



# *From Microscale Network Topology to Macroscale Mechanics of Folded protein Hydrogels*

**Alexander James Wright**

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

The University of Leeds

Astbury Centre for Structural Molecular Biology

**March 2021**

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

© 2021 The University of Leeds and Alexander J. Wright

## Acknowledgements

I would firstly like to express my sincere thanks to my two supervisors Professors David Brockwell and Lorna Dougan for making this PhD project possible. Their constant support, patience, and knowledge has combined to make the last 4 years perhaps the most enjoyable and stimulating working experience of my career, and I will forever be grateful for their belief in me.

Secondly I would like to thank all the colleagues I have met and worked with across the Brockwell, Radford, Berry, and Dougan groups, as well as many others. The various expertise and knowledge of the people included in these groups has acted as a crucial source of support, ideas, and problem solving solutions, and without them I would not have had the energy, resources, and self-belief to complete this work. I am thankful and eternally grateful to you all. I would like to make special thanks to both Nasir Khan, lab manager, and Sophie Cussons, research technician, for their technical and emotional support through the tough times, and their limitless patience with my breaking of things and not doing as I was told. From a scientific standpoint I would like to give special thanks to Matt Hughes for his patience in teaching me some (to him) really basic physics and mathematics without which my understanding of these subjects would not have been adequate to complete this thesis. Thirdly I would like to thank my Mum and Dad for their support and enthusiasm despite having no idea what I was talking about when describing my work. Your support despite me becoming an eternal student was crucial in me making it through the difficult periods.

Finally and most of all I would like to thank and dedicate this thesis to Steph. Without you this all would have been utterly impossible and utterly pointless. You make making an effort worthwhile. Love you and thank you forever.

# Contents

Acknowledgements .....	ii
Abbreviations.....	ix
List of Figures.....	xi
List of Tables .....	xv
Abstract .....	xvi
Section 1: Introduction .....	1
1.1 Hydrogels.....	1
1.1.1 Introduction and Definition .....	1
1.1.2 <i>In vivo</i> Hydrogels.....	4
1.1.2.1 The Extracellular Matrix.....	4
1.1.3 Synthetic Hydrogels .....	10
1.1.3.1 3D Cell Culture, Wound Dressings, and Implants .....	10
1.1.3.2 Other Novel Hydrogel Applications .....	14
1.1.4 Folded Protein Hydrogels .....	17
1.1.4.1 Dynamic Activity and Smart Biomaterials.....	17
1.1.4.2 The Rheology of Folded Protein Hydrogels .....	20
1.1.4.3 FGP Hydrogel Network Design.....	23
1.2 Hydrogel Networks .....	25
1.2.1 Physical Hydrogels .....	26
1.2.2 Chemical Hydrogels .....	29
1.2.3 How to Form Protein Networks via Crosslinking .....	32
1.2.4 Mechanisms of Hydrogel Formation .....	41
1.3 Immunoglobulin 27.....	49
1.3.1 Titin .....	49
1.3.2 Immunoglobulin Domain 27 .....	51
1.4 Rheology .....	55
1.4.1 Introduction .....	55
1.4.2 Viscosity and Structured Fluids.....	56
1.4.3 Elasticity and Perfectly Elastic Materials .....	58
1.4.4 Viscoelastic Materials and Models .....	60

1.5 Aims of This Thesis.....	65
1.5.1 Measuring the Folded Fraction of Protein in an FGP Hydrogel .....	67
1.5.2 Measuring Crosslinking Efficiency and its Correlation to Macromechanical Hydrogel properties.....	68
Section 2: Materials and Methods.....	70
2.1 Materials.....	70
2.1.1 Centrifuges .....	70
2.1.2 Incubators .....	70
2.1.3 Protein Purification Equipment.....	70
2.1.4 Spectrophotometers.....	71
2.1.4.1 Cuvettes.....	71
2.1.5 PCR/Cloning Assembly Thermocycler .....	71
2.1.6 Circular Dichroism.....	71
2.1.7 Fluorimeter .....	71
2.1.8 Microplate Fluorescence Readers .....	71
2.1.9 Mass Spectrometer.....	72
2.1.10 Gel Electrophoresis.....	72
2.1.11 Gel Ladders and Dyes .....	72
2.1.12 DNA Purification Kits.....	72
2.1.13 DNA Primers .....	72
2.1.14 Plasmids .....	72
2.1.15 Bacterial Strains .....	73
2.1.16 Rheometer .....	73
2.1.17 Design of Experiment Software .....	73
2.1.18 Refractometer .....	73
2.1.19 Buffers, Solutions, and Growth Medias .....	73

2.1.19.1 Antibiotic stocks.....	73
2.1.19.2 SDS-PAGE Tris-Tricine Gel Solutions.....	74
2.1.19.3 Bacterial Growth Medias.....	74
2.1.19.4 DoE Protein Purification Buffers.....	74
2.1.19.5 Large-scale Protein Purification Solutions.....	75
2.1.19.6 DNA Agarose Gel Solutions.....	75
2.1.19.7 Folded Fraction Measurement Solutions.....	75
2.1.20 Kits and Products.....	75
2.1.21 Transilluminators.....	76
2.1.22 Gelation Lamp.....	76
2.1.23 Chemicals.....	77
2.2 Methods.....	80
2.2.1 Molecular Biology.....	80
2.2.1.1 Plasmid DNA Preparation.....	80
2.2.1.2 Site Directed PCR Mutagenesis.....	81
2.2.1.3 Golden Gate Polyprotein Assembly.....	83
2.2.2 Polyprotein Expression Optimisation.....	89
2.2.2.1 Design of Experiment Methodology.....	89
2.2.2.2: Test Culture Growth:.....	90
2.2.2.3: Protein Purification and Response measurement.....	90
2.2.2.4: Densitometry:.....	91
2.2.2.5 Subsequent Modelling of Expression Landscape.....	91
2.2.3 Large Scale Expression and Purification of All Protein/Polyproteins.....	91
2.2.3.1 Starter Cultures and Media.....	91
2.2.3.2 Inoculation and Growth.....	92
2.2.3.3 Cell Harvesting and Lysis.....	92
2.2.3.4 Chromatography.....	92
2.2.3.5 Size Exclusion Chromatography.....	93
2.2.3.6 Lyophilisation.....	93
2.2.3.7 TEV Protease Production.....	94
2.2.4 Biochemistry techniques.....	94
2.2.4.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE).....	94
2.2.4.2 DNA Agarose Gel Electrophoresis.....	95

2.2.4.3	Protein Concentration Determination .....	96
2.2.4.4	Optical Density 600 Measurements .....	96
2.2.5	Biophysical Techniques .....	97
2.2.5.1	Hydrogel Preparation .....	97
2.2.5.2	Thermodynamic Stability Measurements of Polyproteins by Urea Denaturation .....	98
2.2.5.3	Hydrogel Folded Fraction Measurement .....	98
2.2.5.4	Measurement of Crosslinking Efficiency .....	100
2.2.5.5	Circular dichroism .....	100
2.2.5.6	Mass spectrometry .....	102
2.2.5.7	Rheology .....	102
Section 3: Optimisation of Polyprotein Cloning, Design, and Expression .....		104
3.1	Introduction .....	104
3.3	Polyprotein Cloning .....	106
3.3.1	Linker Region Design .....	107
3.3.2	Gibson Assembly Method Limitations .....	109
3.3.3	Golden Gate Assembly Design .....	110
3.3.4	Test Expression and Mass Spectrometry Analysis .....	113
3.3.4.1	Protein Test Expression and ID Conformation .....	113
3.3.4.2	Circular Dichroism of Novel Constructs .....	115
3.4	Optimisation of Polyprotein Expression .....	118
3.4.1	Design of Experiments Methodology .....	118
3.4.2	Validation of Repeatability of (I27) <sub>5</sub> Expression Measurement .....	122
3.4.3	Polyprotein Expression Response Measurement .....	123
3.4.3.1	Autoinduction Medium .....	123
3.4.3.2	Terrific Broth Medium .....	127
3.4.3.3	2YT Medium .....	130
3.4.4	Design of Experiment Analysis .....	133
3.4.5	Large-Scale Protein Expression and Yield .....	134
3.5	Discussion .....	136
Section 4: Thermodynamic Characterization of I27 Polyproteins .....		139
4.1	Introduction .....	139

4.2 Thermodynamic Characterization of Polyproteins .....	140
4.3.1 Chemical Denaturation .....	140
4.3.2 Thermal Denaturation .....	144
4.3 Solid-State Circular Dichroism .....	148
4.4 Discussion .....	151
4.4.1 Thermodynamics .....	151
4.4.2 Protein Secondary Structure in Hydrogels.....	152
Section 5: Quantification of Hydrogel Crosslinking Efficiency and Post-Gelation Folded Fraction .....	153
5.1 Introduction .....	153
5.2 Quantifying Crosslinking Post-Gelation .....	154
5.2.1: Investigating Crosslinking Reaction Specificity .....	154
5.2.2: Measuring the Photometric Equivalency of Monomeric Dityrosine and Dityrosine Adducts within Protein Hydrogels .....	158
5.2.3: Subtracting IAEDANS/Dityrosine Fluorescence from Full Hydrogel Fluorescence Spectra .....	161
5.2.4: Crosslinking Efficiency/Crosslinks per Monomer versus Geometry .....	163
5.3: Quantification of Hydrogel Folded Protein Fraction.....	169
5.3.1 Measuring IAEDANS Fluorescence Labelling versus Unfolded Fraction .....	170
5.3.2 Quantification of Native Hydrogel Folded Fractions .....	172
5.4 Discussion .....	173
5.4.1 Unfolded Fractions .....	173
5.4.2 Crosslinking.....	174
Section 6: Rheological Characterisation of I27 Polyprotein Hydrogels .....	176
6.1 Introduction.....	176
6.1.1 Small Amplitude Oscillatory Tests .....	177
6.1.2 Recovery and Relaxation .....	182
6.2 Gelation Analysis .....	184
6.2.1 Network Assembly.....	185
6.2.2 Network Relaxation .....	187
6.3 Macromechanics.....	191
6.3.1 Final Storage Moduli ( $G'_{\infty}$ ) and Frequency Response .....	191
6.3.2 Hysteresis.....	194
6.4 Discussion .....	197

6.4.1 Assembly Kinetics .....	197
6.4.2 Hydrogel Mechanics .....	201
6.4.3 Conclusions .....	204
Section 7: Final Conclusions and Future Directions .....	205
7.1: Microscale Network Topology .....	205
7.2 Translation of Microscale Network Topology to Macroscale Mechanics .....	206
7.3 Final Thoughts and Future Directions .....	207
Appendix .....	209
Appendix 1: Section 5 T-Test Results Summary Tables .....	209
Appendix 2: DNA and Protein Sequences .....	210
Appendix 2.1: I27 Monomers in pGem-T-Easy Vectors .....	210
Appendix 2.2: I27 Pentamers in pET14b Vectors .....	211
Appendix 2.3: Vectors .....	221
Appendix 3: SDS and Agarose Gel Ladders .....	223
References .....	224



## Abbreviations

$T_m$	Melting temperature
$\Delta C_p$	Change in heat capacity
(I27) <sub>5</sub>	I27 pentamer in pET23a vector (Appendix 2)
AFM	Atomic Force Microscopy
AI	Autoinduction
CD	Circular Dichroism
DLA	Diffusion Limited Aggregation
DLCA	Diffusion Limited Cluster Aggregation
DoE	Design of Experiment
ECM	Extracellular Matrix
FDA	Food and Drug Administration
FGP	Folded Globular Protein
FN	Fibronectin
I27	Immunoglobulin Domain 27
IAEDANS	5-({2-[(iodoacetyl)amino]ethyl}amino)naphthalene-1-sulfonic acid
Ig's	Immunoglobulin-like Domains
IPG	Immediately Post Gelation
$k_{B_b}$	Rate of Bond Breakage
$k_{B_f}$	Rate of Bond Formation
KD	Rate of Particle Diffusion
MMP	Matrix Metalloproteinase
MRE	Mean Residue Ellipticity
MWCO	Molecular Weight Cut Off
NaPi	25mM Sodium Phosphate Buffer pH 7.4
NEB	New England Biosciences
OD <sub>600</sub>	Optical density at 600nm
$P$	Probability of Bond Formation
PCR	Polymerase Chain Reaction
PDMA	poly(N,N-dimethylacrylamide)

PEG	Polyethylene Glycol
PVA	Polyvinyl Alcohol
<i>rd</i>	Reaction Distance
RLA	Reaction Limited Aggregation
RLCA	Reaction Limited Cluster Aggregation
$R_{MAX}$	Maximum Assembly Rate
SANS	Small Angle Neutron Scattering
SAOT	Small Amplitude Oscillatory Testing
SDS	Sodium Dodecyl Sulphate
SEM	Scanning Electron Microscopy
TB	Terrific Broth
TEMED	Tetramethylethylenediamine
UF50	Urea concentration at which 50% of a protein population is unfolded
$\Delta D$	Decline Time
$\Delta G$	Gibbs free Energy
$\Delta H$	Enthalpy of Denaturation
$\Delta T$	Lag Time

## List of Figures

- Figure 1.1: Schematic summary of sol-gel transition of hydrogel formation \_\_\_\_\_ Page 3
- Figure 1.2: Illustration of some of the major components and interactions of the extracellular matrix \_\_\_\_\_ Page 5
- Figure 1.3: Schematic summary of the hierarchical structure of mature collagen fibres \_\_\_\_\_ Page 6
- Figure 1.4 (A): The structure of Elastin, (B): A desmosine crosslink formed by the  $\epsilon$ -amino groups of 4 lysine residues \_\_\_\_\_ Page 7
- Figure 1.5 (A): Chemical structure of Hyaluronan chains, (B): Illustration of the structure and binding motifs of Fibronectin \_\_\_\_\_ Page 9
- Figure 1.6: Schematic summary of the differences in lattice structure between polyionic (physical) hydrogels and chemically crosslinked hydrogels \_\_\_\_\_ Page 29
- Figure 1.7: The reaction pathways involved in (poly)acrylamide gelation \_\_\_\_\_ Page 30
- Figure 1.8: Reaction mechanism summary for diazirine-mediated photocrosslinking \_\_\_\_\_ Page 35
- Figure 1.9: Effect of additives on the  $\text{Ru(II)(bpy)}_3^{2+}$ /light-mediated cross-linking of the Gal4 activation domain and Gal80 protein in the presence of APS \_\_\_\_\_ Page 39
- Figure 1.10: Simulation models of DLA and RLA network formation regimes \_\_\_\_\_ Page 46
- Figure 1.11: Simulation models of network formation by RLCA and DLCA dominated regimes \_\_\_\_\_ Page 48
- Figure 1.12: Schematic representation of the composition of Titin \_\_\_\_\_ Page 51
- Figure 1.13: A; the hydrogen bond clamp/zipper region between the A' and G strands of I27. B; The computationally predicted mechanically important residues of I27 \_\_\_\_\_ Page 52
- Figure 1.14: Top; primary sequence of I27 mutant domain used in this study, Bottom; X-ray crystallographic structure of I27 \_\_\_\_\_ Page 54
- Figure 1.15: Schematic representation of the relationship between system energy input, enthalpy, and entropy \_\_\_\_\_ Page 59
- Figure 1.16: Stress and strain amplitude versus time for an ideal fluid \_\_\_\_\_ Page 61

Figure 1.17: Schematic workflow undertaken during this thesis for the production of folded protein hydrogels_____	Page 66
Figure 2.1: Summarised mechanism of Golden Gate Assembly cloning_____	Page 83
Figure 3.1: Schematic Representation of I27 Mutant Domain Library_____	Page 106
Figure 3.2: Summarized schematic polyprotein cloning workflow for both Gibson and Golden Gate Assembly_____	Page 107
Figure 3.3: Nucleotide and amino acid linker sequences linker sequences_____	Page 108
Figure 3.4: Electrophoresis agarose gel image of the colony screen PCR products_____	Page 110
Figure 3.5: Summarised mechanism of Golden Gate Assembly cloning_____	Page 111
Figure 3.6: Agarose electrophoresis gel images showing the PCR colony screen products of each polyprotein DNA construct_____	Page 112
Figure 3.7. Schematic illustrations showing the number and distribution of tyrosine residues throughout the various polyprotein constructs_____	Page 114
Figure 3.8: Overlaid CD spectra of all polyprotein constructs designed for hydrogel formation_____	Page 115
Figure 3.9: Overlaid native fluorescence spectra of all constructs_____	Page 117
Figure 3.10: Distribution of test expression condition values within the 3-dimensional expression landscape cube_____	Page 119
Figure 3.11: Summary of variance of protein expression ratio measured from multiple gels_____	Page 122
Figure 3.12: Autoinduction Raw Response Data and protein expression SDS gel images for replicates 1 and 2_____	Page 124
Figure 3.13: Autoinduction DoE cubes_____	Page 126
Figure 3.14: Terrific Broth Media (TB) SDS-PAGE gels for Densitometric Analysis_____	Page 127
Figure 3.15: DoE cubes for Terrific Broth Medium_____	Page 129

