Cold Spring Harbor perspectives in biology*, 4(5), 2012. doi: 10.1101/cshperspect.a003608. URL <http://www.ncbi.nlm.nih.gov/pubmed/20739415>.
- [10] H. M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235. URL <https://doi.org/10.1093/nar/28.1.235>.
- [11] John Tooze Carl Ivar Branden. *Introduction to Protein Structure*. Garland Pub, 1999. ISBN 9780815323051.
- [12] G B McGaughey, M Gagné, and A K Rappé. pi-Stacking interactions. Alive and well in proteins. *The Journal of biological chemistry*, 273(25):15458–63, 1998. URL <http://www.ncbi.nlm.nih.gov/pubmed/9624131>.
- [13] S Improta, A S Politou, and A Pastore. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure (London, England : 1993)*, 4(3):323–37, 1996. URL <http://www.ncbi.nlm.nih.gov/pubmed/8805538>.
- [14] Patrick G Telmer and Brian H Shilton. Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *The Journal of biological chemistry*, 278(36):34555–67, 2003. doi: 10.1074/jbc.M301004200. URL <http://www.ncbi.nlm.nih.gov/pubmed/12794084>.
- [15] Cristiano L. Dias, Tapio Ala-Nissila, Jirasak Wong-ekkabut, Ilpo Vattulainen, Martin Grant, and Mikko Karttunen. The hydrophobic effect and its role in cold denaturation, 2010. URL <https://doi.org/10.1016/j.cryobiol.2009.07.005>.
- [16] P L Privalov. Cold Denaturation of Protein. *Critical Reviews in Biochemistry and Molecular Biology*, 25(5):281–306, 1990. doi: 10.3109/10409239009090613. URL <http://www.ncbi.nlm.nih.gov/pubmed/2225910>.
- [17] Katarzyna M Tych, Matthew Batchelor, Toni Hoffmann, Michael C Wilson, Emanuele Paci, David J Brockwell, and Lorna Dougan. Tuning protein mechanics through an ionic cluster graft from an extremophilic protein. *Soft matter*, 12(10):2688–2699, 2016. doi: 10.1039/c5sm02938d. URL <http://www.ncbi.nlm.nih.gov/pubmed/26809452>.
- [18] Monika Davidovic, Carlos Mattea, Johan Qvist, and Bertil Halle. Protein Cold Denaturation as Seen From the Solvent. *Journal of the American Chemical Society*, 131(3):1025–1036, 2009. doi: 10.1021/ja8056419. URL <http://www.ncbi.nlm.nih.gov/pubmed/19115852>.
- [19] D Thirumalai and Changbong Hyeon. RNA and protein folding: common themes and variations. *Biochemistry*, 44(13):4957–70, 2005. doi: 10.1021/bi047314+. URL <http://pubs.acs.org/doi/full/10.1021/bi047314%}2B>.
- [20] Peter Tompa and Peter Csermely. The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal*, 18(11):1169–1175, 2004. doi: 10.1096/fj.04-1584rev. URL <http://www.fasebj.org/doi/10.1096/fj.04-1584rev>.

- [21] Elan Z. Eisenmesser, Oscar Millet, Wladimir Labeikovskiy, Dmitry M. Korzhnev, Magnus Wolf-Watz, Daryl A. Bosco, Jack J. Skalicky, Lewis E. Kay, and Dorothee Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, 2005. doi: 10.1038/nature04105. URL <http://www.nature.com/articles/nature04105>.
- [22] Franz Narberhaus, Torsten Waldminghaus, and Saheli Chowdhury. RNA thermometers. *FEMS Microbiology Reviews*, 30(1):3–16, 2006. doi: 10.1111/j.1574-6976.2005.004.x. URL <https://doi.org/10.1111/j.1574-6976.2005.004.x>.
- [23] Anna Maria Giuliodori, Fabio Di Pietro, Stefano Marzi, Benoit Masquida, Rolf Wagner, Pascale Romby, Claudio O Gualerzi, and Cynthia L Pon. The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Molecular cell*, 37(1):21–33, 2010. doi: 10.1016/j.molcel.2009.11.033. URL <https://doi.org/10.1016/j.molcel.2009.11.033>.
- [24] I. Chen Hung, Yu Chuan Hsiao, H. Sunny Sun, Tsung Ming Chen, and Shyh Jye Lee. MicroRNAs regulate gene plasticity during cold shock in zebrafish larvae. *BMC Genomics*, 17(1):922, 2016. doi: 10.1186/s12864-016-3239-4. URL <https://doi.org/10.1186/s12864-016-3239-4>.
- [25] R Jaenicke and G Böhm. The stability of proteins in extreme environments. *Current opinion in structural biology*, 8(6):738–48, 1998. URL <http://www.ncbi.nlm.nih.gov/pubmed/9914256>.
- [26] Fabrizio Pucci and Marianne Rooman. Physical and molecular bases of protein thermal stability and cold adaptation. *Current Opinion in Structural Biology*, 42: 117–128, 2017. doi: 10.1016/j.sbi.2016.12.007. URL <https://doi.org/10.1016/j.sbi.2016.12.007>.
- [27] P Závodszky, J Kardos, Svingor, and G A Petsko. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 95(13):7406–11, 1998. URL <http://www.ncbi.nlm.nih.gov/pubmed/9636162>.
- [28] H Robinson, Y G Gao, B S McCrary, S P Edmondson, J W Shriver, and A H Wang. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature*, 392(6672):202–5, 1998. doi: 10.1038/32455. URL <http://www.ncbi.nlm.nih.gov/pubmed/9515968>.
- [29] Homa Ghalei, Holger von Moeller, Detlef Eppers, Daniel Sohmen, Daniel N Wilson, Bernhard Loll, and Markus C Wahl. Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein. *Nucleic Acids Research*, 42(10): 6698–708, 2014. doi: 10.1093/nar/gku259. URL <https://doi.org/10.1093/nar/gku259>.
- [30] Georges Feller. Cryosphere and Psychrophiles: Insights into a Cold Origin of Life? *Life (Basel, Switzerland)*, 7(2), 2017. doi: 10.3390/life7020025. URL <http://www.ncbi.nlm.nih.gov/pubmed/28604605>.
- [31] Georges Feller. Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica*, 2013:512840, 2013. doi: 10.1155/2013/512840. URL <http://www.ncbi.nlm.nih.gov/pubmed/24278781>.