Figure 3.16: DoE Cubes for 2YT medium	Page 132
Figure 3.17: Large-scale protein expression summary	Page 135
Figure 4.1: Overlaid chemical denaturation curves of I27 polyprotein constructs	Page 142
Figure 4.2: Total number of mutations in a polyprotein construct versus the UF50 concentration	Page 143
Figure 4.3: A; $\Delta G$ of homopolyproteins GC1-, 0-, and 4-A, B; m-value of homopolyproteins GC1-, 0-, and 4-A	Page 144
Figure 4.4: Folded versus unfolded full CD spectra and individual denaturation curves at 222nm for all constructs	Page 146
Figure 4.5: CD spectra pre- and post-gelation	Page 151
Figure 4.6: A; Circular dichroism at 222nm for each construct hydrogel at discrete timepoints	Page 152
Figure 5.1: I27 Domain with Potential Crosslinking Residues Highlighted	Page 155
Figure 5.2: Off-target crosslinking of dicysteine and hexa-his tag motifs	Page 156
Figure 5.3: GC0-A no off target crosslinking	Page 157
Figure 5.4: Photometric equivalency of dityrosine species	Page 160
Figure 5.5: Deconvolution of IAEDANS-labelled hydrogel fluorescence spectra	Page 162
Figure 5.6: Average crosslinking efficiency of same-biological repeat technical repeats of GC1-A and GC6-A hydrogels	Page 163
Figure 5.7: Schematic representation of the number and distribution of tyrosine residues throughout each polyprotein construct	Page 164
Figure 5.8: Summary of average crosslinking efficiency and standard deviation error of 3 biological repeats of GC3-A and GC6-A.	Page 165
Figure 5.9: Summary of all constructs all samples average crosslinking efficiency	Page 166
Figure 5.10: A; Average Crosslinks per Monomer/tyrosines per monomer for each Construct	Page 167
Figure 5.11: Urea concentration versus IAEDANS Fluorescence Intensity from Degraded 100mg/mL GC1-A Hydrogels	Page 171
Figure 5.12: Summarised average unfolded fractions of all constructs	Page 173

Figure 6.1: Geometric relationship between phase angle ( $\delta$ ), $G''$ , $G'$ , and $G^*$	Page 176
Figure 6.2: Example Gelation Curve with Assembly Kinetic Parameters Highlighted	Page 178
Figure 6.3: Example full gelation curves for GC2-A and GC3-A hydrogels with parameters illustrated	Page 180
Figure 6.4: Example Frequency response profiles for; A- a viscoelastic liquid, B- a viscoelastic solid, C- a gel-like material	Page 181
Figure 6.5: Example hysteresis loop of a viscoelastic material generated by a stress-strain cycle	Page 183
Figure 6.6: Schematic representation of the number and distribution of tyrosine residues throughout each polyprotein construct	Page 184
Figure 6.7: A; Summarised Decline Start times ( $\Delta D$ ) with standard deviation. B; Summarised maximum assembly rates ( $R_{MAX}$ ) with standard deviation	Page 185
Figure 6.8: A; Crosslinks per Monomer versus $R_{MAX}$ , B; Crosslinking Efficiency versus $R_{MAX}$ , C; Tyrosines per Monomer versus $R_{MAX}$	Page 186
Figure 6.9: A; Representative example gelation curves of each hydrogel construct. B-F; Example gelation curves with relaxation behaviour fit lines shown in red	Page 188
Figure 6.10: Relaxation time constants versus the average number of crosslinks per monomer of each construct	Page 190
Figure 6.11: A; Final Storage Moduli versus average Crosslinks per Monomer, B; Summarised Storage Moduli for each construct, C; Final Storage Moduli versus Tyrosines per Monomer	Page 192
Figure 6.12: A; Summarised average gradient of $G'$ increase between 0.01-1Hz frequency of each construct. B; Example frequency sweep plots of each construct	Page 193
Figure 6.13: Representative hysteresis loops for all constructs	Page 194
Figure 6.14: A-E; Energy recovery efficiency of each hydrogel construct as a function of the maximum induced strain. F-J; Summarised energy recovery efficiencies for all constructs at each maximum strain with standard deviation	Page 195
Figure 6.16: Efficiency of Energy recovery versus Crosslinks per Monomer	Page 196
Figure 6.17: Schematic summary of the order of constructs between DLA and RLA assembly Regimes	Page 198

## List of Tables

Table 1.1: Side chains groups of the nine most chemically reactive amino acids _____	Page 33
Table 2.1; Mutagenic primers used to create I27 sequence mutants and prepare the pET14b plasmid for use in Golden Gate Assembly _____	Page 81
Table 2.2: PCR Primers for Golden gate Cassette Amplification _____	Page 85
Table 2.3: DoE Input Variables _____	Page 89
Table 3.1: Summarised linker DNA and resulting Amino Acid sequences _____	Page 109
Table 3.2: All I27 pentamer constructs secondary structure composition calculated by CDSSTR algorithm _____	Page 116
Table 3.3: DoE input variables and output measurements _____	Page 119
Table 3.4: Experiment class designations with corresponding variable values _____	Page 120
Table 3.5: Summarised autoinduction medium output results with standard deviation _____	Page 114
Table 3.6: Summarised terrific broth medium output results with standard deviation where available _____	Page 128
Table 3.7: Summarised terrific broth medium output results _____	Page 130
Table 4.1: Summarised Thermodynamic Parameters of I27 Polyprotein Constructs _____	Page 142
Table 4.2: Summarised Melting Temperatures of Polyprotein Constructs Calculated from Thermal Denaturation CD Data _____	Page 147
Table 6.1: Summarised Relaxation Time Constants, and $G'_{\infty}$ 's from Fitted Gelation Curves _____	Page 189
Table 5.1: T-test Results of Crosslinking Efficiency Averages _____	Page 209
Table 5.2: T-test Results of Crosslinks per Monomer Averages _____	Page 209
Table 5.3: T-test Results of Unfolded Fraction Averages _____	Page 209

## **Abstract**

Hydrogels are a unique species of soft matter biomaterial. Generically defined as any hydrophilic polymer network capable of absorbing large volumes of water, their composition and characteristics can be extremely varied. With such an all-encompassing definition, hydrogels represent potentially the most tuneable and adaptable scaffold for biomaterial development in current research, with a myriad of possible applications in both laboratory and clinical settings, ranging from advanced wound dressings to 3-dimensional cell culture models.

The majority of hydrogels to date have been built from non-biological polymer lattices, but over the last 15 years research has shifted towards building hydrogels out of natively folded and dynamically active proteins. Central to the rational design of these materials will be a detailed understanding of the relationship between microscale network topologies and macroscale mechanics. The aim of this thesis is to characterize the relationship between the network crosslinking density and macroscale mechanics of a hydrogel system built from immunoglobulin domain 27 pentamers.

Firstly a facile method is described whereby the number of crosslink sites per monomer network building block can be precisely tuned, and proteins subsequently expressed in a high-yield manner. This was then followed by the development of two assays to measure the unfolded protein fraction post-gelation and the crosslinking efficiency of each hydrogel species. The macroscale mechanics of each hydrogel species was then characterized rheologically. By correlating these microscale measurements with the macroscale mechanics the mechanism of translation between length scales is discussed. I propose that it is possible to rationally tune the mechanics of folded protein hydrogels by the use of precisely situated



crosslink sites in the monomer building block. Furthermore I hypothesise that this tuneability is a result of different crosslink geometries causing changes in the network formation regime of the hydrogel. This leads to differences in network topologies and in turn hydrogel mechanics.

# Section 1: Introduction

## 1.1 Hydrogels

### 1.1.1 Introduction and Definition

Hydrogels are a unique species of soft materials with a staggeringly broad definition; a hydrogel is described as any solid polymeric material containing a large mass fraction of water. This definition encompasses materials as diverse as tofu, hagfish defence slime, the vitreous humour of eyeballs, and SDS polyacrylamide gels [1][2]. These examples illustrate the vast variety in structure, function, and the material properties of substances classed as hydrogels, and demonstrates their diverse and extensive biological and industrial relevance. Whilst the properties of different hydrogels vary hugely, they all share a single characteristic which uniquely defines them; they are formed by the entrapment of large volumes of water within a relatively sparse network [3], [4]. Conceptually hydrogels can be thought of as solid molten water, as they retain the Brownian motion of aqueous water whilst exhibiting the bulk mechanical characteristics of a solid, with their complex modulus dominated by the elastic component [5]. This allows an aqueous environment to exist within a discrete solid material, although diffusion between the internal aqueous environment and an external one is still possible with the rate of osmosis determined by the pore size of the network.

This unique property facilitates the encapsulation and concentration of water soluble molecules and living cells in a movable, protected, solid environment. These encapsulated molecules, varying from flavourings to create different edible jellies to biological molecules and live cells, are able to diffuse through the material according to the laws of osmosis and Brownian motion [6]. This allows biological and chemical activity to occur within a discrete and mechanically stable environment. This unique marriage of two states of matter (solid and

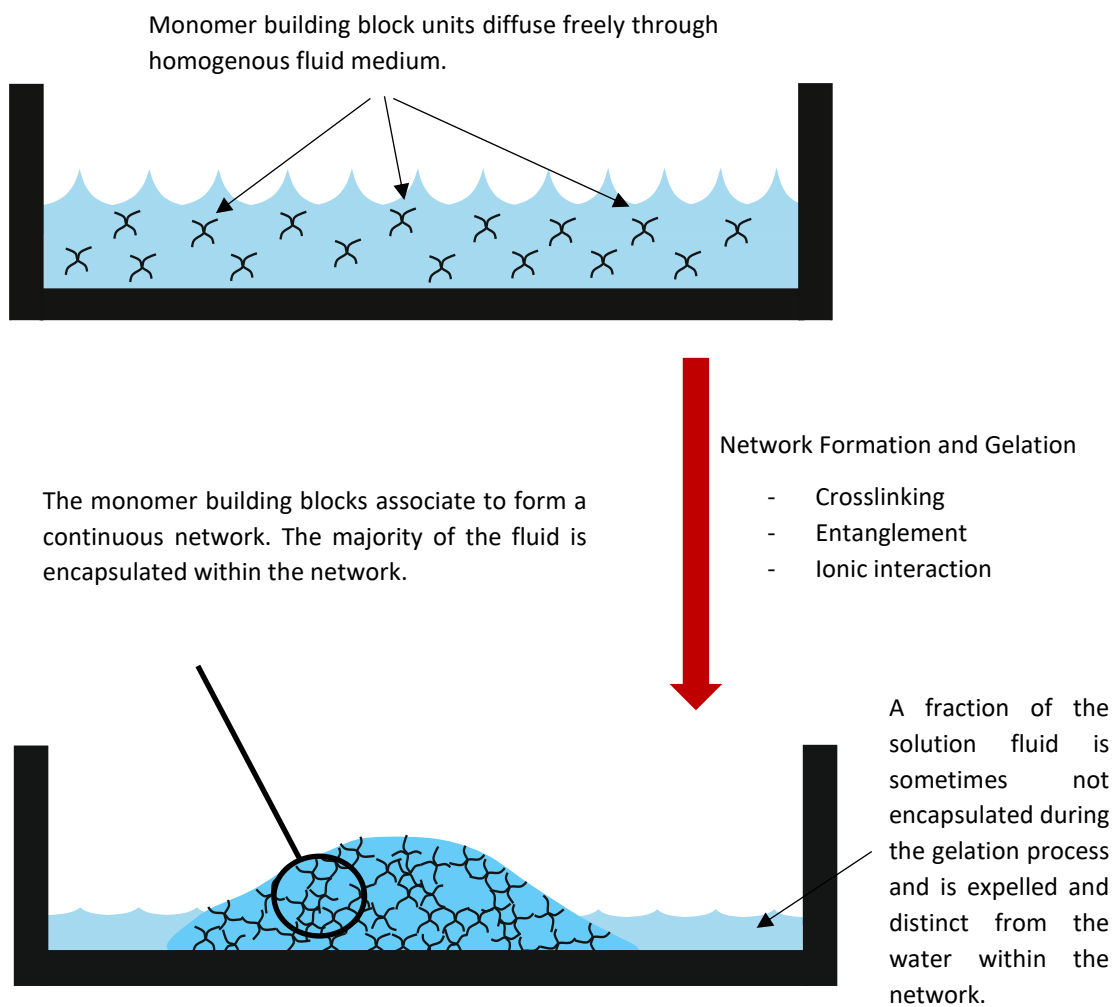
liquid) makes hydrogels a crucial environment without which complex life could not exist as all biochemical reactions are reliant either directly on water as a catalyst, or on Brownian motion and osmosis through a water-based solvent [7], [8]. Without water life as we know it cannot exist, and without a concentration of biological molecules high enough to allow complex chemical reaction pathways to exist, complex life could not evolve. Unicellular organisms achieve this by encapsulating their processes and active molecules within their cell membranes, but for multicellular life to exist extracellular processes such as signalling and diffusion of gasses and molecules between cells and tissues must be possible to allow the coordination of multiple complex tissues and organs. In the animal kingdom this coordination of tissues into complex multicellular life has been achieved by way of aqueous yet solid extracellular mediums of which there are countless examples in biological systems. For example the vitreous humour within mammalian eyes, the extracellular matrix which accounts for the majority of soft tissues in all animals, and almost the entire bodies of Jellyfish [9][4], [10].a

The prevalence of hydrogels throughout nature is replicated in industrial and research science settings. The most common analytical use of synthetic hydrogels is likely SDS polyacrylamide and Agarose gels, used for protein and DNA analysis respectively. The use of hydrogels in this capacity is indicative of what makes them so useful in general; they can retain molecules of various sizes and orientation within a solid network structure allowing them to be moved, concentrated, and separated. Simultaneously these molecules are still able to diffuse and migrate according to natural osmotic principles or as a result of some endemic property such as charge, mass, or size. In addition the properties of these gels are highly tuneable; the macro mechanics, pore-size distribution, and swelling ratio can all be rationally defined in synthetic hydrogels to meet certain requirements [11], [12]. These twin properties have led to the use

of hydrogels for a highly diverse array of industrial and healthcare technologies; used as filters, drug delivery systems, and tissue engineering scaffolds [3][13][14].

---

Figure 1.1: Schematic summary of sol-gel transition of hydrogel formation. Network monomer species undergoes dynamic polymerisation to form a continuous network.



## 1.1.2 *In vivo* Hydrogels

### 1.1.2.1 The Extracellular Matrix

The Extracellular Matrix (ECM) is potentially the most important tissue in the human body that most have never heard of. Not only does it provide crucial connectivity between organs without which a multi-organed organism could not exist, it also provides the mechanical basal lamina and metabolic repository without which individual organ systems could not continuously function. The vascular system is vital for delivering gaseous exchange services, but the ECM is equally important for the delivery of signals and nutrients, and without it nothing beyond unicellular life could have evolved.

The extracellular matrix (ECM) is a solid yet highly aqueous tissue which surrounds and encapsulates all the organs of the mammalian body, and comprises the non-cellular component of all tissues. It is a relaxed mesh network of collagen, fibronectin, and elastin embedded in a hydrogel of glycosaminoglycan-chain-containing proteoglycans [15]. The largest component by mass fraction however is water. The ECM acts as a physical mechanical scaffold for individual cells and organ systems, providing biochemical and biomechanical cues vital for tissue morphogenesis [16]. Genetic abnormalities leading to abnormal ECM formation are related to dozens of pathologies, both metabolic and cellular, most notably several cancers, and scurvy which is caused by a thinning of the collagen network of the ECM [17][18]. The concentration, ratios, and topology of these components is unique to each organ and tissue, but all provide the majority of a tissue/organs tensile and compressive strength and elasticity. The degree of each of these mechanical properties and the composition of the specific ECM directs the morphological organisation of the tissue by binding growth factors, interacting with cell surface receptors, and regulating the diffusion and turnover rate of

soluble signalling molecules [15][19]. The content and some of the interactions of the ECM are highlighted in Figure 1.2.