- [32] Françoise Berger, Nicolas Morellet, Frédéric Menu, and Patrick Potier. Cold shock and cold acclimation proteins in the psychrotrophic bacterium *Arthrobacter globiformis* SI55. *Journal of Bacteriology*, 178(11):2999–3007, 1996. doi: 10.1128/jb.178.11.2999-3007.1996. URL <http://www.ncbi.nlm.nih.gov/pubmed/8655472>.
- [33] Karolina L. Tkaczuk¹ Agnieszka Kaufman-Szymczyk, Arkadiusz Wojtasik, Paweł Parniewski², Aneta Białkowska¹, Marianna Turkiewicz, Agnieszka Kaufman-Szymczyk, Arkadiusz Wojtasik, Paweł Parniewski, Aneta Białkowska, Karolina Tkaczuk, and Marianna Turkiewicz. Identification of the *csp* gene and molecular modelling of the CspA-like protein from Antarctic soil-dwelling psychrotrophic bacterium *Psychrobacter* sp. B6. *Acta biochimica Polonica*, 56(1):63–9, 2009. URL <http://www.ncbi.nlm.nih.gov/pubmed/19238258>.
- [34] Anna Bujacz, Maria Rutkiewicz-Krotewicz, Karolina Nowakowska-Sapota, and Marianna Turkiewicz. Crystal structure and enzymatic properties of a broad substrate-specificity psychrophilic aminotransferase from the Antarctic soil bacterium *Psychrobacter* sp. B6. *Acta Crystallographica Section D Biological Crystallography*, 71(3):632–645, 2015. doi: 10.1107/S1399004714028016. URL <https://doi.org/10.1107/S1399004714028016>.
- [35] Pieter De Maayer, Dominique Anderson, Craig Cary, and Don A Cowan. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO reports*, 15(5):508–17, 2014. doi: 10.1002/embr.201338170. URL <http://www.ncbi.nlm.nih.gov/pubmed/24671034>.
- [36] Geir Villy Isaksen, Johan Åqvist, and Bjørn Olav Brandsdal. Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proceedings of the National Academy of Sciences*, 113(28):7822–7827, 2016. doi: 10.1073/pnas.1605237113. URL <http://www.ncbi.nlm.nih.gov/pubmed/27354533>.
- [37] Marc W. van der Kamp, Erica J. Prentice, Kirsty L. Kraakman, Michael Connolly, Adrian J. Mulholland, and Vickery L. Arcus. Dynamical origins of heat capacity changes in enzyme catalysed reactions. *Nature communications*, 9(1):165324, 2017. doi: 10.1101/165324. URL <http://www.nature.com/articles/s41467-018-03597-y>.
- [38] Tim J. Kamerzell and C. Russell Middaugh. The Complex Inter-Relationships Between Protein Flexibility and Stability. *Journal of Pharmaceutical Sciences*, 97(9):3494–3517, 2008. doi: 10.1002/jps.21269. URL <https://www.ncbi.nlm.nih.gov/pubmed/18186490>.
- [39] James C.A. Bardwell and Ursula Jakob. Conditional disorder in chaperone action. *Trends in Biochemical Sciences*, 37(12):517–525, 2012. doi: 10.1016/j.tibs.2012.08.006. URL <http://www.ncbi.nlm.nih.gov/pubmed/23018052>.
- [40] Youn Hong Jung, Ji-hyun Uh, Kyunghye Lee, and Hana Im. An intrinsically disordered domain in *Polaribacter irgensii* KOPRI 22228 CspB confers extraordinary freeze-tolerance. *Biochemical and Biophysical Research Communications*, 2018. doi: 10.1016/j.bbrc.2018.01.050. URL <https://doi.org/10.1016/j.bbrc.2018.01.050>.
- [41] Salvino D’Amico, Tony Collins, Jean-Claude Marx, Georges Feller, and Charles Gerday. Psychrophilic microorganisms: challenges for life. *EMBO reports*, 7(4):385–389, 2006. doi: 10.1038/sj.embor.7400662. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1456908/>.

- [42] Harry G. Saavedra, James O. Wrabl, Jeremy A. Anderson, Jing Li, and Vincent J. Hilser. Dynamic allostery can drive cold adaptation in enzymes. *Nature*, page 1, 2018. doi: 10.1038/s41586-018-0183-2. URL <http://www.nature.com/articles/s41586-018-0183-2>.
- [43] Kathryn J Boor. Bacterial stress responses: What doesn't kill them can make them stronger, 2006. URL <http://dx.plos.org/10.1371/journal.pbio.0040023>.
- [44] G Horn, R Hofweber, W Kremer, and H R Kalbitzer. Structure and function of bacterial cold shock proteins. *Cellular and Molecular Life Sciences : CMLS*, 64(12):1457–70, 2007. doi: 10.1007/s00018-007-6388-4. URL <http://www.ncbi.nlm.nih.gov/pubmed/17437059>.
- [45] Anna Brandi, Paola Pietroni, Claudio O. Gualerzi, and Cynthia L. Pon. Post-transcriptional regulation of CspA expression in Escherichia coli. *Molecular Microbiology*, 19(2):231–240, 1996. doi: 10.1046/j.1365-2958.1996.362897.x. URL <http://doi.wiley.com/10.1046/j.1365-2958.1996.362897.x>.
- [46] Peter Graumann and Mohamed A. Marahiel. Some like it cold: Response of microorganisms to cold shock, 1996. URL <https://www.ncbi.nlm.nih.gov/pubmed/8929274>.
- [47] J P Etchegaray and M Inouye. CspA, CspB, and CspG, major cold shock proteins of Escherichia coli, are induced at low temperature under conditions that completely block protein synthesis. *Journal of bacteriology*, 181(6):1827–30, 1999. URL <http://www.ncbi.nlm.nih.gov/pubmed/10074075>.
- [48] Barbara Schmid, Jochen Klumpp, Eveline Raimann, Martin J Loessner, Roger Stephan, and Taurai Tasara. Role of cold shock proteins in growth of Listeria monocytogenes under cold and osmotic stress conditions. *Applied and Environmental Microbiology*, 75(6):1621–1627, 2009. doi: 10.1128/AEM.02154-08. URL <http://www.ncbi.nlm.nih.gov/pubmed/19151183>.
- [49] P Graumann, T M Wendrich, M H Weber, K Schröder, and M A Marahiel. A family of cold shock proteins in Bacillus subtilis is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures. *Molecular microbiology*, 25(4):741–56, 1997. URL <http://www.ncbi.nlm.nih.gov/pubmed/9379903>.
- [50] S Fabio Falsone, Michael Weichel, Reto Cramer, Michael Breitenbach, and Andreas J Kungl. Unfolding and double-stranded DNA binding of the cold shock protein homologue Cla h 8 from Cladosporium herbarum. *The Journal of Biological Chemistry*, 277(19):16512–6, 2002. doi: 10.1074/jbc.M200833200. URL <http://www.jbc.org/content/277/19/16512.full>.
- [51] Yumiko Tanabe, Satoru Nagatoishi, Kouhei Tsumoto, Satoru Nagatoishi, and Yumiko Tanabe. Thermodynamic characterization of the interaction between human Y-box binding protein YB-1 and nucleic acids. *Mol. BioSyst.*, 11(9):2441–2448, 2015. doi: 10.1039/C5MB00184F. URL <http://pubs.rsc.org/en/Content/ArticleLanding/2015/MB/C5MB00184F>.
- [52] Dieter Perl, Christine Welker, Thomas Schindler, Katja Schröder, Mohamed A. Marahiel, Rainer Jaenicke, and Franz X. Schmid. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins.