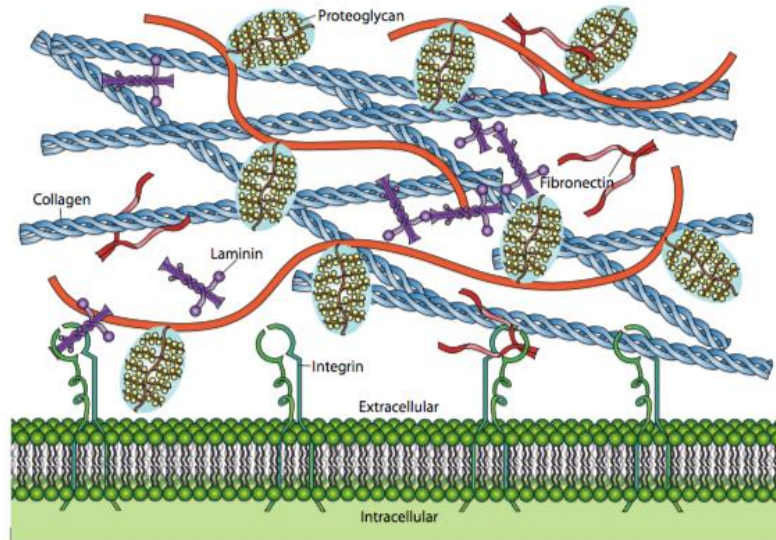


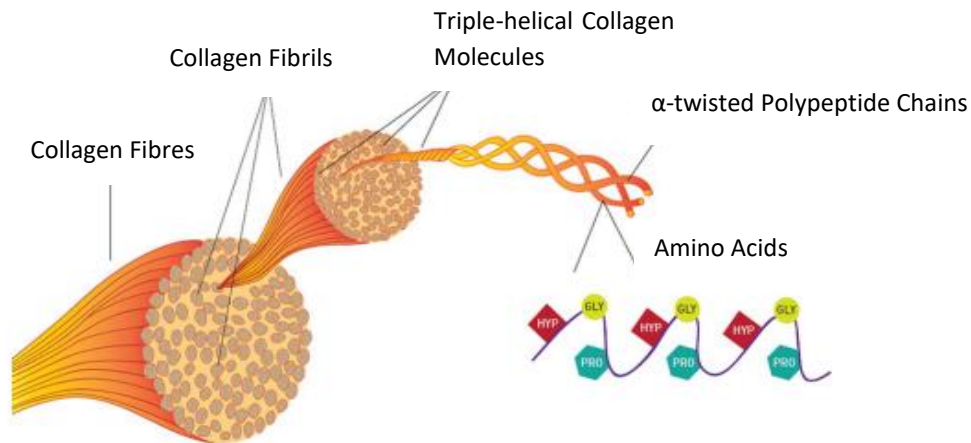
Figure 1.2: Illustration of some of the major components and interactions of the extracellular matrix. Reproduced from *Cells: Molecules and Mechanisms*, by Eric Wong, published 2009.

Collagen comprises approximately 30% of the total protein mass of multicellular animals [20]. Collagen fibres are hierarchical in structure; three disordered polypeptide chains associate and twist in a polyproline type 2 orientation to form a right-handed triple helix tropocollagen fibre as shown in Figure 1.3. All types of collagen exhibit a glycine residue at every third position with the two most common intermediate residues being proline and the modified hydroxyproline. Each glycine forms a hydrogen bond with one of the next two residues thereby stabilising the fibre. These tropocollagen molecules are then assembled into larger bundles of collagen fibres, cables and sheets by mechanical cell traction forces exerted by secretory fibroblasts. The ECM mechanics are distinct to their tissue and function depending on the requirements of the tissue [16][21][22][23][24]. This collagen network provides tensile

strength to the tissue and is the main structural element of the ECM, facilitating cell adhesion via novel binding site motifs, and cell migration via chemotaxis [16].

---

Figure 1.3: Schematic summary of the hierarchical structure of mature collagen fibres. Reproduced from Sibilla et al, 2015.



---

Elastin is the second most abundant species of fibrous protein in the ECM, composed of repeating  $\beta$ -turns stabilised by glycine-proline hydrogen bonds. Elastin is able to stretch several times over its resting length under stress, before recovering elastically upon the release of tension with virtually no viscous dissipation of energy *in vivo* [25]. Tropoelastin monomers assemble into mature elastin fibres and subsequently associate with collagen via lysine-lysine crosslinking [26]. The subsequent networks elastic properties are therefore determined by three components; collagen network density, inherent elastin elasticity, and the number and distribution of lysine crosslinks. Elastin contains on average one lysine per 20 residues. Four lysines combine to form a single crosslink called a desmosine, and it is thought that all lysines in elastin can become involved in a crosslink [27], [28]. However the number of crosslinks present at any given moment is highly dependent on the level of activity and expression of the lysyl oxidase enzymes LOX and LOXL [29]. The relationship between

crosslink density and network elastic properties is therefore unknown, other than a general understanding that the distribution of desmosine crosslinks and the degree of their formation with collagen gives the ECM specific bulk elastic properties. The degree of elastin's inherent stretch is determined by how closely associated (crosslinked) it becomes to collagen meaning that the elastic properties of the ECM can be finely regulated by the formation and breakage of elastin-collagen lysine crosslinks [26]–[28].

Figure 1.4 (A): The structure of Elastin. Adapted from Urry et al, 2002.

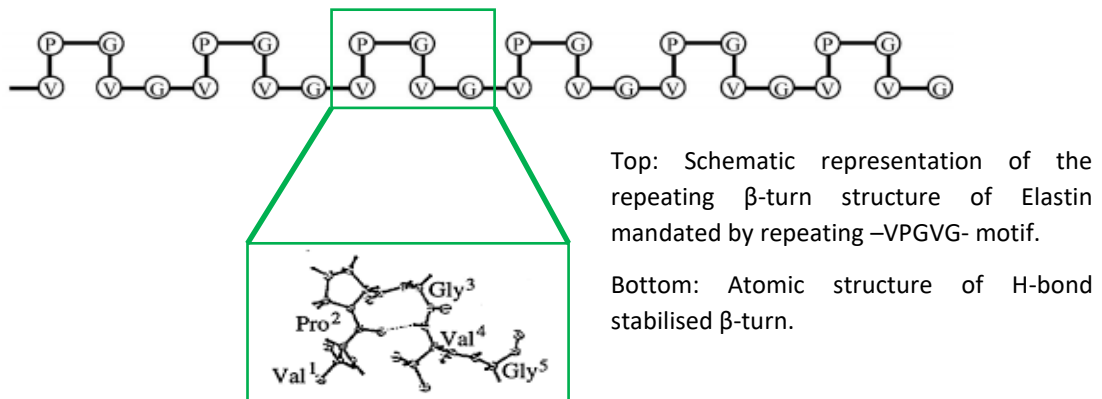
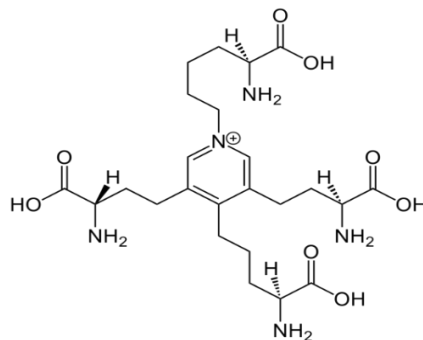


Figure 1.4 (B): A desmosine crosslink formed by the  $\epsilon$ -amino groups of 4 lysine residues.



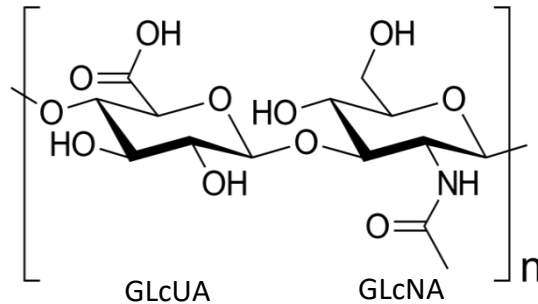
The final major fibrous component of the ECM is fibronectin (FN). The role of FN is as the key mechanosensor of the ECM, relaying mechanical signals both from cell traction forces and from the collagen network to induce both enzymatic remodelling of the network by way of



initiating intracellular signalling cascades, and cellular adhesion and migration around the network as part of tissue remodelling. FN is a modular protein with monomers being 230-270 kDa depending on splice variants [30]. FN is the key binding and signalling constituent the ECM, containing binding sites for collagen, integrin, tenascin, and multiple other cell adhesion sites. FN is composed of three domain types organised into repeating units (Figure 1.5B). Type III modules are 7-stranded  $\beta$ -barrels, and types I and II contain disulphide bonds to stabilise the structure and prevent total mechanical unfolding under stress, which is crucial to FN's role as a mechanosensor [31]. FN monomers associate into dimers by way of disulphide bond formation between two C-terminal cysteines [32]. Once fibrillated, FN acts as a mechanosensory bridge between the collagen ECM network and cells, binding to collagen via modules I<sub>6-9</sub> and II<sub>1-2</sub>, and various cell surface proteins. The key cell adhesion domain are modules III<sub>9-10</sub>, 10 containing an Arg-Gly-Asp (RDG) motif which binds  $\alpha$ 5 $\beta$ 1-integrin, and 9 a synergy site consisting of a His-Ser-Arg-Asn motif [33]. Both these motifs are partially buried in the folded state but under lateral stress FN is able to undergo a degree of conformational change prior to rupture, stretching several times over its length [34]. This stretching exposes both sites for integrin binding, and in this way FN acts as a mechanosensor for cell adhesion to the ECM. Several other binding sites have also been tentatively identified for molecules such as heparin, tenascin-C, and fibrinogen, but almost all are cryptic [30], [32], [35].

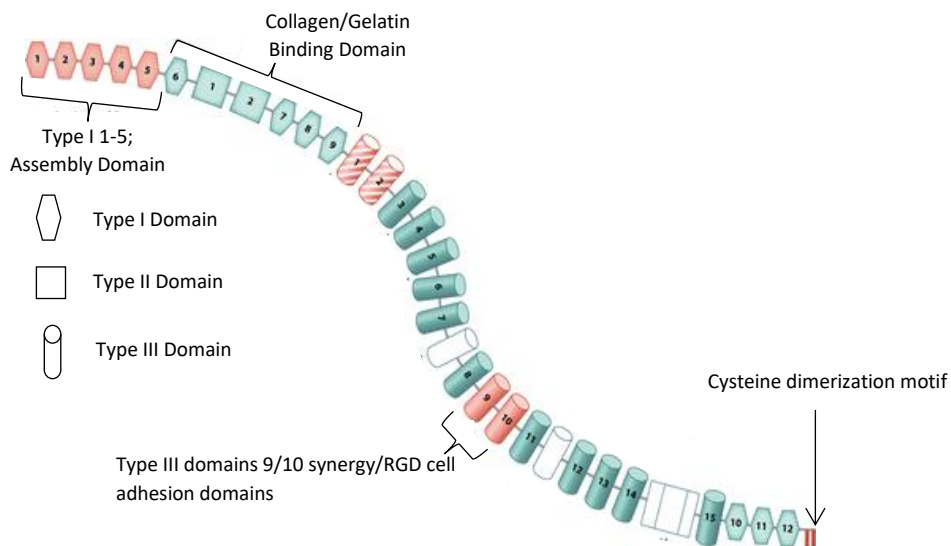
Figure 1.5 (A): Chemical structure of Hyaluronan chains.

The molecular weight of chains varies between  $5 \times 10^5$  to  $5 \times 10^6$  Da [272]. The repeating dimer unit is composed of D-glucuronic acid (GlcUA) bound to N-acetyl-D-glucosamine (GlcNAc) via a  $\beta 1,3$  or  $\beta 1,4$  glycosidic bond [273].



=====

Figure 1.5 (B): Illustration of the structure and binding motifs of Fibronectin. Adapted from Singh et al, 2010.



Remodelling and maintenance of the ECM is continuous. Fibroblasts are the cellular species most associated with the ECM, as they secrete both its primary component (collagen) and the enzymes which cleave it; matrix metalloproteinases (MMPs). MMPs of various types cleave all three of the main ECM fibrous proteins (collagen, elastin, and fibronectin), and are therefore able to breakdown the ECM network [36]–[38]. This degradative effect is counterbalanced by MMP inhibitors secreted by surrounding tissues and organs. In this way

an equilibrium is established between the fibroblasts which constantly secrete ECM and MMPs, and the surrounding tissues which regulate the level of MMP activity by way of MMP inhibitors such as TIMPs [39]. The actual components of the ECM are therefore constantly being turned over, whilst the overall network topology remains the same. Tissues also secrete lysine crosslinking enzymes LOX and LOXL, and can therefore regulate the elastic properties of the ECM by regulating the degree of elastin crosslinking [40]. All this adds up to a finely balanced regulatory system.

### 1.1.3 Synthetic Hydrogels

#### 1.1.3.1 3D Cell Culture, Wound Dressings, and Implants

The development of highly accurate biomimetic models for the purpose of *in vitro* disease research and drug development is extremely complicated and difficult. This is because the ECM in which all mammalian cells have evolved to exist is extremely complex with cellular behaviour effected by almost every aspect of the surrounding microenvironment including substrate stiffness, elasticity, and the biological molecules present [41]. Therefore creating models which can be tuned to accurately mimic the mechanics of various ECM tissues as well as their biological composition is a crucial research aim in the drug discovery industry, as the quality of this mimicry directly affects the rate and accuracy of drug discovery [42]. This is why the world is interested in hydrogels.

As has been discussed earlier, hydrogels are conceptually very similar to the ECM, both being solid network structures containing well over 95% water as a mass fraction. The ECM has a crucial role in providing mechanical, chemical, and binding cues to the cells and tissues it encapsulates, and in its absence cells and tissues will not behave as *in vivo*. Replicating a tissue or cell phenotype *in vitro* in a tissue model is therefore reliant on providing the correct ECM-

derived signals, and by extension upon building a highly biomimetic synthetic ECM. The culturing of stem/mature cells and their subsequent directed differentiation/growth is the basic requirement of a synthetic ECM. Towards this end billions of pounds has been spent on creating novel hydrogel systems and characterising the cells and tissues which can be persuaded to grow in them, and trying to improve their biomimetic accuracy [43].

This has led to a clear evolution in 3D hydrogel cell culture technology; initially plastic polymer systems such as PVC and PEG dominated research and to this day are still being developed. The reason for this is simple; they are cheap to produce, and their network topology can be precisely controlled due to their chemical simplicity. This creates clearly defined relationships between volume fraction and crosslink density and macroscopic mechanical properties and pore size distribution [44]. Plastics (hydrocarbon polymers) are easy to understand and manipulate chemically, and they have had significant success as hydrogel matrices. The most commonly used plastic polymer has been polyethylene glycol (PEG) due to its low immunogenicity and prior FDA approval for use in certain *in vivo* medical applications [45]. PEG and other plastic polymers have been used as the network species for hydrogels capable of supporting pancreatic  $\beta$ -cells [46], vascular invasion as a precursor to tissue formation [47], and bovine chondrocytes [48]. These examples all exhibit different crosslinking mechanisms allowed by the easy incorporation of different chemical moieties into the polymer, which all affect the behaviour of cells within them. This illustrates the success of plastic hydrogels, but because they are not natural products found in the mammalian body they lack the specialised cues associated with scaffold-cell interactions such as integrin binding and MMP remodelling. This led to the next technological leap; the incorporation of short peptides containing cleavage, binding, and other cell-interactive sequences either as the network-forming species or tethered to a plastic matrix.

The incorporation of peptides opened up a far greater design space for tailored erosion rates and responsiveness to environmental change such as pH, enzymatic degradation, and protein-protein interactions [49]. These properties have led to significant successes in the design of ECM-mimicking hydrogels capable of supporting encapsulated cell growth and differentiation in response to external environmental stimuli and certain cell-derived cues. Examples of incorporated peptides aiding tailored cellular morphogenesis include the encapsulation of neural stem cells in a physical protein hydrogel formed by the self-assembling peptide (RADA)<sub>16</sub> conjugated to an –IKVAV motif associated with neural cell attachment. This was transplanted into the site of traumatic brain injuries to aid neural regeneration and reduce cerebral cavitation caused by tissue death [50]. The use of multiple peptide sequences conjugated to a 4-arm PEG norbornene lattice created a multi-responsive gel; a KCGPQG↓IWGQCK MMP-cleavage sequence allowed the network to be remodelled by cells, and cellular adhesion was mediated by a CGRGDS fibronectin-binding sequence. These hydrogels were then used to culture and directionally differentiate human mesenchymal stem cells into chondrogenic, osteogenic, and adipogenic lineages [51]. A final example is the use of polyvinyl alcohol (PVA) incorporating the integrin-binding peptide RGDS, supporting attachment and spreading of fibroblasts. Firstly the mechanics of the gel were highly tuneable as a function of PVA fraction volume and the number of crosslink sites per unit. Secondly the incorporation of the RGDS peptide in a dose-dependent manner increased the cell viability by up to a factor of 4 within the gel for up to 2 weeks [52]. These examples illustrate the advantages of plastic matrices, but that complex cellular cues can only be imparted or received via peptide sequences.

More recently hydrogel technology has moved towards incorporating both peptides and natural polymers such as collagen and hyaluronan into their networks. The reason for this is

obvious; 3D cell cultures aim to mimic the ECM, and the ECM is made of collagen and hyaluronic acid. Therefore making a hydrogel out of these molecules should provide the best model of the ECM. This may or may not be true but has certainly yielded encouraging results. For example peptide/natural polymer hydrogels have been tailored for use in joints as a regenerative treatment for damaged hyaline cartilage, which is composed of chondrocytes and a dense ECM. Hyaluronic acid was used as the lattice backbone, with various collagen species encapsulated to aid the regeneration of cartilage in damaged joints [53]. Further examples include a lysine-crosslinked peptide containing a MMP II cleavage site sensitive to degradation in certain tissues aiding localized cell proliferation [54], a poly(glycolic) acid (PGA) gel incorporating collagen I for use in dental pulp regeneration [55], and multiple constructs aimed at accelerating open wound healing by stimulating increased tissue regeneration [55][56]. These are just a few isolated examples of how hydrogels have been utilised as tissue engineering scaffolds, and serve to demonstrate their clinical potential and importance.