- Nature Structural Biology*, 5(3):229–235, 1998. doi: 10.1038/nsb0398-229. URL <http://dx.doi.org/10.1038/nsb0398-229>.
- [53] H Schindelin, M A Marahiel, and U Heinemann. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature*, 364(6433):164–8, 1993. doi: 10.1038/364164a0. URL <http://dx.doi.org/10.1038/364164a0>.
- [54] Michael J Morten, Jose R Peregrina, Maria Figueira-Gonzalez, Katrin Ackermann, Bela E Bode, Malcolm F White, and J Carlos Penedo. Binding dynamics of a monomeric SSB protein to DNA: a single-molecule multi-process approach. *Nucleic acids research*, 43(22):10907–24, 2015. doi: 10.1093/nar/gkv1225. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4678828>.
- [55] Nan Wang, Kunitoshi Yamanaka, and Masayori Inouye. Acquisition of double-stranded DNA-binding ability in a hybrid protein between *Escherichia coli* CspA and the cold shock domain of human YB-1. *Molecular Microbiology*, 38(3):526–534, 2000. doi: 10.1046/j.1365-2958.2000.02146.x. URL <http://doi.wiley.com/10.1046/j.1365-2958.2000.02146.x>.
- [56] Yan Zhang, David H. Burkhardt, Silvi Rouskin, Gene Wei Li, Jonathan S. Weissman, and Carol A. Gross. A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation, 2018. URL <https://www.sciencedirect.com/science/article/pii/S1097276518301801>.
- [57] Klaas E A Max, Markus Zeeb, Ralf Bienert, Jochen Balbach, and Udo Heinemann. Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from *Bacillus caldolyticus*. *The FEBS Journal*, 274(5):1265–79, 2007. doi: 10.1111/j.1742-4658.2007.05672.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/17266726>.
- [58] Rolf Sachs, Klaas E A Max, Udo Heinemann, and Jochen Balbach. RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA (New York, N.Y.)*, 18(1):65–76, 2012. doi: 10.1261/rna.02809212. URL <http://rnajournal.cshlp.org/content/18/1/65.full>.
- [59] Markus Zeeb, Klaas E A Max, Ulrich Weininger, Christian Löw, Heinrich Sticht, and Jochen Balbach. Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution. *Nucleic Acids Research*, 34(16):4561–71, 2006. doi: 10.1093/nar/gkl376. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1636342>.
- [60] T Schindler, D Perl, P Graumann, V Sieber, M A Marahiel, and F X Schmid. Surface-exposed phenylalanines in the RNP1/RNP2 motif stabilize the cold-shock protein CspB from *Bacillus subtilis*. *Proteins*, 30(4):401–6, 1998. URL <http://www.ncbi.nlm.nih.gov/pubmed/9533624>.
- [61] M M Lopez, K Yutani, and G I Makhatadze. Interactions of the cold shock protein CspB from *Bacillus subtilis* with single-stranded DNA. Importance of the T base content and position within the template. *The Journal of Biological Chemistry*, 276(18):15511–8, 2001. doi: 10.1074/jbc.M010474200. URL <http://www.jbc.org/content/276/18/15511.full>.

- [62] Klaas E A Max, Markus Zeeb, Ralf Bienert, Jochen Balbach, and Udo Heinemann. T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. *Journal of Molecular Biology*, 360(3):702–14, 2006. doi: 10.1016/j.jmb.2006.05.044. URL <http://www.sciencedirect.com/science/article/pii/S0022283606006449>.
- [63] Enrico Rennella, Tomáš Sára, Michael Juen, Christoph Wunderlich, Lionel Imbert, Zsofia Solyom, Adrien Favier, Isabel Ayala, Katharina Weinhäupl, Paul Schanda, Robert Konrat, Christoph Kreutz, and Bernhard Brutscher. RNA binding and chaperone activity of the E. coli cold-shock protein CspA. *Nucleic Acids Research*, 45(7):gkx044, 2017. doi: 10.1093/nar/gkx044. URL <https://doi.org/10.1093/nar/gkx044>.
- [64] Tai Fei Yu, Zhao Shi Xu, Jin Kao Guo, Yan Xia Wang, Brian Abernathy, Jin Dong Fu, Xiao Chen, Yong Bin Zhou, Ming Chen, Xing Guo Ye, and You Zhi Ma. Improved drought tolerance in wheat plants overexpressing a synthetic bacterial cold shock protein gene SeCspA. *Scientific Reports*, 7:44050, 2017. doi: 10.1038/srep44050. URL <http://www.nature.com/articles/srep44050>.
- [65] A Brandi. Massive presence of the Escherichia coli ‘major cold-shock protein’ CspA under non-stress conditions. *The EMBO Journal*, 18(6):1653–1659, 1999. doi: 10.1093/emboj/18.6.1653. URL <http://emboj.embopress.org/cgi/doi/10.1093/emboj/18.6.1653>.
- [66] M H Weber, A V Volkov, I Fricke, M A Marahiel, and P L Graumann. Localization of cold shock proteins to cytosolic spaces surrounding nucleoids in Bacillus subtilis depends on active transcription. *Journal of Bacteriology*, 183(21):6435–43, 2001. doi: 10.1128/JB.183.21.6435-6443.2001. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=100140>.
- [67] Werner Kremer, Benjamin Schuler, Stefan Harrieder, Matthias Geyer, Wolfram Gronwald, Christine Welker, Rainer Jaenicke, and Hans R. Kalbitzer. Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima. *European Journal of Biochemistry*, 268(9):2527–2539, 2001. doi: 10.1046/j.1432-1327.2001.02127.x. URL <http://doi.wiley.com/10.1046/j.1432-1327.2001.02127.x>.
- [68] Juho Lee, Ki-Woong Jeong, Bonghwan Jin, Kyoung-Seok Ryu, Eun-Hee Kim, Joong-Hoon Ahn, and Yangmee Kim. Structural and dynamic features of cold-shock proteins of Listeria monocytogenes, a psychrophilic bacterium. *Biochemistry*, 52(14):2492–504, 2013. doi: 10.1021/bi301641b. URL <http://dx.doi.org/10.1021/bi301641b>.
- [69] Yeongjoon Lee, Chulhee Kwak, Ki Woong Jeong, Prasannavenkatesh Durai, Kyoung Seok Ryu, Eun Hee Kim, Chaejoon Cheong, Hee Chul Ahn, Hak Jun Kim, and Yangmee Kim. Tyr51: Key Determinant of the Low Thermostability of Colwellia psychrerythraea Cold Shock Protein. *Biochemistry*, page acs.biochem.8b00144, 2018. doi: 10.1021/acs.biochem.8b00144. URL <http://pubs.acs.org/doi/10.1021/acs.biochem.8b00144>.
- [70] Astrid Jung, Christian Bamann, Werner Kremer, Hans Robert Kalbitzer, and Eike Brunner. High-temperature solution NMR structure of TmCsp. *Protein Science : a publication of the Protein Society*, 13(2):342–50, 2004. doi: 10.1110/

ps.03281604. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2286716>.

- [71] Jörg Schönfelder, David De Sancho, and Raul Perez-Jimenez. The Power of Force: Insights into the Protein Folding Process Using Single-Molecule Force Spectroscopy. *Journal of Molecular Biology*, 428(21):4245–4257, 2016. doi: 10.1016/j.jmb.2016.09.006. URL <https://www.sciencedirect.com/science/article/pii/S0022283616303709?via=ihub#f0020>.
- [72] Keir C Neuman and Attila Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, 5(6):491–505, 2008. doi: 10.1038/nmeth.1218. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3397402>.
- [73] Iddo Heller, Tjalle P Hoekstra, Graeme A King, Erwin J G Peterman, and Gijs J L Wuite. Optical tweezers analysis of DNA-protein complexes. *Chemical Reviews*, 114(6):3087–119, 2014. doi: 10.1021/cr4003006. URL <http://dx.doi.org/10.1021/cr4003006>.
- [74] John S Graham, Reid C Johnson, and John F Marko. Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic acids research*, 39(6):2249–59, 2011. doi: 10.1093/nar/gkq1140. URL <https://doi.org/10.1093/nar/gkq1140>.
- [75] Toshio Ando, Takayuki Uchihashi, Noriyuki Kodera, Daisuke Yamamoto, Atsushi Miyagi, Masaaki Taniguchi, and Hayato Yamashita. High-speed AFM and nano-visualization of biomolecular processes. *Pflügers Archiv : European journal of physiology*, 456(1):211–25, 2008. doi: 10.1007/s00424-007-0406-0. URL <http://www.ncbi.nlm.nih.gov/pubmed/18157545>.
- [76] Yves F Dufrêne and Peter Hinterdorfer. Recent progress in AFM molecular recognition studies. *Pflügers Archiv : European journal of physiology*, 456(1):237–45, 2008. doi: 10.1007/s00424-007-0413-1. URL <http://www.ncbi.nlm.nih.gov/pubmed/18157727>.
- [77] G. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978. doi: 10.1126/science.347575. URL <http://www.sciencemag.org/content/200/4342/618.abstract>.
- [78] Toni Hoffmann and Lorna Dougan. Single molecule force spectroscopy using polyproteins. *Chemical Society Reviews*, 41(14):4781, 2012. doi: 10.1039/c2cs35033e. URL <http://xlink.rsc.org/?DOI=c2cs35033e>.
- [79] C. Hyeon and D. Thirumalai. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proceedings of the National Academy of Sciences*, 100(18):10249–10253, 2003. doi: 10.1073/pnas.1833310100. URL <http://www.pnas.org/content/100/18/10249.full>.
- [80] Changbong Hyeon, Michael Hinczewski, and D Thirumalai. Evidence of disorder in biological molecules from single molecule pulling experiments. *Physical Review Letters*, 112(13):138101, 2014. doi: 10.1103/PhysRevLett.112.138101. URL <https://doi.org/10.1103/PhysRevLett.112.138101>.