The ECM interacts with cells via binding motifs and cleavage sites. These can be incorporated into hydrogels simply by conjugating short peptides to the network forming backbone. The utilisation of such short peptide sequences has been a major step towards better 3D cell cultures [57]. However the mechanical properties of the ECM are equally as important in determining cellular morphology, and understanding and tuning this rationally is a major hurdle which must be overcome if a perfect 3D cell culture is to be developed. The mechanical properties of plastic polymer hydrogels are generally defined by their volume fraction; add more network molecules and it becomes stiffer [44]. Stiffness can also be mediated by altering the crosslink density be they covalent, physical or ionic, and is easy to achieve in plastics [58]. However, as has been discussed earlier, plastic polymers can never be perfect biomimetics because they do not exist in the human body and cells and tissues cannot interact

with them. Organic polymers such as hyaluronan and collagen have also been used to form biomimetic hydrogels which has allowed a more complex system of interactions to exist between cells and the matrix, but our understanding of how to tune the mechanical stiffness of such gels is poor. In general terms collagen/hyaluronan gels exhibit the same relationship between volume fraction and stiffness as plastic gels, but the effect of changes in crosslink density is still unclear and difficult to investigate. In the case of collagen-based hydrogels, the length of the collagen polypeptides has a significant effect on mechanical properties, as longer triple helix tropocollagen fibrils leads to a higher degree of entanglement. Lysine crosslinking has been utilised to simulate the desmosine formation present in the ECM to chemically crosslink collagen fibres together [59][60]. These models have had significant success but their macro-mechanical design is still course grained. Contributory factors such as increasing fibre length, volume fraction, and hyaluronan : collagen ratio are known, but it is not understood how these relationships are defined and the full range of interplay between them. This means that at the time of writing a perfect 3D cell culture system does not yet exist using the technologies discussed above, and our ability to build a perfect simulation of the ECM is still lacking [61].

#### 1.1.3.2 Other Novel Hydrogel Applications

Whilst hydrogels have been developed most intensely for clinical applications for the reasons discussed in Section 1.1.3.1, their natural properties have led to their use for a range of non-clinical applications. Hydrogels are defined by the marriage they represent between the aqueous and the solid phase; they are mostly aqueous water through which molecules can freely diffuse, but are solid and contain a network which can incorporate any number of

chemical functionalities. This gives hydrogels a massive potential design space; if a process can take place in an aqueous environment and requires an interaction with some kind of solid phase substrate then it can likely be incorporated into a hydrogel.

The term hydrogel was first coined in the terms of the patent for the first contact lenses in 1960 [62]. Since then contact lenses have been developed which are more durable, less irritating to the eye, and less of an infection risk. The swelling ratio and transparency of a contact lens are the most important aspects of their design as it is these which allow them to be tuned to any prescription and revolutionised optical treatments [63]. This aspect of their function has more or less been optimised as exemplified by their ubiquity in advanced healthcare [64]. What has yet to be optimised is the prevention of persistent eye infections caused by extended lens use. As one would predict, placing a foreign body in one's eye dramatically increases the risk of eye irritation or pathogenic infection [65]. This has led to continued development of novel hydrogels with properties to reduce the risk of infection for use as contact lenses. One example of a novel hydrogel system for contact lenses is a PDMS hydrogel with a top layer of PEGMA. The PEGMA layer is resistant to protein absorption and thereby acts to reduce the potential for bacterial survival on the lens. The PEGMA layer has high oxygen permeability creating a bactericidal environment at the lens surface. All these novel properties make this system potentially useful in reducing eye infections associated with lens use [66]. A second example illustrates the crossover between 3D cell culture advances and contact lens technology. Peptides can be used to stimulate eukaryotic cell health, but can conversely be used to promote prokaryotic cell death. The bactericidal peptide KRWWKWIRW was tethered to a PEG lattice. This rendered the hydrogel inhospitable to >99.9% of bacteria and represents a promising mechanism of maintaining lens sterility long-



term [67]. Contact lenses have been the most successful use of hydrogels to date, and are still undergoing dramatic technological advances.

Another more recent utilisation of hydrogels has been as novel filters. Filtration by definition involves the removal of one or more solutes from a solvent. Removal of specific solutes requires some form of specific interaction with the filter membrane, and a filter membrane through which water can move freely. Traditional plastic membrane filters are generally defined by their pore size and filter water by size exclusion of solutes. Hydrogels in contrast can incorporate almost any chemical moiety and can therefore be designed to interact in very specific ways with the filtrate. One such example is the design of bactericidal filters. Access to clean drinking water is one of the most fundamental of human rights and vital to quality and continuation of life. In both highly industrialised and less industrialised nations the process of water treatment requires bactericidal processes, which require the addition of harmful chemicals such as chlorine dioxide and monochloramine that can damage public health if not properly controlled [68]–[71]. Hydrogels have been developed which contain bactericidal chemicals, for example a graphene oxide-silver hydrogel has been developed which kills 100% of bacteria passing through it up to  $10^5$  cells/ml. This could be used as a chemical-free final disinfection stage, or as a one-step bactericidal filtration for untreated water [72]. Hydrogel filters have also been developed for the separation of solvents, for example for use in oil spill clean-up; a PVA hydrogel crosslinked with glutaraldehyde was used to coat a standard PDMA filter paper. Once saturated with water the hydrogel creates a water emulsification barrier through which the water could flow but any form of oil could not. This allowed the separation of oil from water at >99% efficiency, and was stable in up to 8M sulphuric acid, 10M sodium hydroxide, and saturated concentrations of sodium chloride [73]. These two examples

illustrate the potential for hydrogel design to filter out/separate any number of solutes/solvents, and their wide range of industrial and personal applications.

Other novel applications of hydrogels have included as sensor platforms. By incorporating various moieties into a hydrogel either as part of the lattice or encapsulated in its pores has allowed quantitative calorimetric measurements for various molecules. For example L-Glutamate oxidase was encapsulated within a polycarbamylsulfonate hydrogel. When immersed in a solution containing L-Glutamate, a colour change occurred as a result of the enzymatic degradation of L-glutamate + O<sub>2</sub> + H<sub>2</sub>O → α-oxoglutarate + NH<sub>3</sub> + H<sub>2</sub>O<sub>2</sub>. An absorbance measurement at 492nm was then used to calculate the concentration of α-oxoglutarate in the solution, with a suggested application in the testing of levels of MSG in soy sauce. A second example of a highly complex hydrogel sensor platform is a simple agarose hydrogel embedded with chitosan carbon dots. These carbon dots are able to chelate heavy metal ions and induce a photometric change. Investigation of the reflectance spectra of the hydrogel before and after exposure to a solution containing Cr<sup>6+</sup>, λ = 380 nm, Cu<sup>2+</sup> λ = 290 nm, Fe<sup>3+</sup> λ = 360 nm, Pb<sup>2+</sup> λ = 215 nm, and Mn<sup>2+</sup> λ = 250 nm, allowed the accurate quantification of the concentration of each of these ions [74].

#### 1.1.4 Folded Protein Hydrogels

##### 1.1.4.1 Dynamic Activity and Smart Biomaterials

Section 1.1.3 discusses how hydrogels have been used so far and the materials from which they have been built. Chemical functionalities have been incorporated in polymer matrices, and organic polymers have been utilised to provide cellular signalling cues. All these technologies have shown promise in certain fields and applications and it is likely that all will undergo continued research. However few of these technologies incorporate what is perhaps

the most remarkable capability of the molecular scale world; the enzymatic activity and dynamic mechanics of folded globular proteins (FGPs). Every process which keeps an organism alive is performed or regulated by a protein exhibiting secondary structure in isolation or combination. The number of different catalytic, synthetic, transport, mechanical, and signalling processes in nature are beyond imagination. Added to this, similar processes in different organisms are often carried out by utterly different proteins. This scale of diversity illustrates what makes folded globular proteins so powerful; any biological function one can imagine, a protein likely exists that can facilitate it. This means by extension that if one understood how to design or isolate proteins to perform specific tasks, we could in theory apply them to improve every industry and life on the planet. Towards this end the cutting edge of biological research has for decades become the characterisation and understanding of how to get from primary sequence to tertiary structure and function [75].

Hydrogel research over the last ten years or so has similarly begun to shift in focus towards the incorporation of FGPs in order to create a new generation of highly biocompatible and functionally diverse smart biomaterials [76]. By integrating FGPs into hydrogels their myriad enzymatic functions can become a dynamic functional property of the hydrogel. For example the entrapment of L-Glutamate oxidase to make a hydrogel sensitive to L-Glutamate concentration [77]. The potential enzymatic functions are limitless, though finding uses for them in industry is still an ongoing process. FGPs are sensitive to environment in ways that polymers and natural secondarily-unstructured polymers such as collagen and hyaluronan are not. FGPs can be extremely sensitive to pH, temperature, salt, and most notably ligand binding, all in a concentration dependent manner yielding a scale of response magnitude [78]–[81]. Most such changes are reversible. These factors and more can all alter FGP structure, shape, and mechanics. This ability to respond to environmental change or the

presence/absence of various molecules means that FGPs can alter a hydrogel's properties/activity in a smart responsive manner. The entrapment of proteins in a hydrogel or their tethering to the lattice can imbue the material with enzymatic functionality, but does not allow full incorporation of their sensitivity to environmental stimuli. The incorporation of responsive FGP mechanics requires hydrogel lattices to be built from them, and is discussed in depth in Section 1.4.7.

It is important at this stage to make a distinction for the purposes of this thesis between types of protein hydrogel. Hydrogels which consist of plastic-polymer or peptide networks with proteins entrapped in the water-filled pockets, or entangled/tethered to the polymer lattice, will be termed pseudo-protein hydrogels as the proteins do not have a structural role in the hydrogel lattice. Hydrogels built solely from a polymeric folded protein lattice will be referred to as true protein hydrogels. This distinction must be made in order to distinguish the majority of protein-containing hydrogels described in the literature from the protein-lattice hydrogels which are described in this thesis.

Development of hydrogels built exclusively from FGPs has over the last 10 years begun to accelerate. This next generation of smart biomaterials will hold an advantage over previous species of hydrogel; they will be more tuneable and smartly responsive to environment and the presence/absence of various molecules. They will be mechanically unique as not only will stiffness be mediated by volume fraction and crosslinking density, but also by the elasticity and dynamic unfolding-refolding equilibrium of the protein domains. Finally and most crucially they will have infinitely more and diverse catalytic activity, which will allow their adaptation to perform new tasks and processes which may have been impossible before or required the use of harmful and expensive chemical processes. The diversity of folded protein

hydrogels could be as great as that of FGPs in nature. The disadvantages are also significant however. FGP hydrogels will be more expensive to produce as design, expression, and purification of recombinant proteins on an industrial scale is vastly more expensive than plastic production or collagen extraction, although this is improving rapidly [82][83]. The rational design of their mechanics will be far more complex as the varying mechanical properties of different protein domains and crosslinking geometries will require optimisation on an individual basis. Finally they will likely degrade faster than plastics because proteins have not evolved to persist for long, with protein half-lives *in vivo* rarely exceeding days [84], [85]. Time will tell if protein hydrogels find an industrial or healthcare niche, but it is certain that those which are developed will be more dynamic, tuneable, and adaptable than any that have come before.

#### 1.1.4.2 The Rheology of Folded Protein Hydrogels

As discussed above the incorporation of FGP's into a hydrogel has the potential to imbue the material with a choice of dynamic properties. These can include catalytic or synthetic enzymatic activity, ligand and environmental sensitivity, and sensitivity to specific enzymatic degradation. These properties have the potential to make true-protein hydrogels smarter and more responsive than anything that has come before. In addition to these chemical and enzymatic capabilities folded proteins have far more complex micro-scale mechanics than synthetic polymers, peptides, or even disordered organic fibres such as collagen and hyaluronan [86]. Static polymers do not exhibit complex microscale mechanics because they contain no 3-dimensional structure beyond a tendency to coil under certain conditions, and can be described as semi-flexible chains. This gives them a persistence length, a bending modulus, and in some cases a lengthening under stress value, but little elastic or responsive

capability [87], [88]. In contrast folded proteins have elastic, viscous, and dynamic micromechanical diversity which we are only beginning to understand.

A single polypeptide chain could adopt billions of discrete 3-dimensional conformations, but will ultimately settle into the one with the lowest entropy in a specific environment. How they are able to adopt the most energetically favourable conformation without sampling all other possibilities still defies our understanding. The sequentially-determined fold is usually stabilised by a complex system of inter-residue hydrogen bonds, disulphide bridges, the hydrophobic effect, and other less specific ionic interactions. This means that the final folded structure of a protein is an entropic energy minima, and is maintained in its shape by the inherent packing of its side chains and multiple inter-residue bonds. In essence polypeptide chains are able to spontaneously transition from a 2-dimensional semi-flexible chain little different to a polymer, to a specific 3-dimensional shape resistant to rearrangement and with a unique bond network holding it in a folded state. This makes proteins viscoelastic on the microscale, and inherently mechanically complex [89].

FGPs exist in an entropic energy minima and therefore require energy input, be it chemical kinetic or thermal, to alter or destroy their structure. As discussed in Section 1.4.3, in order to move a structure into an adjacent energy minima state sufficient energy input must occur over a small enough timescale to overcome the free energy barrier between the two states [90]. In the case of protein folds this means that the secondary structure networks of proteins can store and re-emit kinetic energy, allowing the structure to resist permanent mechanically-induced damage and recover in an elastic manner. In this way FGPs are viscoelastic on the microscale to a degree not shared by synthetic polymers or other organic chains. The mechanical characteristics of FGPs are as varied as those found on the macroscale; some are

stiffer than others, some are more highly elastomeric with a greater or lesser degree of recoverable 3-dimensional rearrangement, some are non-mechanically labile thanks to inter-residue covalent bonds, and some exhibit higher viscosity caused by internal friction between chains [91]. Hydrogels built from FGP lattices are therefore predicted to have distinct microscale mechanical properties which can be rationally tuned by the choice of building block and chemical environment. These will translate to the macroscale bulk mechanics of the gel, and will be distinct and more complex than the basic volume-fraction and crosslink density laws which govern polymer and fibrous hydrogel bulk mechanics [92].

The macromechanical properties of a hydrogel are in essence determined by its ability to spread stress throughout the entire network and prevent localised breakdown. This is a function of the degree of interconnectedness across the network, or in other words the crosslink density. Mechanical force (stress) is propagated through network structures and bonds, and the more there are, the higher the energy storage capacity of the network. In the case of FGP hydrogels this is further complicated by the relationship between crosslink density and the elastic response of the folded protein domains. The crosslink density will act as a function of two processes; the ability of the network to spread and store stress, and the efficiency of the network to re-emit this energy and recover elastically. Therefore understanding the relationship between the crosslink density of an FGP hydrogel lattice and the macromechanical gel properties is a crucial aspect of the rational design of protein hydrogels. It has been demonstrated that the nature of the protein building block used affects hydrogel mechanical properties [93], [94]. It is predicted that the crosslink density will also affect these properties but has so far not been rationally demonstrated. Nor have the laws governing the translation of microscale network interconnectedness to bulk mechanical properties been rationally investigated. Characterising the relationship between microscale

crosslink density and macroscale hydrogel mechanics is the subject of this thesis. It has been demonstrated that alterations to the microscale mechanics of FGPs either by ligand binding or chemical denaturation translates to the bulk micromechanics of the hydrogel [95], [96]. The rational design of these mechanics is still coarse grained however; large alterations can be induced but poorly predicted, and the microscale alterations in network topology which lead to them are not understood. As a result FGP hydrogel mechanical design sits at an exciting stage; its is clear what can be done, but understanding how to predict and design the mechanical properties of the gel is crucial before rational design can become possible.

#### 1.1.4.3 FGP Hydrogel Network Design

The design of FGP hydrogels is still in its infancy. The stages of progress which have so far been achieved can be summarised as follows; firstly proteins have been crosslinked together in the absence of any additional physical linker species to form a solid hydrogel at 1-10% w/v [49], [97], [98]. Secondly it has been demonstrated that the macromechanics of a protein hydrogel can be altered by changing the volume fraction, the protein used, and the chemical environment [99]–[101]. Furthermore it has been shown that a fraction of protein is able to remain folded post-gelation, and that this can be altered by chemical environment [102]. The list of advances not yet achieved is far longer; precise measurement of the degree of folded protein post-gelation, measurement of the efficiency of crosslinking and the subsequent characterisation of the effect of crosslinking density on gel mechanics, characterisation of the chemical and environmental durability of FGPs as part of hydrogel networks over extended periods of time, and observing the ability of FGP hydrogels to promote encapsulated cellular growth/differentiation. Finally and perhaps most importantly



demonstration of the continued functional activity of an FGP post-gelation has yet to be demonstrated.