- [81] Changbong Hyeon and D Thirumalai. Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments. *Journal of Physics: Condensed Matter*, 19(11):113101, 2007. doi: 10.1088/0953-8984/19/11/113101. URL <http://iopscience.iop.org/article/10.1088/0953-8984/19/11/113101>.
- [82] Reinat Nevo, Vlad Brumfeld, Ruti Kapon, Peter Hinterdorfer, and Ziv Reich. Direct measurement of protein energy landscape roughness. *EMBO reports*, 6(5):482–6, 2005. doi: 10.1038/sj.embor.7400403. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1299309>.
- [83] Toni Hoffmann, Katarzyna M Tych, David J Brockwell, and Lorna Dougan. Single-molecule force spectroscopy identifies a small cold shock protein as being mechanically robust. *The Journal of Physical Chemistry. B*, 117(6):1819–26, 2013. doi: 10.1021/jp310442s. URL <http://dx.doi.org/10.1021/jp310442s>.
- [84] Jörg Schönfelder, Raul Perez-Jimenez, and Victor Muñoz. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nature communications*, 7:11777, 2016. doi: 10.1038/ncomms11777. URL <http://www.nature.com/ncomms/2016/160601/ncomms11777/full/ncomms11777.html{#}ref47>.
- [85] Benjamin Schuler, Everett A. Lipman, and William A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002. doi: 10.1038/nature01060. URL <http://www.nature.com/articles/nature01060>.
- [86] Chad Ray, Jason R Brown, and Boris B Akhremitchev. Single-molecule force spectroscopy measurements of "hydrophobic bond" between tethered hexadecane molecules. *The journal of Physical Chemistry. B*, 110(35):17578–83, 2006. doi: 10.1021/jp063517r. URL <http://dx.doi.org/10.1021/jp063517r>.
- [87] R Merkel, P Nassoy, A Leung, K Ritchie, and E Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–3, 1999. doi: 10.1038/16219. URL <http://dx.doi.org/10.1038/16219>.
- [88] Ali A Almaqwashi, Thayaparan Paramanathan, Ioulia Rouzina, and Mark C Williams. Mechanisms of small molecule-DNA interactions probed by single-molecule force spectroscopy. *Nucleic Acids Research*, pages gkw237–, 2016. doi: 10.1093/nar/gkw237. URL <http://nar.oxfordjournals.org/content/early/2016/04/15/nar.gkw237.long>.
- [89] Marcel Ander, Sivaraman Subramaniam, Karim Fahmy, A. Francis Stewart, and Erik Schäffer. A Single-Strand Annealing Protein Clamps DNA to Detect and Secure Homology. *PLOS Biology*, 13(8):e1002213, 2015. doi: 10.1371/journal.pbio.1002213. URL <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002213>.
- [90] David R Jacobson and Omar A Saleh. Measuring the differential stoichiometry and energetics of ligand binding to macromolecules by single-molecule force spectroscopy: an extended theory. *The Journal of Physical Chemistry. B*, 119(5):1930–8, 2015. doi: 10.1021/jp511555g. URL <http://dx.doi.org/10.1021/jp511555g>.

- [91] Pierre Burgos, Zhenyu Zhang, Ramin Golestanian, Graham J Leggett, and Mark Geoghegan. Directed single molecule diffusion triggered by surface energy gradients. *ACS nano*, 3(10):3235–43, 2009. doi: 10.1021/nn900991r. URL <http://dx.doi.org/10.1021/nn900991r>.
- [92] Chih-Kung Lee, Yu-Ming Wang, Long-Sun Huang, and Shiming Lin. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. *Micron (Oxford, England : 1993)*, 38(5):446–61, 2007. doi: 10.1016/j.micron.2006.06.014. URL <http://www.sciencedirect.com/science/article/pii/S0968432806001375>.
- [93] Wei Zhang, Xiujuan Lü, Wenke Zhang, and Jiacong Shen. EMSA and Single-Molecule Force Spectroscopy Study of Interactions between *Bacillus subtilis* Single-Stranded DNA-Binding Protein and Single-Stranded DNA. *Langmuir*, 27(24):15008–15015, 2011. doi: 10.1021/la203752y. URL <http://dx.doi.org/10.1021/la203752y>.
- [94] Luda S Shlyakhtenko, Samrat Dutta, Ming Li, Reuben S Harris, and Yuri L Lyubchenko. Single-Molecule Force Spectroscopy Studies of APOBEC3A-Single-Stranded DNA Complexes. *Biochemistry*, 2016. doi: 10.1021/acs.biochem.6b00214. URL <http://dx.doi.org/10.1021/acs.biochem.6b00214>.
- [95] Junping Yu, Yaxin Jiang, Xinyong Ma, Yi Lin, and Xiaohong Fang. Energy landscape of aptamer/protein complexes studied by single-molecule force spectroscopy. *Chemistry, an Asian journal*, 2(2):284–9, 2007. doi: 10.1002/asia.200600230. URL <http://www.ncbi.nlm.nih.gov/pubmed/17441163>.
- [96] Venkatraman Ramanujam, Hema Chandra Kotamarthi, and Sri Rama Koti Ainavarapu. Ca²⁺ binding enhanced mechanical stability of an archaeal crystallin. *PloS one*, 9(4):e94513, 2014. doi: 10.1371/journal.pone.0094513. URL <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094513>.
- [97] Joanna F Swain and Lila M Gierasch. The changing landscape of protein allostery. *Current Opinion in Structural Biology*, 16(1):102–8, 2006. doi: 10.1016/j.sbi.2006.01.003. URL <http://www.sciencedirect.com/science/article/pii/S0959440X06000042>.
- [98] K Gunasekaran, Buyong Ma, and Ruth Nussinov. Is allostery an intrinsic property of all dynamic proteins? *Proteins*, 57(3):433–43, 2004. doi: 10.1002/prot.20232. URL <http://www.ncbi.nlm.nih.gov/pubmed/15382234>.
- [99] H Li, A F Oberhauser, S B Fowler, J Clarke, and J M Fernandez. Atomic force microscopy reveals the mechanical design of a modular protein. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6527–31, 2000. doi: 10.1073/pnas.120048697. URL <http://www.pnas.org/content/97/12/6527>.
- [100] Xiaotang Hu and Hongbin Li. Force spectroscopy studies on protein-ligand interactions: A single protein mechanics perspective. *FEBS Letters*, 588(19):3613–3620, 2014. doi: 10.1016/j.febslet.2014.04.009. URL <http://www.sciencedirect.com/science/article/pii/S0014579314002956>.
- [101] Sri Rama Koti Ainavarapu, Lewyn Li, Carmen L Badilla, and Julio M Fernandez. Ligand binding modulates the mechanical stability of dihydrofolate reductase.