Aggregation of proteins is easy to achieve at high concentrations, and historically avoiding this has been a key research aim in the development of recombinant protein products [103], [104]. Protein aggregation is undesirable because it almost universally results in an increase in unfolded fraction and a loss of function, both of which are equally undesirable in an FGP hydrogel, as is the potential increase in immunogenicity [105]–[107]. Therefore physical crosslinking methods are inappropriate for FGP hydrogels as an environmental change resulting in gelation will likely be due to aggregation. Instead chemical crosslinking methods have become the norm using the chemistry of amino acid side chains to form covalent crosslinks directly between proteins. The methods by which this has been achieved are discussed in Section 1.2.3, and several mechanisms have been described which result in the polymerisation of protein monomers to form a hydrogel network [94][108][109]. The networks formed via these methods have been partially characterised; SEM imaging shows variations in pore size, SANS has been used to measure the particle size distribution and the fractal dimension, and fluorescence to demonstrate the presence of crosslinks [94][95]. All these observations and measurements so far lack context as no understanding exists regarding their relationship to crosslinking density. Crosslink density is a complex number as it comprises information of multiple parameters; the efficiency of crosslinking, the crosslink geometry, and the total number of crosslinks formed. Understanding this is crucial to understanding how to rationally design network topology and by extension the gels mechanical properties. The presence of folded protein post-gelation has been demonstrated by three distinct methods; chemical labelling, circular dichroism, and chemical denaturation [95][96], [110]. However a precise measurement of the folded fraction of protein, and

investigation of the thermodynamic stability of FGPs in a network has yet to be performed. Until these parameters are fully understood rational design of smarter biomaterials will be limited.

## **1.2 Hydrogel Networks**

All synthetic hydrogels begin as solutions with constituent molecules diffusing freely in a fluid environment which will flow to fill its container. In order to transition into a solid hydrogel state these molecules must associate into a network dense enough to create an average pore size small enough to prevent the diffusion of macromolecular structures whilst allowing the continued diffusion of water into the external environment. The surface tension of the water therefore acts as the primary barrier between the water encapsulated in the gel and the external environment, but achieving a suitably high surface tension is reliant upon the network being dense enough and stiff enough. Once a network of sufficient density is achieved, the water is unable to flow outside the network and the solution has become a solid material and will maintain a certain shape. The mechanical properties of the gel are defined by the ability of this network to resist and recover from deformation by mechanical stress and maintain a network continuous and dense enough to retain the water within it. If sufficient stress is applied the network will break down and the water will again be able to flow and the viscous component of the complex modulus will exceed the elastic modulus.

There are two mechanisms by which a network capable of forming a hydrogel can be built; the polyionic interaction or entanglement of one or more species of molecule to form a physical gel network defined by transient interactions, or via the permanent chemical crosslinking of molecules to form a continuous and far less dynamic network [58]. Natural hydrogels are formed by combinations of physical interactions and crosslinking. There are no

exclusively chemical natural hydrogels in biology. Organic molecules such as collagen, hyaluronic acid, and alginate exhibit ionic and physical interactions at high concentrations, certain pH's, and in response to thermal stimuli in order to undergo a sol-gel transition *in vitro* [111][112][113]. Synthetic hydrogels have been built using both mechanisms; agarose electrophoresis gels are formed by the heat-induced molecular entanglement of agarose molecules to form a physical gel [114]. In contrast polyacrylamide gels are formed by the incorporation of crosslinking bisacrylamide subunits into growing chains of acrylamide monomers to form crosslink bridges. The unique properties of both classes of hydrogel have multiple industrial and therapeutic applications, and the bulk rheological properties of both are highly tuneable and useful.

### 1.2.1 Physical Hydrogels

Hydrogels are formed by the polymerisation of either homo- or heterogeneous monomeric subunits. Most synthetic gels require covalent crosslinking between these monomeric units but this is not always necessary, as a class of hydrogels known as physical hydrogels can be formed from a network of physical entanglements and/or ionic interactions using a mixture of anionic and cationic monomers (Figure 1.9) [2][58]. These structures are non-homogenous, with areas of high and low lattice density, and are highly sensitive to alterations in ionic strength and temperature. For this reason physical gels have been used as environment-sensitive delivery systems for everything from drugs to flavour oils [6], [115], [116].

Physical hydrogels often have far more viscous rheological properties due the transient nature of the interactions which form their networks. The mechanical properties of hydrogels are defined by the relationship between the strength of the interactions which form their network, the reversibility or otherwise of these interactions, and the degree of

interconnectedness which defines the timescales upon which that network can rearrange in response to mechanical stress and the degree of energy dissipation under strain [117]. In the case of physical hydrogels their network is formed either by molecular entanglement (making the rate at which they can rearrange dependent on the coefficient of friction between interacting molecules), transient ionic interactions such as van der Waals forces, or ionic bonds [118], [119]. All three of these interaction species are to a greater or lesser degree transient, meaning that the network is fundamentally more dynamic as these monomers are able to diffuse, continuously breaking and forming bonds at a defined rate. A physical network therefore does not possess a constant force dissipation pathway, meaning that force cannot be distributed across the entire network efficiently and is concentrated far more locally and not distributed globally. The result of this is that the viscous component of a physical gel has a far greater effect on the complex modulus. A physical bulk network is generally less elastic as stress is localised and the viscous component permanently dissipates more energy as heat due to internal friction, meaning the gel is less able to recover after stress. This makes physical gels generally weaker than chemical ones, with lower storage moduli, a greater degree of hysteresis as energy is lost during permanent network alteration under strain, and a response to lower frequencies of stress as the viscous properties slow down the recovery of strain.

This, combined with the sensitivity to environmental pH and temperature, makes physical hydrogels relatively short-lived and highly sensitive to environment. DNA agarose gels for example will not form 1 pH unit either side of its optimum pH 8, and will dissolve within hours if immersed in pure water [114]. This is due once again to the nature of the interaction allowing the formation of the network; ionic interactions generally depend on the ratio of the protonated/deprotonated state of various chemical groups. The pH environment determines the likelihood of a chemical group being protonated or not, and therefore an alteration in pH

will affect the ratio of protonated: deprotonated groups logarithmically. This can allow a very subtle change in pH environment to ablate ionic interactions as the protonation state ratio of the bonding groups changes by an order of magnitude. In a similar fashion an increase in the thermal energy of a system of non-covalently bound molecules will lead to an expansion and dissipation of the network as it melts. Gels made from very weakly associated molecules such as cellulose/acrylic acid are able to undergo this transition at temperatures associated with inflammation and have therefore been used as drug delivery vehicles [120].

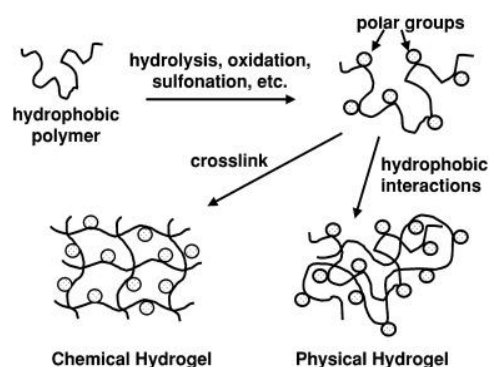
Certain molecules also have the novel property of undergoing an alteration in solubility in response to temperature. Whilst almost all solutes will crash out of solution at a suitably low temperature, some molecules with specific ratios of hydrophilic:hydrophobic groups behave inversely becoming less soluble as temperature increases. This is another mechanism which has been exploited as a delivery method for drugs and other molecules. This environmental sensitivity has made physical hydrogels a prime candidate for certain drug delivery systems, as their ability to encapsulate and protect a drug, then release it in response to a specific pH or temperature makes them ideal for site-specific drug delivery *in vivo* [121], [122][6].

## 1.2.2 Chemical Hydrogels

Chemical hydrogels are formed by polymerisation as monomers become covalently cross-linked to form a regular lattice structure (Figure 1.6) [56].

---

Figure 1.6: Schematic summary of the differences in lattice structure between polyionic (physical) hydrogels and chemically crosslinked hydrogels [77]. The incorporation of polar groups onto the monomer is not necessary for all monomer species.

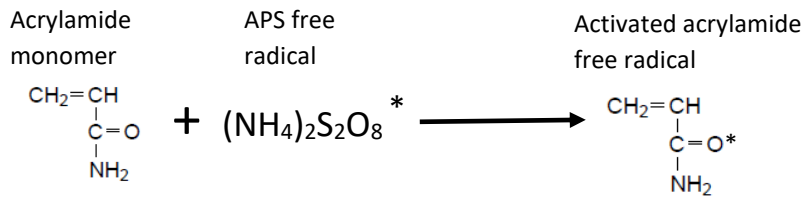


---

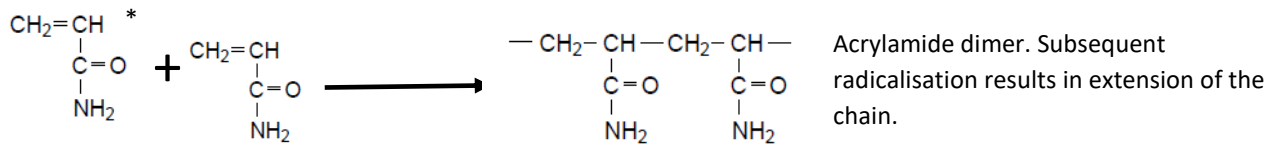
Chemical gels are generally more homogenous than physical gels with a regular crosslinked structure prevailing. The two key components of chemical hydrogel formation are a polymer in great excess, and a crosslinking agent. The most commonly used example of chemical gel formation via radical polymerisation is that of SDS polyacrylamide gels; acrylamide and bisacrylamide are the two lattice monomer species, ammonium persulfate (APS) acts as a crosslinking agent, with tetramethylethylenediamine (TEMED) additionally present to act as a catalyst to increase the rate of polymerisation. Upon dissolving in water APS forms a free radical, which in turn generates monomeric acrylamide radicals. This reaction is catalysed by TEMED which acts as an electron chaperone between APS and acrylamide. The acrylamide radical is then able to react with another acrylamide monomer to form a dimer, which is extended into a growing chain by subsequent rounds of acrylamide radical formation. These acrylamide chains are randomly crosslinked by the incorporation of bisacrylamide units to form an interconnected acrylamide lattice [122][123]. This process is summarised in Figure 1.7.

## Figure 1.7: The reaction pathways involved in (poly)acrylamide gelation.

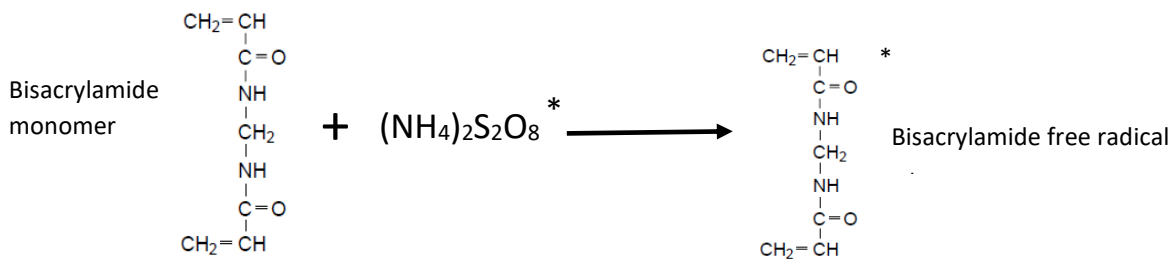
Reaction 1: Acrylamide monomer radicals are formed by APS free radicals in water. This allows the acrylamide radical to polymerise with a non-activated acrylamide monomer (Reaction 2). Subsequent radicalisation of this acrylamide chain can be repeated indefinitely to extend the length of the chain.



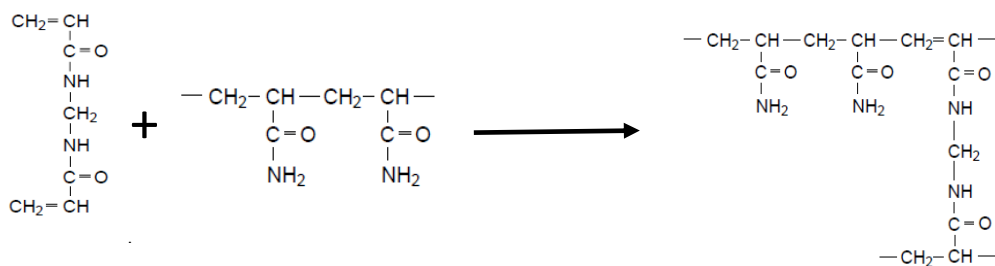
Reaction 2: Acrylamide radical polymerises with a non-activated acrylamide monomer. Subsequent radicalisation of this acrylamide chain can be repeated indefinitely to extend the length of the chain.



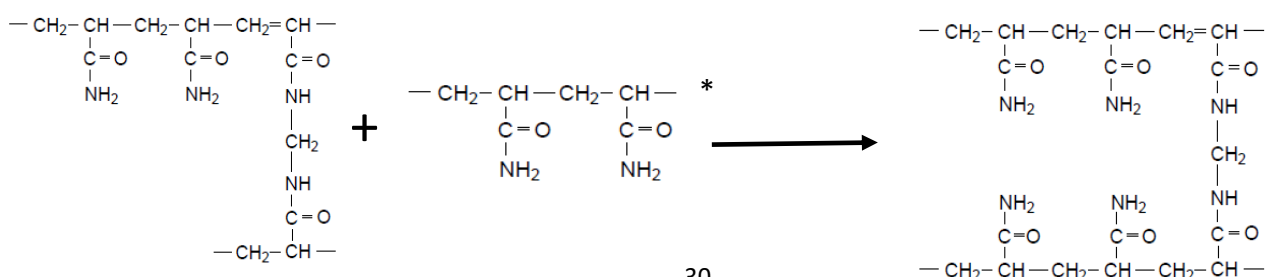
Reaction 3: Bisacrylamide monomer is radicalised by APS free radical in water.



Reaction 4: Bisacrylamide radical is incorporated as next polymeric unit into growing acrylamide chain.



Reaction 5: The un-reacted end of the bisacrylamide unit reacts with acrylamide radical species, forming crosslink between two acrylamide chains. This pattern continues until all acrylamide and bisacrylamide monomers have been incorporated into the polyacrylamide lattice, or all APS free radical species have been used up as APS free radicals cannot be recycled. TEMED acts as an electron chaperone for all electron transfers between species, thus increasing the rate of polymerisation.



This basic pattern is followed in all chemically crosslinked hydrogels; a monomeric species of the lattice backbone (either one or more species) polymerises to form long chains. These are randomly crosslinked either directly to each other via reactive groups, or via a second crosslinking monomer species [123]. This can be via reactive chemical groups, incorporation of a cross-linkable monomeric unit, or a number of other methods, yet all rely on covalent crosslinking between polymer chains. A crosslinking agent is usually employed to generate a radical lattice monomer species, though in certain cases such as poly(N-vinyl-2-pyrrolidone) (PVP) the lattice monomer species can be directly radicalised with UV illumination [124]. The favourable polymeric properties of plastics has made plastics, rather than proteins, the species of choice for hydrogel design. However building hydrogel lattices from polypeptide backbones is an intriguing notion, because it has the potential to revolutionise the tuneability of multiple hydrogel characteristics. The degradation rate of a hydrogel for example, is of crucial interest when designing a drug delivery vehicle, and the environmental parameters upon which it depends include temperature, pH, and time [122]. Plastic hydrogels can be tailored to degrade in certain regions of the body after certain periods of time in response to physiological conditions such as stomach acid or the high temperatures around the heart, but this is limited by the subtlety with which the hydrogel can sense changes in the environment. Proteins are inherently more sensitive to environmental changes, and in theory the degradation rate of a true protein hydrogel is far more highly tuneable than any crosslinked plastic. For this reason the building of hydrogels using fully-folded and active protein lattices presents the possibility of multiple characteristics which are tuneable to an extent hitherto impossible with the use of plastics alone.



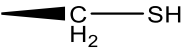
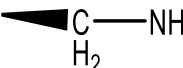
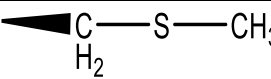
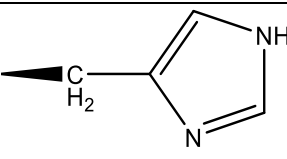
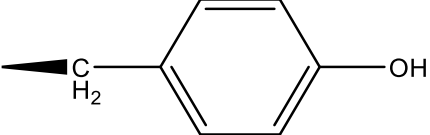
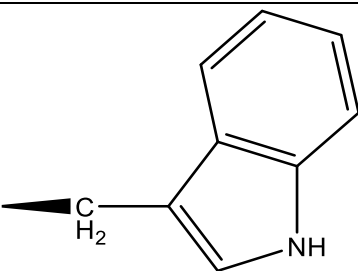
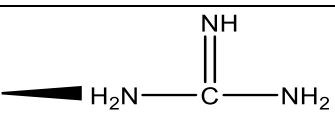
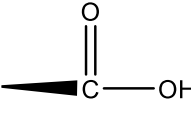
## 1.2.3 How to Form Protein Networks via Crosslinking

### 1.2.3.1 Amino Acid Crosslinking Candidates

Protein cross-linking is dependent upon the formation of covalent bonds between the side chains of their constituent amino acids. Amino acids exhibit many different side chain functional groups such as amines, sulfhydryl, and hydroxyl groups to name a few, and all can be modified and crosslinked by a range of chemistries. However no side chain species exhibits unique chemistry or an utterly distinct degree of reactivity, and so specific residue-residue crosslinking reactions are extremely challenging to achieve. Instead most commercially available crosslinking agents are capable only of favouring certain linkages [125], for example amine-to-amine linkages or thiol-thiol linkages, and so multiple crosslink species will always be generated during a reaction containing all 20 amino acids.

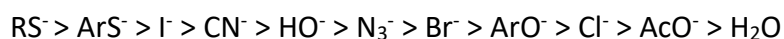
The specificity of amino acid crosslinking is simplified by the differences in reactivity between different side chains. All the alkyl side chains of the hydrophobic residues are almost chemically inert, the hydroxyl groups of threonine and serine are equivalent to water derivatives and therefore have very low chemical reactivity, and the remaining amide groups are uncharged and therefore also exhibit low reactivity. This yields only nine residues which can be considered to be highly chemically reactive. These include the guanidinyll group of arginine, the carboxyl groups of glutamic acid and aspartic acid, the sulfhydryl group of cysteine, the imidazolyl group of histidine, the  $\epsilon$ -amino group of lysine, the thioether moiety of methionine, the indolyl group of tryptophan, and the phenolic hydroxyl group of tyrosine [126]. The various modification chemistries of these residues are summarised in Table 1.1:

Table 1.1: Side chains groups of the nine most chemically reactive amino acids. Other Reactions: (A) Iodination; insertion of a single iodine atom, (B) Nitration; addition/substitution of a nitro group, (C) Diazotization; formation of an R-N=N-R azo linkage between R-N and R-OH groups, (D) Esterification, and (E) Amidation [126].

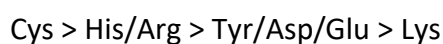
Amino Acid	Active Side Chain	Alkylation	Acylation	Oxidation	Other Reactions
Cysteine		+	+	+	A,D,
Lysine		+	+	-	C,E,
Methionine		+	-	+	I
Histidine		+	+	+	A,C
Tyrosine		+	+	+	A,B,C,D
Tryptophan		+	-	+	H
Arginine		-	-	-	G
Aspartic and Glutamic Acid		-	+	-	D,E

The choice of crosslinkable residues is reduced yet further by the hydrophobicity of both tryptophan and methionine which are usually buried in the centre of folded globular proteins. This reduces their reactivity with the solvent and sterically impairs crosslink formation. Therefore we are ultimately left with seven residues which are suitable for protein-protein

crosslinking. The suitability and specificity of each residue then becomes dependent upon their inherent reactivity; the quality of their leaving group. Most side chain modification reactions are nucleophilic, meaning that the relative reactivity is directly proportional to the nucleophilicity of their leaving group. A general order of nucleophilicity has been determined as [127]:



From this ranking we can deduce a general order of reactivity for the seven candidate crosslinking residues:



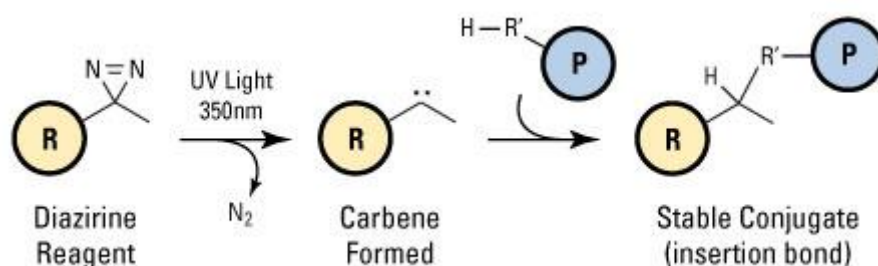
This general order is not constant and is totally dependent on the specific chemistry being used and the reaction conditions, but illustrates the potential difficulties in crosslinking a single residue species. Cysteine's dominant reactivity is such that thiol-based click chemistry has become the mechanism of choice for a wide range of analytical experiments based on protein labelling and multimerization [128]. This has included the labelling of cysteines with thiol-reactive molecules such as maleimides, and disulphide bridge formation between cysteine residues [129], [130].

An additional limitation in the design of specific protein crosslinking reactions is the need to perform them in generally physiological conditions. This is necessitated by the need to maintain the protein's folded state, as an unfolded protein is no more interesting than a semi-flexible polymer. For this reason more highly specific chemistries such as reactions with mercurials and cyanogen bromide, and the use of extreme pH is not suitable for forming folded protein hydrogels as they will likely yield a significant unfolded fraction post-gelation.

The most successful method for increasing target residue specificity whilst avoiding the use of harmful catalysts or additives has been photo-initiated mechanisms. With the addition of photo-reactive crosslinking agents such as diazirine-derived molecules it is possible to promiscuously crosslink any amino acid species to any other [131], [132]. This is achieved by the generation of a highly reactive carbene species via photolytic cleavage of an  $N_2$  molecule from the diazirine molecule. This carbene species is then able to insert itself into any close C-H bond [133]. This mechanism is summarised in Figure 1.8 [133]:

---

Figure 1.8: Reaction mechanism summary for diazirine-mediated photocrosslinking.



Mechanisms such as this can be carried out at physiological conditions and do not result in permanently radicalised, oxidised or reduced proteins meaning they generally remain folded. However the promiscuity of such reactions is difficult to reduce as the high reactivity of carbenes means they can react with the chiral carbon atom of all amino acids, though lysine is the preferential species [134]. The solution to this promiscuity therefore lies in reducing the overall reactivity of the activated photo-catalyst, leading to preferential reaction with the more highly reactive amino acids. The reactivity of an amino acid side chain is proportional to its degree of nucleophilicity, which is a function of the strength moment of the permanent dipole of their leaving group. An extreme dipole can be generated either by a significant

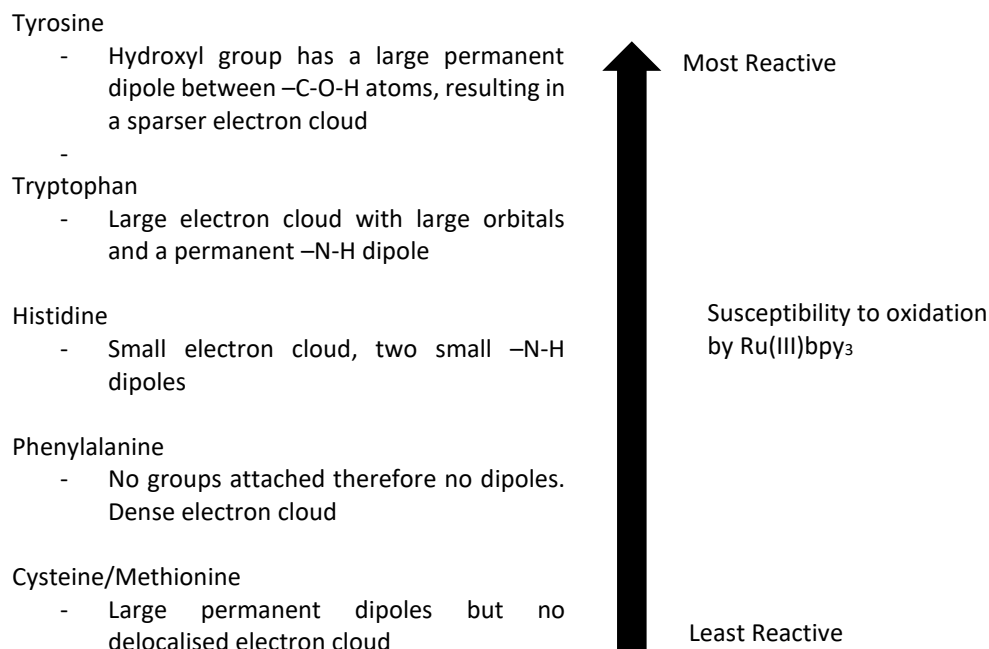
difference between the electronegativity of two bonded atoms, or by a delocalisation of electrons into an electron cloud as part of a ring structure. The prime examples of this are cysteine ( $-S^{\delta-}-H^{\delta+}$ ), methionine ( $-S^{\delta-}-CH_3^{\delta+}$ ), and the aromatic residues tyrosine, tryptophan, phenylalanine, and histidine. The chemical crosslinking strategy utilised throughout this thesis relies on the differential reactivity of these residues with tyrosyl radicals, and is described in detail in Section 1.2.3.2.

### 1.2.3.2 Tyrosine-Tyrosine Crosslinking

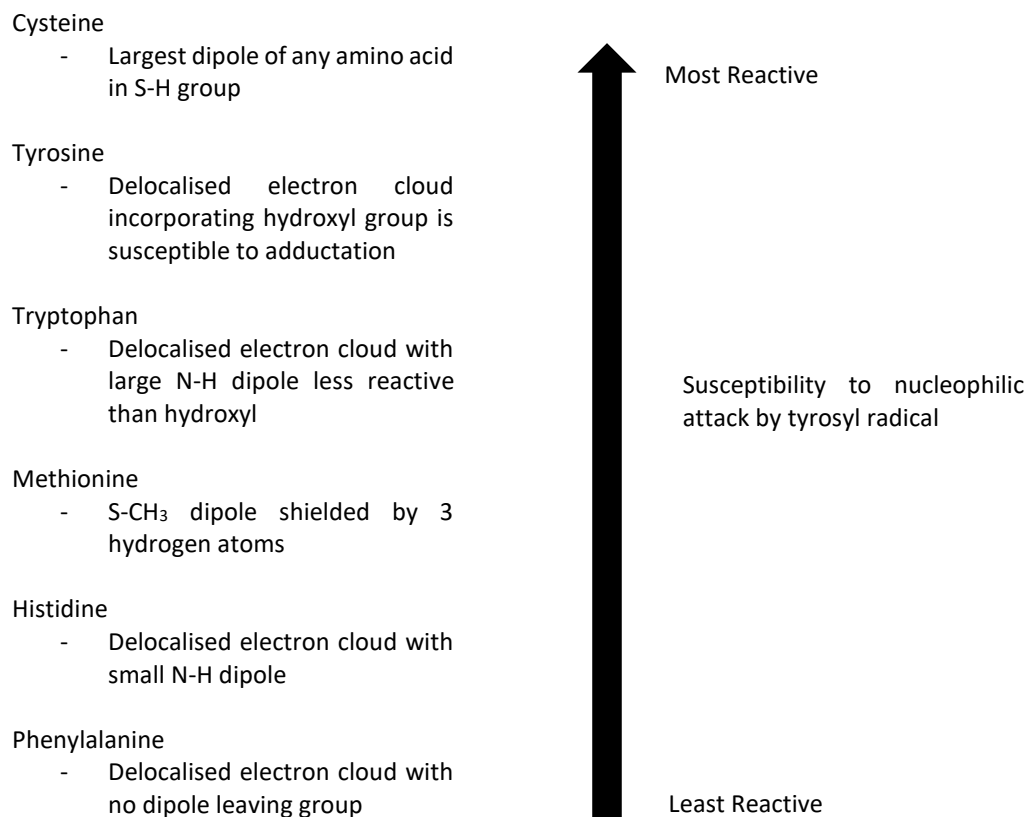
A crosslinking reaction with a high degree of specificity is photoactivated tris-bipyridylruthenium (II) ( $Ru(II)bpy_3$ )-mediated tyrosine crosslinking, whereby the phenolic rings of two spatially close tyrosines become covalently linked to form a dityrosine adduct [135].  $Ru(II)bpy_3$  is a highly photoactive molecule with a  $\lambda_{max}$  of 452nm . Upon irradiation the absorbed energy raises the complex to a higher energy state leading to a single electron oxidation of the molecule [136][137]. This new  $Ru(III)bpy_3$  complex is a highly oxidative radical. The reaction is performed in the presence of an excess of ammonium persulfate which is radicalised upon dissolving in water [138]. This combination of water-associated radicalisation of the persulphate and the photoactivated radicalisation of the  $Ru(II)bpy_3^{2+}$  complex generates two reactive species;  $Ru(III)bpy_3$  ( $Ru_{(III)}$ ) and a  $SO_4^{\cdot-}$  sulphate radical.

$Ru_{(III)}$  is a strong oxidant and is therefore able to scavenge an electron from the delocalised electron cloud of an aromatic ring to return to its stable  $Ru(II)bpy_3$  state [135]. The propensity for an aromatic amino acid's oxidation is determined by the dipoles its electron cloud creates. A ring with no attached groups contains no dipoles and therefore is resistant to oxidation,

whilst a ring with an attached group will likely have a dipole and be more susceptible. This creates an order of amino acid reactivity with Ru(III):



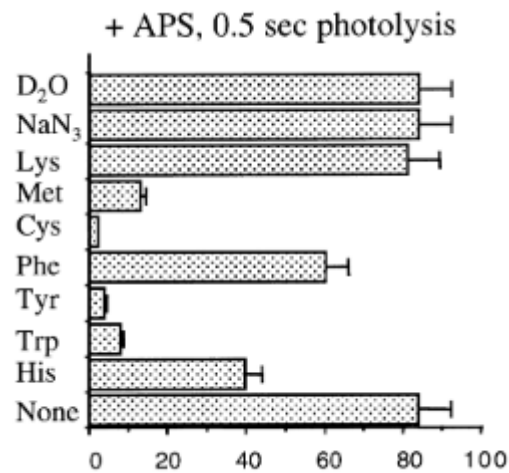
This order of reactivity means that tyrosine is significantly more favourable for oxidation by Ru(III) than any other residue, leading to the preferential generation of tyrosyl radicals. These tyrosyl radicals (and also theoretically histidine and tryptophan) are then able to initiate a nucleophilic attack, reacting with residues based on their nucleophilicity. This yields a second order of reactivity as a function of the propensity of a residue to act as an electron donor in the formation of a new covalent bond with the aromatic radical:



The  $\text{SO}_4^{\cdot-}$  sulphate radical acts as a terminal proton scavenger to complete the nucleophilic substitution of the tyrosyl radical and stabilise the resultant dityrosine adduct. This reaction mechanism yields a predictable order of crosslink species formation efficiency; tyrosine-tyrosine crosslinks or tyrosine-cysteine crosslinks are equally the most likely to undergo nucleophilic substitution by a radical as evidenced by Figure 1.9. Beyond this an exact order is less predictable but a rough order of the theoretical crosslink species is likely to be cysteine-tryptophan/histidine/phenylalanine, tryptophan-tryptophan/histidine/phenylalanine, methionine-tyrosine/tryptophan/histidine/phenylalanine, histidine-histidine/phenylalanine, and finally phenylalanine/phenylalanine. The actual formation of any crosslink species other than tyrosine-tyrosine has not been observed quantitatively, but the ability of all these residues to reduce tyrosine-tyrosine crosslinking efficiency has been demonstrated (Figure 1.9) [139]. Figure 1.9 shows that the addition of certain monomeric amino acids into the

reaction solution will quench protein-protein crosslinking to a certain extent as they compete for crosslinking sites on the surface of the proteins.

Figure 1.9 Reproduced from Fancy & Kodadek, 2000: Effect of additives on the  $\text{Ru(II)(bpy)}_3^{2+}$ /light-mediated crosslinking of the Gal4 activation domain and Gal80 protein in the presence of APS. The numbers on the horizontal axes represent percentage yield of crosslinked products.



As a strategy for forming chemically crosslinked hydrogels this mechanism has been successful. Dityrosine crosslinking has been demonstrated to yield protein lattices containing folded protein using a range of protein constructs and species. These include a polyprotein construct consisting of the ECM-native Fibronectin II domain and an integrin-binding RGD domain, an  $(I27)_5$  polyprotein constructed from the giant muscular protein titin, and GB1-Resilin [94], [97], [99]. These studies have demonstrated that the crosslinking reaction is sufficiently gentle to prevent total ablation of protein structure. However as discussed above off-target crosslinking can occur between residues other than tyrosine. This necessitates that any accurate study of crosslinking density be performed using a system with demonstrably no side-reaction crosslink formation, and is the subject of discussion and demonstration in Section 4, but in short relies on the rational selection and modification of a protein containing no surface exposed highly reactive side chains other than tyrosine hydroxyl groups.