- Biophysical Journal*, 89(5):3337–44, 2005. doi: 10.1529/biophysj.105.062034. URL <http://www.sciencedirect.com/science/article/pii/S0006349505729763>.
- [102] Yi Cao, Teri Yoo, Shulin Zhuang, and Hongbin Li. Protein-protein interaction regulates proteins’ mechanical stability. *Journal of Molecular Biology*, 378(5):1132–41, 2008. doi: 10.1016/j.jmb.2008.03.046. URL <http://www.sciencedirect.com/science/article/pii/S0022283608003665>.
- [103] Jian-Min Yuan, Chia-Lin Chyan, Huan-Xiang Zhou, Tse-Yu Chung, Haibo Peng, Guanghui Ping, and Guoliang Yang. The effects of macromolecular crowding on the mechanical stability of protein molecules. *Protein science : a publication of the Protein Society*, 17(12):2156–66, 2008. doi: 10.1110/ps.037325.108. URL <http://www.ncbi.nlm.nih.gov/pubmed/18780817>.
- [104] Yi Cao, Teri Yoo, and Hongbin Li. Single molecule force spectroscopy reveals engineered metal chelation is a general approach to enhance mechanical stability of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32):11152–7, 2008. doi: 10.1073/pnas.0803446105. URL <http://www.pnas.org/content/105/32/11152.full>.
- [105] Hema Chandra Kotamarthi, Anju Yadav, and Sri Rama Koti Ainavarapu. Small Peptide Binding Stiffens the Ubiquitin-like Protein SUMO1. *Biophysical Journal*, 108(2):360–7, 2015. doi: 10.1016/j.bpj.2014.11.3474. URL <http://www.sciencedirect.com/science/article/pii/S0006349514046888>.
- [106] Yi Cao and Hongbin Li. Engineered elastomeric proteins with dual elasticity can be controlled by a molecular regulator. *Nature Nanotechnology*, 3(8):512–516, 2008. doi: 10.1038/nnano.2008.168. URL <http://www.nature.com/doifinder/10.1038/nnano.2008.168>.
- [107] Joseph A Marsh, Sarah A Teichmann, and Julie D Forman-Kay. Probing the diverse landscape of protein flexibility and binding. *Current Opinion in Structural Biology*, 22(5):643–50, 2012. doi: 10.1016/j.sbi.2012.08.008. URL <http://www.sciencedirect.com/science/article/pii/S0959440X1200142X>.
- [108] J P Junker, K Hell, M Schlierf, W Neupert, and M Rief. Influence of substrate binding on the mechanical stability of mouse dihydrofolate reductase. *Biophysical Journal*, 89(5):L46–8, 2005. doi: 10.1529/biophysj.105.072066. URL <http://www.cell.com/article/S0006349505729349/fulltext>.
- [109] Yi Cao, M M Balamurali, Deepak Sharma, and Hongbin Li. A functional single-molecule binding assay via force spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40):15677–81, 2007. doi: 10.1073/pnas.0705367104. URL <http://www.pnas.org/content/104/40/15677.full>.
- [110] Johannes Stigler and Matthias Rief. Ligand-Induced Changes of the Apparent Transition-State Position in Mechanical Protein Unfolding. *Biophysical journal*, 109(2):365–72, 2015. doi: 10.1016/j.bpj.2015.06.009. URL <http://www.sciencedirect.com/science/article/pii/S0006349515005913>.
- [111] Yukinori Taniguchi and Masaru Kawakami. Variation in the mechanical unfolding pathway of p53DBD induced by interaction with p53 N-terminal region or DNA. *PloS one*, 7(11):e49003, 2012. doi: 10.1371/journal.pone.0049003. URL <https://doi.org/10.1371/journal.pone.0049003>.

- [112] Yong Jian Wang, Palma Rico-Lastres, Ainhoa Lezamiz, Marc Mora, Carles Solsona, Guillaume Stirnemann, and Sergi Garcia-Manyes. DNA Binding Induces a Nanomechanical Switch in the RRM1 Domain of TDP-43. *The Journal of Physical Chemistry Letters*, 9(14):3800–3807, 2018. doi: 10.1021/acs.jpcllett.8b01494. URL <http://pubs.acs.org/doi/10.1021/acs.jpcllett.8b01494>.
- [113] Morten Bertz and Matthias Rief. Ligand binding mechanics of maltose binding protein. *Journal of Molecular Biology*, 393(5):1097–105, 2009. doi: 10.1016/j.jmb.2009.08.066. URL <http://www.sciencedirect.com/science/article/pii/S0022283609010924>.
- [114] David J Brockwell, Emanuele Paci, Rebecca C Zinober, Godfrey S Beddard, Peter D Olmsted, D Alastair Smith, Richard N Perham, and Sheena E Radford. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nature Structural Biology*, 10(9):731–7, 2003. doi: 10.1038/nsb968. URL <http://dx.doi.org/10.1038/nsb968>.
- [115] Morten Bertz and Matthias Rief. Mechanical unfoldons as building blocks of maltose-binding protein. *Journal of Molecular Biology*, 378(2):447–58, 2008. doi: 10.1016/j.jmb.2008.02.025. URL <http://www.sciencedirect.com/science/article/pii/S0022283608002106>.
- [116] Lewis E. Kay. Protein dynamics from NMR. *Nature Structural Biology*, 5(7):513–517, 1998. doi: 10.1038/755. URL <http://www.nature.com/doi/10.1038/755>.
- [117] Jeremy K. M. Sanders and Brian K. Hunter. *Modern NMR spectroscopy: A guide for chemists. Second edition*, volume 15. Elsevier, 1994. ISBN 0198555679. doi: 10.1016/0165-6147(94)90288-7.
- [118] John Cavanagh, W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton. *Protein NMR spectroscopy: principles and practice*. Academic Press, 1996. ISBN 9780080471037.
- [119] David G. Reid. *Protein NMR Techniques*, volume 60. Humana Press, New Jersey, 1997. ISBN 0-89603-309-0. doi: 10.1385/0896033090. URL <http://link.springer.com/10.1385/0896033090>.
- [120] Ian R Kleckner and Mark P Foster. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta*, 1814(8):942–68, 2011. doi: 10.1016/j.bbapap.2010.10.012. URL <http://www.ncbi.nlm.nih.gov/pubmed/21059410>.
- [121] Bruce N. Smith. Natural Abundance of the Stable Isotopes of Carbon in Biological Systems. *BioScience*, 22(4):226–231, 1972. doi: 10.2307/1296391. URL <https://doi.org/10.2307/1296391>.
- [122] Joseph B Lambert, Gerhard Binsch, and John D Roberts. NITROGEN-15 MAGNETIC RESONANCE SPECTROSCOPY. I. CHEMICAL SHIFTS. *Proceedings of the National Academy of Sciences of the United States of America*, 51(76):735–7, 1964. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC300152/>.
- [123] Brian Whitehead, C. Jeremy Craven, and Jonathan P. Waltho. Double and Triple Resonance NMR Methods for Protein Assignment. *Protein NMR Techniques*, 60:

- 29–52, 1997. doi: 10.1385/0-89603-309-0:29. URL <http://link.springer.com/10.1385/0-89603-309-0:29>.
- [124] Shiou-Ru Tzeng and Charalampos G. Kalodimos. Protein activity regulation by conformational entropy. *Nature*, 488(7410):236–240, 2012. doi: 10.1038/nature11271. URL <http://www.nature.com/doifinder/10.1038/nature11271>.
- [125] Yangmee Kim. NMR spin relaxation approaches for characterizing protein motions. *Bio Design*, 3(1), 2015. URL http://www.bdjn.org/APP_{_}PDF/BDJN003-01-02.pdf.
- [126] Z. Luz and S. Meiboom. Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution-order of the reaction with respect to solvent. *The Journal of Chemical Physics*, 39(2):366–370, 1963. doi: 10.1063/1.1734254. URL <https://doi.org/10.1063/1.1734254>.
- [127] J. P. Carver and R. E. Richards. A general two-site solution for the chemical exchange produced dependence of T₂ upon the carr-Purcell pulse separation. *Journal of Magnetic Resonance (1969)*, 6(1):89–105, 1972. doi: 10.1016/0022-2364(72)90090-X. URL <https://www.sciencedirect.com/science/article/pii/002223647290090X>.
- [128] Xiang J. Song, Peter F Flynn, Kim A Sharp, and A Joshua Wand. Temperature dependence of fast dynamics in proteins. *Biophysical Journal*, 92(6), 2007. doi: 10.1529/biophysj.106.102160. URL <https://www.ncbi.nlm.nih.gov/pubmed/17218465>.
- [129] Arthur M. Mandel, Mikael Akke, and Arthur G. Palmer. Dynamics of ribonuclease H: Temperature dependence of motions on multiple time scales. *Biochemistry*, 35(50):16009–16023, 1996. doi: 10.1021/bi962089k. URL <http://pubs.acs.org/doi/abs/10.1021/bi962089k>.
- [130] M. M. Lopez, K. Yutani, and G. I. Makhatadze. Interactions of the Major Cold Shock Protein of *Bacillus subtilis* CspB with Single-stranded DNA Templates of Different Base Composition. *Journal of Biological Chemistry*, 274(47):33601–33608, 1999. doi: 10.1074/jbc.274.47.33601. URL <http://www.jbc.org/content/274/47/33601.full>.
- [131] Brian P. Austin, Sreedevi Nallamsetty, and David S. Waugh. Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner for the Production of Soluble Recombinant Proteins in *Escherichia coli*. In *Methods in molecular biology (Clifton, N.J.)*, volume 498, pages 157–172. 2009. doi: 10.1007/978-1-59745-196-3_11. URL <http://www.ncbi.nlm.nih.gov/pubmed/18988025>.
- [132] Andrea Pennati, Jiusheng Deng, and Jacques Galipeau. Maltose-binding protein fusion allows for high level bacterial expression and purification of bioactive mammalian cytokine derivatives. *PloS one*, 9(9):e106724, 2014. doi: 10.1371/journal.pone.0106724. URL <http://www.ncbi.nlm.nih.gov/pubmed/25198691>.
- [133] Paul G. Blommel and Brian G. Fox. A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expression and Purification*, 55(1):53–68, 2007. doi: 10.1016/j.pep.2007.04.013. URL <http://www.ncbi.nlm.nih.gov/pubmed/17543538>.

- [134] G. V. Sherbet, M. S. Lakshmi, and F. Cajone. Isoelectric characteristics and the secondary structure of some nucleic acids. *Biophysics of Structure and Mechanism*, 10(3):121–128, 1983. doi: 10.1007/BF00537554. URL <http://link.springer.com/10.1007/BF00537554>.
- [135] J A Glasel. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *BioTechniques*, 18(1):62–63, 1995. URL <http://www.ncbi.nlm.nih.gov/pubmed/7702855>.
- [136] Norma J Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1(6):2876–90, 2006. doi: 10.1038/nprot.2006.202. URL <http://www.ncbi.nlm.nih.gov/pubmed/17406547>.
- [137] Ewa Folta-Stogniew. Oligomeric States of Proteins Determined by Size-Exclusion Chromatography Coupled With Light Scattering, Absorbance, and Refractive Index Detectors. In *New and Emerging Proteomic Techniques*, pages 97–112. Humana Press, New Jersey, 2006. doi: 10.1385/1-59745-026-X:97. URL <http://link.springer.com/10.1385/1-59745-026-X:97>.
- [138] Allen P. Minton. Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. *Analytical Biochemistry*, 501:4–22, 2016. doi: 10.1016/j.ab.2016.02.007. URL <https://www.sciencedirect.com/science/article/pii/S0003269716000683?via=ihub>.
- [139] Berry Birdsall, Rodney W. King, Miriam R. Wheeler, Charles A. Lewis, Scott R. Goode, R. Bruce Bruce Dunlap, and Gordon C.K. Roberts. Correction for light absorption in fluorescence studies of protein-ligand interactions. *Analytical Biochemistry*, 132(2):353–361, 1983. doi: 10.1016/0003-2697(83)90020-9. URL <http://www.sciencedirect.com/science/article/pii/0003269783900209>.
- [140] Sunish K. Sugunan, Benjamin Robotham, Ryan P. Sloan, Jdrzej Szmytkowski, Kenneth P. Ghiggino, Matthew F. Paige, and Ronald P. Steer. Photophysics of untethered ZnTPP-fullerene complexes in solution. *Journal of Physical Chemistry A*, 115(44):12217–12227, 2011. doi: 10.1021/jp2082853. URL <http://pubs.acs.org/doi/abs/10.1021/jp2082853>.
- [141] Alexander G Kozlov and Timothy M Lohman. SSB binding to ssDNA using isothermal titration calorimetry. *Methods in molecular biology (Clifton, N.J.)*, 922:37–54, 2012. doi: 10.1007/978-1-62703-032-8_3. URL <http://www.ncbi.nlm.nih.gov/pubmed/22976176>.
- [142] Joel Tellinghuisen. Optimizing isothermal titration calorimetry protocols for the study of 1:1 binding: Keeping it simple. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(5):861–867, 2016. doi: 10.1016/J.BBAGEN.2015.10.011. URL <https://www.sciencedirect.com/science/article/pii/S0304416515002755>.
- [143] Moran Jerabek-Willemsen, Chistoph J. Wienken, Dieter Braun, Philipp Baaske, and Stefan Duhr. Molecular Interaction Studies Using Microscale Thermophoresis. *ASSAY and Drug Development Technologies*, 9(4):342–353, 2011. doi: 10.1089/adt.2011.0380. URL <https://doi.org/10.1089/adt.2011.0380>.
- [144] Jeffrey L. Hutter and John Bechhoefer. Calibration of atomic force microscope tips. *Review of Scientific Instruments*, 64(7):1868–1873, 1993. doi: 10.1063/1.1143970. URL <https://doi.org/10.1063/1.1143970>.