### 1.2.3.3 Previous Hydrogel Crosslinking Strategies

The disparity in reactivity between certain side chain groups can greatly reduce the random crosslinks formed, and previous FGP hydrogels have been crosslinked with agents such as glutaraldehyde, which crosslinks lysine residues via its two aldehyde groups, and photo-initiated disulphide bridging between cysteine residues [140]–[142]. These two strategies have yielded certain levels of success but neither is perfect; glutaraldehyde has been shown to react with other less reactive amide groups if it is not spatially favourable to link two lysine residues [143], and lysine is too common to be used to regulate crosslink sites. Cysteine residues are often associated with protein function either by conferring fold strength via disulphide bridges, or as part of a metal ion-binding motif [144]. This makes their use as crosslinking points potentially detrimental to the function and mechanical strength of the protein, depriving the hydrogel of dynamic function. In addition, due to their association with specific functions cysteines are a rare amino acids species (observed occurrence of 2.28%), and are generally found spatially close to one another in fully folded proteins forming disulphide bonds [145]. This makes native cysteine crosslinking a sub-optimal crosslinking method. However the rational insertion of non-native cysteines into a protein has the potential to be a useful tool for increasing crosslinking density and strength in more advanced, designed hydrogel lattice constructs. However their insertion into proteins is often associated with association, aggregation and folding issues making their rational insertion difficult to achieve [146]. A novel cysteine crosslinking strategy has been devised capable of yielding a folded keratin hydrogel linked by allyl thioether bonds demonstrating that such strategies can be useful [141]. This construct was also capable of supporting stem cell growth, indicating that true protein hydrogels could function as scaffolds for tissue engineering. Many other crosslinking strategies have been tested with various levels of success, utilising various

residue species, with various degrees of specificity, such as thermally triggered transglutaminase-mediated lysine-glutamine crosslinking [147]. But as discussed in Section 1.2.3.1 achieving high specificity is extremely difficult, and without a method to allow exactly tuneable crosslinking efficiency rational design of hydrogel mechanics is not possible. No mechanism has yet been described for the targeted crosslinking of the side chains of tryptophan, histidine, methionine, or phenylalanine though all have been demonstrated as modifiable and capable of forming random crosslinks in the presence of certain linkage molecules.

#### 1.2.4 Mechanisms of Hydrogel Formation

The formation of a hydrogel is synonymous with the formation of a network. Hydrogels begin as discrete constituent monomer units diffusing freely through an aqueous medium according to the laws of Brownian motion. All monomers diffuse at the same rate and are homogenous throughout the solution volume. The transition from this solution state to a gel state occurs as a result of monomers becoming bound to each other via a sufficiently permanent mechanism whereby a continuous network comes into existence spanning the entire volume of the solution. This network is subsequently able to persist on a timescale exceeding the viscous movement of fluid within it, meaning that the elastic properties of the network dominate the viscous properties of the solution. This classifies the material as solid.

##### 1.2.4.1 The Requirements of Network Formation

The mechanism of network formation and persistence is reliant on the rate of bond formation ( $kB_f$ ) exceeding the rate of bond breakage ( $kB_b$ ). If the rate of bond formation is lower or equal to the rate of bond breakage then a network cannot form as at no instant will sufficient

monomers be bound together to span the solution volume. The degree to which  $k_{B_f}$  exceeds  $k_{B_b}$  in essence determines the rate of network formation. This does not necessarily lead to certain mechanical characteristics or network topologies. The  $k_{B_f}:k_{B_b}$  ratio is determined firstly by the species of bond being formed. A low ratio example is van der Waals forces, which are transient ionic bonds between short-lived dipoles [148]. The probability of a dipole existing between any two covalently-bonded atoms at a given moment is a function of the difference in their electronegativities; the greater the difference, the longer-lived the dipole. Commensurately the longer-lived a dipole is the greater the  $k_{B_f}$  and the lower the  $k_{B_b}$ , meaning that a network is more likely to form over time. A high ratio example is the formation of chemical covalent bonds. The rate of covalent bond formation can vary significantly, but the rate of covalent bond breakage under normal conditions is always extremely low, and a covalent bond can almost be considered permanent on physiological timescales [149]. All other bond species exist somewhere in between these two extremes with a general rank order of  $k_{B_f}:k_{B_b}$  magnitude of: ionic>metallic>>frictional [150]. In all cases the bond species is to a greater or lesser degree transient, but if  $k_{B_f}$  remains greater than  $k_{B_b}$  then the network will exist constantly, but will be forever changing on the microscale. This leads to micromechanical variations over time, but on the bulk scale the mechanics of the gel will remain constant in the absence of environmental insult. In systems which undergo dynamic network polymerisation in response to an external cue removal of this cue halts bond formation, meaning  $k_{B_f}$  will drop to zero. In systems such as this over time the network will degrade no matter the nature of the bonds, as a  $k_{B_b}$  no matter how small is greater than a  $k_{B_f}$  of zero, meaning a net loss of bonds per unit time [151]. For gels formed by physical interactions such as molecular entanglement (friction) or ionic interactions the  $k_{B_f}$  can be reduced by alterations in pH, temperature, or co-solute concentration all of which can be very

rapidly induced. This coupled with the high  $kB_b$  of such bond species makes physical gels highly sensitive to environmental stimuli, and generally short lived as environmental homeostasis is challenging to achieve [152]. Hydrogels formed by covalent bonds are longer lived and less sensitive to environment as extreme conditions are required to increase the  $kB_b$  of covalent bonds.

#### 1.2.4.2 Models of Network Formation

The formation rate and resultant topology of a network is dictated by two things; the rate of monomer diffusion ( $kD$ ), which is a function of the volume fraction that the network species occupies and the solvent viscosity, and the probability of bond formation ( $p$ ) when two particles come within the reaction distance ( $rd$ , the distance over which electron dissociation can occur and a covalent bond form). In this section all bonds will be assumed to be covalent and so only the rate of bond formation will be considered as the rate of covalent bond breakage is insignificant under the considered timescales. The ratio between the rate of diffusion and the probability of bond formation causes two distinct regimes of network formation. If  $kD \gg p$  then it can be assumed that the rate limiting step in network formation is the formation of bonds; network formation is reaction rate limited. If  $p \gg kD$  then network formation is limited by the chance of a particle diffusing close enough to another to form a bond; this is diffusion rate limited. For the sake of simplicity network formation is conceptually simplified to aggregation; the gain of density from a dilute solution. This gives us the two key models of network formation: reaction limited aggregation, and diffusion limited aggregation. These different models and their applicability to hydrogel formation are discussed [153][154].

#### 1.2.4.2.1 Diffusion Limited Aggregation (DLA)

All particles fully dissolved in aqueous solution diffuse freely, continuously, and eternally according to the laws of Brownian motion. The directionality of particle movement in the absence of fluid current is random, and the rate of movement is determined by three properties; the size of the particle, the viscosity of the solvent, and the total energy of the system. Larger particles experience more drag as they collide with more molecules of solvent and therefore move more slowly, more viscous solvents exert more resistance to solute movement, and the amount of energy in the system affects the convection force of particles and the viscosity of the solvent [155]. In a DLA regime  $p$  is given the value of 1. This means that if two monomers come within the  $rd$  there is a 100% chance that they will form a bond. This means that to form bonds all that has to happen is for monomers to diffuse close enough each other, and this happens more often the faster they move through the solution; the faster they diffuse the more likely they are to get close enough to each other to react. This rule of aggregation can be simulated by visualising the movement of particles through a finite medium volume and observing the pattern in which they are captured and added to the aggregate [156].

Imagine a box with a single particle in its centre. Into this box a second particle enters from a random direction. The second particles movement within the box is random and if tracked would resemble a tangled ball of string. At some point however the second particle will come within the  $rd$  of the first particle. A bond is instantly formed and a static dimer exists at the centre of the box. This is the nucleus of the aggregate. An infinite number of subsequent particles enter the box, diffuse freely for various lengths of time before coming close enough

to the growing aggregate to form a bond and become part of it. This causes a phenomenon called network branching. As the nucleus grows from its original dimer each new particle will react with the first part of the aggregate it comes within  $rd$  of. The result of this is that incoming particles are captured by the outermost units of the aggregate furthest from the nucleus and never reach the centre of the aggregate. This has been shown via simulation to give rise to sparse branched chains which collect incoming particles at their outermost edges, and shield the core like the canopy of a tree catching [157], [158].

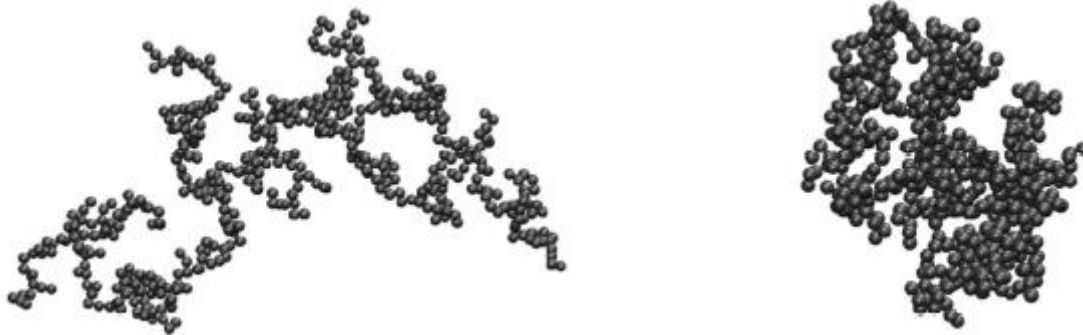
#### 1.2.4.2.2 Reaction Limited Aggregation (RLA)

In an RLA system the rate of diffusion is constant and can be fast or slow. Instead the  $p$  of bond formation when two particles come within the  $rd$  is  $<1$ , meaning that there is a chance that no bond will be formed and the particles will drift apart again. This is a result of the rate of diffusion exceeding the timescale of bond formation ( $kB_f$ ). This means that particles entering an imaginary box containing an aggregate nucleus will not necessarily form a bond with the first aggregate unit it comes within the  $rd$  of. Instead particles have a greater chance of diffusing past the outermost edges of the aggregate and continuing closer to the nucleus before capture. This favours the addition of monomers closer to the core, and has been computationally demonstrated to yield dense non-branching aggregates spanning a smaller volume than a DLA system [157][159]. This effect can vary in magnitude as a function of the ratio between  $kD$  and  $p$ ; the lower the  $p$  value the greater the chance of not forming a bond with the first aggregate unit, and the greater the chance of diffusing closer to the nucleus. The greater the  $kD$  the less time a particle will spend within the  $rd$  of another thereby preventing bond formation, and favouring diffusion closer to the nucleus. Therefore the

greater the degree by which  $kD$  exceeds  $\rho$  the denser the aggregate will be. Figure 1.10 demonstrates the topological differences between aggregation regimes dominated by either DLA or RLA:



Figure 1.10: Simulation models of DLA and RLA network formation regimes [157].



Left: 2-dimensional projection of a 256 particle aggregate formed under a DLA regime ( $\rho \sim 1$ ).

Right: 2-dimensional projection of a 256 particle aggregate formed under a RLA regime ( $\rho \gg 1$ ).



#### 1.2.4.2.2 The Real-World DLCA and RLCA Regimes of Hydrogel Network Assembly

In a real-world hydrogel network formation system neither a DLA nor an RLA model can adequately capture the true complexity of the process because both occur simultaneously with each dominating at different times. This is due to two key facts; firstly there is not a single aggregate nucleus, instead multiple dimers form independently in the reaction volume and subsequently scavenge further diffusing monomers leading to multiple clusters across a range of length scales. Secondly the number of monomers is finite meaning that over time monomer concentration decreases. This leads to a changing ratio between  $\rho$  and  $kD$  over time, leading to a crossover between the two regimes at some point during network formation. In essence a real-world system is a hybrid of the two regimes with both dominating

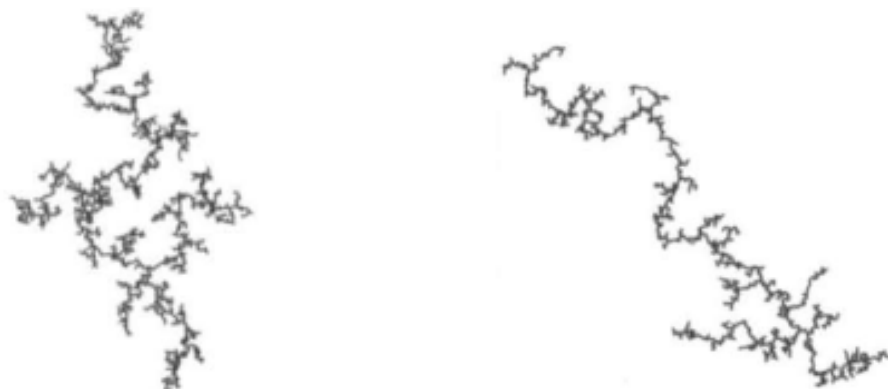
at different stages during the assembly. Furthermore a hydrogel system is not a constant addition of monomers to a single nucleus, but rather an addition of monomers to multiple clusters in combination with cluster aggregation. This is called diffusion/reaction limited cluster aggregation (DLCA/RLCA) [160].

Monomers will have a constant  $kD$  throughout the aggregation process, but the multiple discrete clusters they form during the early stages of assembly will develop a range of unique  $kD$ 's. This means that multiple sized clusters will be diffusing at different speeds. The  $p$  of monomer addition is constant but the probability of a cluster binding to another cluster is a function of the number of bond sites the cluster has giving each cluster a unique probability of cluster-cluster bond formation ( $pC$ ). The early stages of network formation will be dominated by monomer addition to multiple clusters. As the monomer concentration falls the dominant process will shift to cluster aggregation. This leads to a change in overall rate of network formation. Once again cluster aggregation can be either diffusion or reaction rate limited, but neither the  $kD$  or the  $pC$  are constant and are instead constantly changing as assembly progresses. As the average cluster size increases the average  $kD$  will decrease, and as the average number of available bonds per cluster increases the  $pC$  will also increase. The interplay between all these factors leads to different network topologies; an RLCA-dominated system will exhibit multiple large dense regions connected by short branched chains, and a global tendency towards higher density. A DLCA-dominated system will exhibit fewer and smaller dense regions and longer more highly branched chains and a global tendency towards greater sparsity [160], [161][162]. This is illustrated in Figure 1.11:



---

Figure 1.11: Simulation models of network formation by RLCA and DLCA dominated regimes [274].



Left: 1000 particle simulation of extended cluster formation by RLCA.

Right: 1000 particle simulation of extended cluster formation by DLCA.

---

Hydrogel networks formed by covalent bond formation (chemical gels) will exhibit network topologies determined by the  $kD$  of the monomer and the  $p$  of monomer-monomer bond formation. The closer  $p$  is to 1 then the more the system will tend towards a DLCA regime, and the further from 1 the more it will tend to RLCA. The subsequent cluster  $kD$ 's and  $pC$ 's will likely be highly variable and affected by the density of crosslink sites in the monomer and the degree of their continued availability once incorporated into a growing cluster. This means that a system dominated by DLA initially may not be dominated DLCA later. Understanding the relationship between crosslink density, crosslink efficiency, and ultimate macro mechanical gel properties is the central aim of this thesis. Chemical protein hydrogels exhibit a network topology which can broadly be described as interconnected clusters [102]. Changes in average cluster size would be expected from proteins with different crosslink densities as clusters with few accessible crosslink sites will tend towards RLCA and larger clusters, whilst

those which retain many accessible crosslink sites during cluster growth will tend towards DLCA and smaller clusters. The interplay between crosslink density, assembly regime, network topology, and final gel properties is highly complex and variable by a thousand methods, and represents the ultimate future question of rational protein hydrogel design.