- [145] Megan L Hughes and Lorna Dougan. The physics of pulling polypeptides: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Reports on Progress in Physics*, 79(7):76601, 2016. doi: 10.1088/0034-4885/79/7/076601. URL <https://iopscience.iop.org/article/10.1088/0034-4885/79/7/076601>.
- [146] R L Willett, K W Baldwin, K W West, and L N Pfeiffer. Differential adhesion of amino acids to inorganic surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7817–22, 2005. doi: 10.1073/pnas.0408565102. URL <http://www.ncbi.nlm.nih.gov/pubmed/15901900>.
- [147] Evan Evans. Probing the Relation Between Force-Lifetime- and Chemistry in Single Molecular Bonds. *Annual Review of Biophysics and Biomolecular Structure*, 30(1):105–128, 2001. doi: 10.1146/annurev.biophys.30.1.105. URL <https://doi.org/10.1146/annurev.biophys.30.1.105>.
- [148] David P Sadler, Eva Petrik, Yukinori Taniguchi, James R Pullen, Masaru Kawakami, Sheena E Radford, and David J Brockwell. Identification of a mechanical rheostat in the hydrophobic core of protein L. *Journal of Molecular Biology*, 393(1):237–48, 2009. doi: 10.1016/j.jmb.2009.08.015. URL <http://www.sciencedirect.com/science/article/pii/S0022283609009991>.
- [149] Peter Hinterdorfer and Yves F Dufrene. Detection and localization of single molecular recognition events using atomic force microscopy. *Nature Methods*, 3(5):347–55, 2006. doi: 10.1038/nmeth871. URL <http://www.ncbi.nlm.nih.gov/pubmed/16628204>.
- [150] Frank Delaglio, Stephan Grzesiek, Geerten W. Vuister, Guang Zhu, John Pfeifer, and Ad Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, 1995. doi: 10.1007/BF00197809. URL <http://link.springer.com/10.1007/BF00197809>.
- [151] Wim F. Vranken, Wayne Boucher, Tim J. Stevens, Rasmus H. Fogh, Anne Pajon, Miguel Llinas, Eldon L. Ulrich, John L. Markley, John Ionides, and Ernest D. Laue. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4):687–696, 2005. doi: 10.1002/prot.20449. URL <http://www.ncbi.nlm.nih.gov/pubmed/15815974>.
- [152] Alexandra Ahlner, Mats Carlsson, Bengt Harald Jonsson, and Patrik Lundström. PINT: A software for integration of peak volumes and extraction of relaxation rates. *Journal of Biomolecular NMR*, 56(3):191–202, 2013. doi: 10.1007/s10858-013-9737-7. URL <http://link.springer.com/10.1007/s10858-013-9737-7>.
- [153] Markus Niklasson, Renee Otten, Alexandra Ahlner, Cecilia Andresen, Judith Schlagnitweit, Katja Petzold, and Patrik Lundström. Comprehensive analysis of NMR data using advanced line shape fitting. *Journal of Biomolecular NMR*, 69(2):93–99, 2017. doi: 10.1007/s10858-017-0141-6. URL <http://link.springer.com/10.1007/s10858-017-0141-6>.
- [154] Ian R Kleckner and Mark P Foster. GUARDD: User-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. *Journal of Biomolecular NMR*, 52(1):11–22, 2012. doi: 10.1007/s10858-011-9589-y. URL <http://www.ncbi.nlm.nih.gov/pubmed/22160811>.

- [155] Dmitry M. Korzhnev, Karin Kloiber, and Lewis E. Kay. Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: Theory and application. *Journal of the American Chemical Society*, 126(23):7320–7329, 2004. doi: 10.1021/ja049968b. URL <http://www.ncbi.nlm.nih.gov/pubmed/15186169>.
- [156] Toni Hoffmann, Katarzyna M. Tych, Thomas Crosskey, Bob Schiffrin, David J. Brockwell, and Lorna Dougan. Rapid and Robust Polyprotein Production Facilitates Single-Molecule Mechanical Characterization of β -Barrel Assembly Machinery Polypeptide Transport Associated Domains. *ACS Nano*, page 150826162530004, 2015. doi: 10.1021/acsnano.5b01962. URL <http://dx.doi.org/10.1021/acsnano.5b01962>.
- [157] Mariano Carrion-Vazquez, Andres F Oberhauser, Thomas E Fisher, Piotr E Marszalek, Hongbin Li, and Julio M Fernandez. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Progress in Biophysics and Molecular Biology*, 74(1-2):63–91, 2000. doi: 10.1016/S0079-6107(00)00017-1. URL <https://www.sciencedirect.com/science/article/pii/S0079610700000171>.
- [158] W A Linke. Stretching molecular springs: elasticity of titin filaments in vertebrate striated muscle. *Histology and histopathology*, 15(3):799–811, 2000. doi: 10.14670/HH-15.799. URL <http://www.ncbi.nlm.nih.gov/pubmed/10963124>.
- [159] David J Brockwell, Godfrey S Beddard, John Clarkson, Rebecca C Zinober, Anthony W Blake, John Trinick, Peter D Olmsted, D Alastair Smith, and Sheena E Radford. The effect of core destabilization on the mechanical resistance of I27. *Biophysical Journal*, 83(1):458–72, 2002. doi: 10.1016/S0006-3495(02)75182-5. URL <http://www.sciencedirect.com/science/article/pii/S0006349502751825>.
- [160] Yukinori Taniguchi, David J Brockwell, and Masaru Kawakami. The effect of temperature on mechanical resistance of the native and intermediate states of I27. *Biophysical journal*, 95(11):5296–305, 2008. doi: 10.1529/biophysj.108.141275. URL <http://www.sciencedirect.com/science/article/pii/S0006349508789549>.
- [161] Michael Schlierf and Matthias Rief. Temperature softening of a protein in single-molecule experiments. *Journal of Molecular Biology*, 354(2):497–503, 2005. doi: 10.1016/j.jmb.2005.09.070. URL <http://www.sciencedirect.com/science/article/pii/S0022283605011496>.
- [162] Edward C Hulme and Mike A Trevethick. Ligand binding assays at equilibrium: validation and interpretation. *British journal of pharmacology*, 161(6):1219–37, 2010. doi: 10.1111/j.1476-5381.2009.00604.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/20132208>.
- [163] Arieh Ben-Naim. The Binding Isotherm. In *Cooperativity and Regulation in Biochemical Processes*, pages 25–49. Springer US, Boston, MA, 2001. doi: 10.1007/978-1-4757-3302-0_2. URL http://link.springer.com/10.1007/978-1-4757-3302-0_{_}2.
- [164] Katja Schröder, Peter Graumann, Arndt Schnuchel, Tad A. Holak, and Mohamed A. Marahiel. Mutational analysis of the putative nucleic acid-binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of single-stranded DNA containing the Y-box motif. *Molec-*

- ular Microbiology*, 16(4):699–708, 1995. doi: 10.1111/j.1365-2958.1995.tb02431.x. URL <http://doi.wiley.com/10.1111/j.1365-2958.1995.tb02431.x>.
- [165] P S Aguilar, A M Hernandez-Arriaga, L E Cybulski, A C Erazo, and D de Mendoza. Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*. *The EMBO Journal*, 20(7):1681–91, 2001. doi: 10.1093/emboj/20.7.1681. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=145467>.
- [166] Stefan Duhr and Dieter Braun. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences of the United States of America*, 103(52):19678–82, 2006. doi: 10.1073/pnas.0603873103. URL <http://www.pnas.org/content/103/52/19678.full>.
- [167] Christoph J Wienken, Philipp Baaske, Ulrich Rothbauer, Dieter Braun, and Stefan Duhr. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, 1:100, 2010. doi: 10.1038/ncomms1093. URL <http://dx.doi.org/10.1038/ncomms1093>.
- [168] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, 31(13):3406–15, 2003. URL <http://www.ncbi.nlm.nih.gov/pubmed/12824337>.
- [169] Taekjip Ha, Alexander G. Kozlov, and Timothy M. Lohman. Single-Molecule Views of Protein Movement on Single-Stranded DNA. 2012. URL <https://doi.org/10.1146/annurev-biophys-042910-155351>.
- [170] Vickery L. Arcus, Erica J. Prentice, Joanne K. Hobbs, Adrian J. Mulholland, Marc W. Van Der Kamp, Christopher R. Pudney, Emily J. Parker, and Louis A. Schipper. On the Temperature Dependence of Enzyme-Catalyzed Rates. *Biochemistry*, 55(12):1681–1688, 2016. doi: 10.1021/acs.biochem.5b01094. URL <http://pubs.acs.org/doi/10.1021/acs.biochem.5b01094>.
- [171] Nicholas P George and James L Keck. Molecular biology: Slip sliding on DNA. *Nature*, 461(7267):1067–8, 2009. doi: 10.1038/4611067a. URL <http://dx.doi.org/10.1038/4611067a>.
- [172] Jianyi Yang, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. The I-TASSER suite: Protein structure and function prediction, 2014. URL <http://www.nature.com/articles/nmeth.3213>.
- [173] Ambrish Roy, Alper Kucukural, and Yang Zhang. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4):725–738, 2010. doi: 10.1038/nprot.2010.5. URL <http://www.nature.com/articles/nprot.2010.5>.
- [174] Yang Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9(1):40, 2008. doi: 10.1186/1471-2105-9-40. URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-40>.
- [175] Geoffrey Bodenhausen and David J. Ruben. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1):185–189, 1980. doi: 10.1016/0009-2614(80)80041-8. URL [https://doi.org/10.1016/0009-2614\(80\)80041-8](https://doi.org/10.1016/0009-2614(80)80041-8).

- [176] N J Baxter and M P Williamson. Temperature dependence of ^1H chemical shifts in proteins. *Journal of biomolecular NMR*, 9(4):359–69, 1997. URL <http://www.ncbi.nlm.nih.gov/pubmed/9255942>.
- [177] Jennifer H. Tomlinson and Mike P. Williamson. Amide temperature coefficients in the protein G B1 domain. *Journal of Biomolecular NMR*, 52(1):57–64, 2012. doi: 10.1007/s10858-011-9583-4. URL <http://www.ncbi.nlm.nih.gov/pubmed/22076570>.
- [178] Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171):913–916, 2007. doi: 10.1038/nature06407. URL <http://www.nature.com/doifinder/10.1038/nature06407>.
- [179] Philipp Neudecker, Patrik Lundström, and Lewis E Kay. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophysical Journal*, 96(6):2045–2054, 2009. doi: 10.1016/j.bpj.2008.12.3907. URL <http://www.ncbi.nlm.nih.gov/pubmed/19289032>.
- [180] Sneha Munshi, Soundhararajan Gopi, Sandhyaa Subramanian, Luis A Campos, and Athi N Naganathan. Protein plasticity driven by disorder and collapse governs the heterogeneous binding of CytR to DNA. *Nucleic Acids Research*, 2018. doi: 10.1093/nar/gky176. URL <https://doi.org/10.1093/nar/gky176>.
- [181] Michael J. Morten, Roland Gamsjaeger, Liza Cubeddu, Ruvini Kariawasam, Jose Peregrina, J. Carlos Penedo, and Malcolm F. White. High-affinity RNA binding by a hyperthermophilic single-stranded DNA-binding protein. *Extremophiles*, pages 1–11, 2017. doi: 10.1007/s00792-016-0910-2. URL <http://link.springer.com/10.1007/s00792-016-0910-2>.
- [182] Franz X. Schmid, Dieter Perl, Uwe Mueller, and Udo Heinemann. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Structural Biology*, 7(5):380–383, 2000. doi: 10.1038/75151. URL <http://www.nature.com/doifinder/10.1038/75151>.
- [183] Kate A. Stafford, Nikola Trbovic, Joel A. Butterwick, Robert Abel, Richard A. Friesner, and Arthur G. Palmer. Conformational preferences underlying reduced activity of a thermophilic ribonuclease H. *Journal of Molecular Biology*, 427(4):853–866, 2015. doi: 10.1016/j.jmb.2014.11.023. URL <https://doi.org/10.1016/j.jmb.2014.11.023>.
- [184] Dieter Perl, Georg Holtermann, and Franz X. Schmid. Role of the Chain Termini for the Folding Transition State of the Cold Shock Protein. *Biochemistry*, 40(51):15501–15511, 2001. doi: 10.1021/bi011378s. URL <http://dx.doi.org/10.1021/bi011378s>.
- [185] E. D. Watt, H. Shimada, E. L. Kovrigin, and J. P. Loria. The Mechanism of Rate-Limiting Motions in Enzyme Function. *Proceedings of the National Academy of Sciences*, 104(29):11981–11986, 2007. doi: 10.1073/pnas.0702551104. URL <http://www.ncbi.nlm.nih.gov/pubmed/17615241>.
- [186] José A Caro, Kyle W Harpole, Vignesh Kasinath, Jackwee Lim, Jeffrey Granja, Kathleen G Valentine, Kim A Sharp, and A Joshua Wand. Entropy in molecular recognition by proteins. *Proceedings of the National Academy of Sciences of the*

United States of America, 114(25):6563–6568, 2017. doi: 10.1073/pnas.1621154114. URL <http://www.ncbi.nlm.nih.gov/pubmed/28584100>.

- [187] Davoud Mozhdehi, Kelli M. Luginbuhl, Joseph R. Simon, Michael Dzuricky, Rüdiger Berger, H. Samet Varol, Fred C. Huang, Kristen L. Buehne, Nicholas R. Mayne, Isaac Weitzhandler, Mischa Bonn, Sapun H. Parekh, and Ashutosh Chilkoti. Genetically encoded lipidpolypeptide hybrid biomaterials that exhibit temperature-triggered hierarchical self-assembly. *Nature Chemistry*, page 1, 2018. doi: 10.1038/s41557-018-0005-z. URL <http://www.nature.com/articles/s41557-018-0005-z>.
- [188] João Conde, Nuria Oliva, Mariana Atilano, Hyun Seok Song, and Natalie Artzi. Self-assembled RNA-triple-helix hydrogel scaffold for microRNA modulation in the tumour microenvironment. *Nature Materials*, 15(3):353–363, 2016. doi: 10.1038/nmat4497. URL <http://www.nature.com/doifinder/10.1038/nmat4497>.
- [189] Chun Gwon Park, Christina A Hartl, Daniela Schmid, Ellese M Carmona, Hye-Jung Kim, and Michael S Goldberg. Extended release of perioperative immunotherapy prevents tumor recurrence and eliminates metastases. *Science Translational Medicine*, 10(433):eaar1916, 2018. doi: 10.1126/scitranslmed.aar1916. URL <http://www.ncbi.nlm.nih.gov/pubmed/29563317>.
- [190] Marcelo A. Da Silva, Samuel Lenton, Matthew Hughes, David J. Brockwell, and Lorna Dougan. Assessing the Potential of Folded Globular Polyproteins As Hydrogel Building Blocks. *Biomacromolecules*, 18(2):636–646, 2017. doi: 10.1021/acs.biomac.6b01877. URL <http://pubs.acs.org/doi/10.1021/acs.biomac.6b01877>.

-