### **1.3 Immunoglobulin 27**

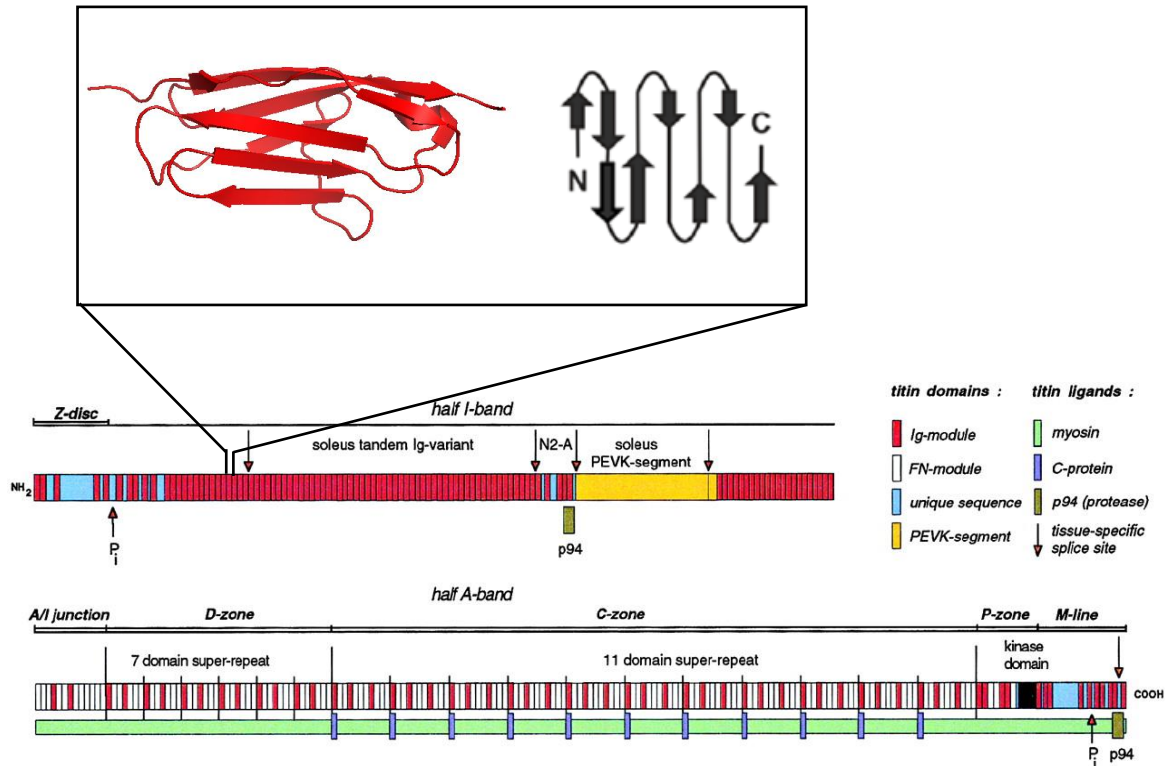
#### **1.3.1 Titin**

Titin is the largest single chain protein in the human proteome, with a mass of between 3 and 4.2 megadaltons depending on the splice variant, and a length of almost 1 $\mu$ m and width of 4nm [163]. Its primary function is as an elastomeric spring in the sarcomeres of striated muscle, enabling contracted or extended sarcomeres to return to their passive length post-contraction in an elastic fashion. Titin acts as a kinetic energy store, storing energy generated during the ATP-myosin driven contraction process. Upon the dissociation of the myosin head/actin crossbridge complex the elastic potential force stored in the titin causes titin to spring back to its resting length, either re-shortening or re-lengthening the sarcomere depending on the direction of Z-band movement relative to the relaxed state [164]. Titin is a polyprotein composed of approximately 300 distinct Ig-like domains, and spans half the entire sarcomere (there are two titin molecules per sarcomere) between the Z-disk and the M-line. Titin is an approximately 27000 residue fibril comprising two distinct domains; the I-band domain which contains 297 copies of repeating Immunoglobulin-like domains (Ig's) and fibronectin-like domains (FN<sub>III</sub>), which are folded into globular  $\beta$ -sheet structures. This I-band region acts as a passive elastic spring during muscle relaxation, extending then recoiling via its elastomeric Ig repeats. The second region of Titin is the A-band which comprises highly ordered and stiff repeats of Ig and FN<sub>III</sub> domains to provide specific binding sites for Myosin

and C protein [165]. At high forces *in vitro* the individual domains of titin lengthen before unfolding sequentially [166]–[169]. The dynamic refolding of these domains *in vivo* may have an elastic recovery effect, but there is some debate as to whether sarcomeres can become sufficiently extended to unfold these domains [170], [171]. In either case the ability of Ig domains to undergo a significant degree of lengthening without unfolding plays a crucial elastic role in titin's ability to return to its passive length post-contraction.

Each Ig domain of titin is sequentially unique and all are believed to be mechanically unique, yet all are nearly structurally identical. One specific domain of cardiac titin classified as Immunoglobulin Domain 27 (I27) has become the most highly mechanically characterised protein in history, and has been utilised for the purposes of this thesis as the building block from which to build folded globular protein hydrogels due to its high mechanical stability, known structure, and its natural sequence advantages for use of the tyrosine crosslinking strategy described in Section 1.2.3.3.

Figure 1.12: Schematic representation of the composition of Titin (bottom [165]) illustrating the size and repetitive structure of the 3.7 megadalton polypeptide. Enlarged (top): 1.8Å X-ray crystallographic structure [181] of a single Immunoglobulin domain 27 unit.



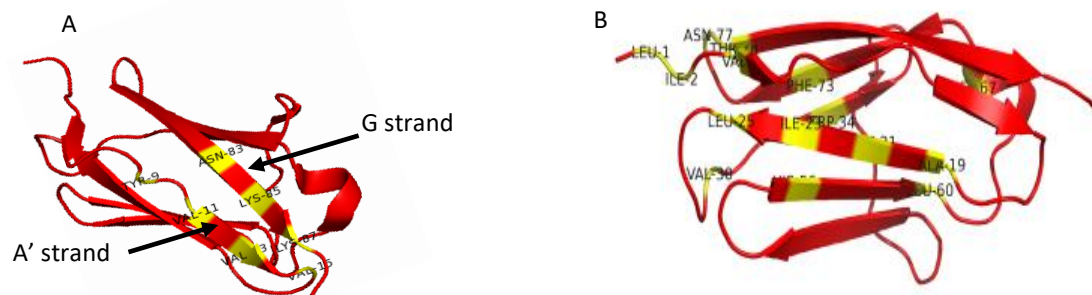
### 1.3.2 Immunoglobulin Domain 27

Immunoglobulin domain 27 (I27) is a non-repeating elastomeric domain isolated from the I-band of the giant cardiac muscle protein Titin [172], [173]. Due to its role as an elastomeric unit *in vivo* and the early solving of its crystal structure, I27 became a paradigm for early mechanical unfolding experiments using AFM [174], [175]. Like almost all mechanically strong proteins the secondary structure of I27 consists almost exclusively of  $\beta$ -sheets, the most mechanically stable secondary protein structure species [176], [177]. I27 consists of 7 parallel  $\beta$ -sheet strands connected by random coils and  $\beta$ -turns, with a zipper region of hydrogen bonds acting as an H-bond clamp between the A' and G strands (Figure 1.13A yellow residues), with residues Y9, V11, V13, and V15 on the A' strand being H-bonded to residues N83, K85, and K87 on the G strand. The shearing of this zipper region is believed to be the rate limiting

step in the complete mechanical unfolding of I27 modules in the absence of a disulphide staple [178]. T78 is also believed to be H-bonded to L1 to confer additional mechanical strength [179]–[181]. This cluster of residues is an obvious candidate for mechanically-weakening mutations as proline substitutions of all 4 of the A' strand residues have been shown to reduce the peak unfolding force of the protein relative to wild type at relatively high speeds, though this may not be true at pulling speeds lower than  $0.001 \text{ nm ms}^{-1}$  [180]. Additionally, by pulling *in silico* at a force just below peak unfolding force, several residues were identified which shifted orientation in space to the greatest degree, thereby indicating that they are involved in the force propagation network of the protein, a key component of mechanical strength [181]. These include I2, V4, N77, I23, L25, V30, W34, H56, F73, and T78 on the N-terminal side of the protein, and A19, F21, L60, M67, and L84 on the C-terminal side, as well as predicting a further H-bond between T78 and L1 (Figure 1.13B).

---

Figure 1.13: A; the hydrogen bond clamp/zipper region between the A' and G strands of I27. B; The computationally predicted mechanically important residues of I27.



---

Making hydrogels from folded globular proteins is of interest for one key reason; if the hydrogel lattice can be constructed from almost universally folded protein, then the gel itself can be designed to incorporate many of the mechanical, responsive, and catalytic properties

of the protein. Towards this end the first property that a prospective protein hydrogel building block must exhibit is an ability to retain its folded state post-gelation. This likely requires a significant degree of mechanical stability as the swelling forces associated with gelation are suspected to be considerable, leading to unfolding of weak proteins and rendering them little more than semi-flexible polymer chains. I27's high degree of mechanical strength and elastic behaviour at sub-rupture forces makes it a prime candidate for hydrogel formation, as does its ability to maintain its mechanical properties in polyprotein chains of various lengths [182]. The importance of maintaining a fully folded protein state once in a gel phase is of crucial importance both for maintaining protein function, and to rationally tune the viscoelastic properties of the hydrogel.

In addition to its mechanical stability, I27 is an ideal candidate for investigating the relationship between crosslinking density of the hydrogel lattice and the macroscopic properties of the gel. This is as a result of the chosen tyrosine-crosslinking strategy used throughout this study as detailed in Section 1.2.3. As discussed earlier, activated tyrosine residues can also crosslink to methionine, cysteine, tryptophan, and histidine side chains [139]. The I27 variant designed and created for use in this thesis (Appendix 2), descended from the previously reported C47S/C63S mutant [183], contains one buried residue each of cysteine, tryptophan, and methionine. In addition it contains four histidines; two are buried, one forms a hydrogen bond with a glutamic acid and is therefore likely to be highly unfavourable for crosslinking, leaving one solvent exposed non-bonded residue potentially available for crosslinking. Histidine is however the least efficient off-target crosslinking residue after phenylalanine, and likely has a rate constant of bond formation far lower than tyrosine. As a result, the primary sequence of I27 lends itself well to the avoidance of off-target crosslinking, making it accurate to state that the number of available crosslink sites is equal to the number

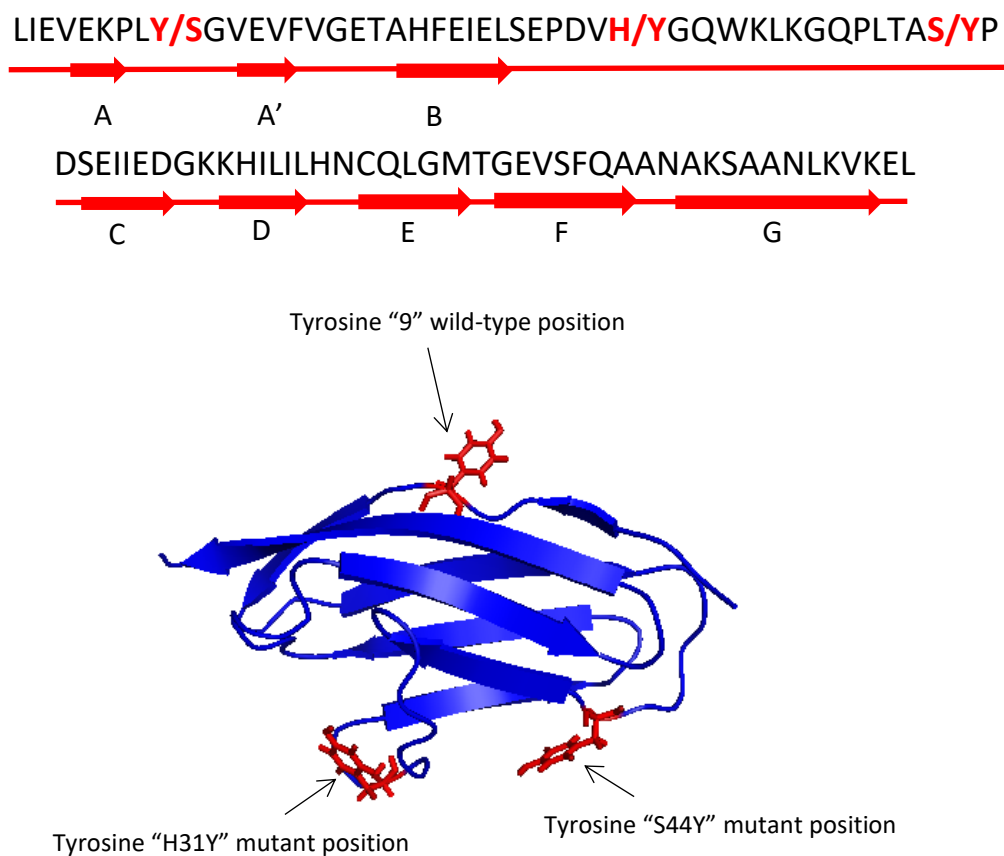
of tyrosines in the construct. This is discussed in greater detail in Section 1.2.3.2. I27 contains a single natural tyrosine residue at position 9. The substitution of this residue has no effect on the mechanical strength of the protein [184], and as such will have no effect on the properties of the hydrogel, with any alterations measured due solely to the altered density of crosslinking [185]. I27 consists of  $\beta$ -sheets joined by random coil/ $\beta$ -turn bends, which provide ideal locations for the insertion of additional tyrosine residues with minimal disruption to the structure and mechanical stability of the protein.

---

Figure 1.14:

Top; primary sequence of I27 mutant domain used in this study. Tyrosine substitutions in **bold red**.

Bottom; X-ray crystallographic structure of I27 (1TIT, [172]) with the position of each potential mutagenically inserted tyrosine residue.



The extensive characterisation of the force propagation network of I27 has yielded a thorough understanding of how to tune its mechanical stability. This understanding, together with its mechanical stability and a lack of potential off-target crosslinking residues makes I27 an ideal candidate for use as a model system to study the relationship between crosslinking density and hydrogel macromechanics.

## **1.4 Rheology**

### **1.4.1 Introduction**

The word rheology comes from the Greek *panta rhei* meaning "everything flows", and the analytical technique of rheology is the study of the deformation, or flow, of materials in response to shear stress [186]. Rheometric experiments act to apply stress to materials whereby the geometry head of the instrument is rotated across the surface of the material to induce torsional deformation. Several types of measurement can be taken from this interaction depending on the pattern of force applied to the material, all of which summated can describe the viscoelastic properties of a material. Viscoelasticity is the relationship between two competing mechanisms of resistance to deformation; the  $G'$  storage modulus, and the  $G''$  loss modulus [5]. The elastic  $G'$  storage modulus of a material or structured fluid can be defined as the degree to which a material acts like a solid and is able to absorb stress and recover without permanent alteration of the materials structure. The viscous  $G''$  loss modulus is the opposite of the storage modulus, and is a measure of how much the material behaves like a liquid with no ability to recover its structure after stress [187], [188]. The deforming force applied to hydrogels is in the form of shear stress, as the geometry moves in a circular motion across the top surface of the material and the bottom remains static. This induces torsion in the material as the top of the gel moves relative to the bottom. By



deforming the material with various degrees of shear force and in various patterns the viscoelastic properties of a hydrogel can be determined as a function of amplitude, frequency, time, and hysteresis amongst others. From this general network properties can be inferred, and an understanding of the network topology can be teased out. To perform rheological measurements on viscoelastic materials only three parameters need to be defined beforehand; the sample height, the sample area, and the torque applied to the sample [187].

#### 1.4.2 Viscosity and Structured Fluids

Viscosity is the resistance of a fluid to flow. A structured fluid is a fluid which contains a degree of structure which requires energy input to allow to flow [189]. All fluids are structured and therefore have viscosity, only an imaginary “perfect fluid” has none and exists only as a concept to describe all real fluids relative to its perfectly unstructured state. There are two types of flow; extensional and shear. Extensional flow is an elongation of the fluid volume, and is characterised by the simultaneous stretching and thinning of the fluid. Measurement of this is vital for fluid mechanics, but is complicated for the understanding of viscoelastic materials. This is why rheology utilises shear flow, though shear and extensional are equivalent and can provide the same structural and mechanical information.

Shear flow in essence is the flow of layers of fluid at different rates inside a single volume [190]. Via friction between the geometry head and the fluid stress is applied to the top layer of the fluid, causing it to flow. The shear stress is defined in Equation 1.1 as the force imparted over the area of the sample.

$$\text{Equation 1.1: } \sigma = F/A \text{ (Pa)}$$

The force moves the top layer of fluid a certain distance (X), whilst the bottom layer remains stationary. In between these two layers all other layers exist in a gradient of displacement, which is equal to distance/height (h) and is called the shear strain (Equation 1.2).

$$\text{Equation 1.2: } \gamma = X/h \text{ (\%)}$$

The gradient of displacement through a column of fluid as exists in a rheometric measurement is determined by friction versus viscosity. A known force is applied to the top layer of the fluid by the rheometers motor in torque. This is converted into Pascals for ease of analysis. The rate at which force is applied is called the shear rate and is defined in Equation 1.3 where V= shear rate, r= radius of the geometry, and  $\omega$ = angular velocity.

$$\text{Equation 1.3: } V = r\omega$$

This rate is known only for the top layer, as each subsequent layer experiences diminishing stress as a result of the imperfect transfer of momentum [191]. Momentum is transferred by the collision of molecules (friction), and every collision results in loss of energy dissipated mostly as heat, causing a loss of momentum [192]. This energy loss is the key parameter inferred during the characterisation of a viscoelastic material. This gradient of displacement and loss of momentum through the fluid is called the shear strain rate, and is defined in Equation 1.4 as the rate of change (Y) of change in strain ( $\Delta\gamma$ )/time ( $\Delta t$ ).

$$\text{Equation 1.4: } \dot{\gamma} = \Delta\gamma/\Delta t$$

This rate is intrinsically linked to viscosity as it is a direct measure of the time between the application of stress and the beginning of strain, i.e. how resistant to flow the fluid is. Commensurately viscosity becomes a measure of the internal friction of a fluid. This relationship is defined in Equation 1.5 where  $\eta$ = shear viscosity.

$$\text{Equation 1.5: } \eta = \sigma / \dot{\gamma} \text{ (Pa/S)}$$

The viscosity of a fluid is not constant, and varies with stress and frequency [193]. Fluids which exhibit a linear relationship between shear stress and strain are defined as classic Newtonian fluids; double the stress and the fluid flows twice as fast. These fluids have an almost constant viscosity as an increase in stress causes an increase in strain and rate meaning that once divided by each other according to equations 1.4 and 1.5 the final viscosity remains the same [194]. Fluids which do not conform to this linear relationship and instead exhibit a non-linear dependence between stress and strain shear rate have variable viscosity, which varies as a function of shear rate/stress. These fluids can become either more or less viscous in response to increased stress or shear rate, and are called non-Newtonian fluids [195].

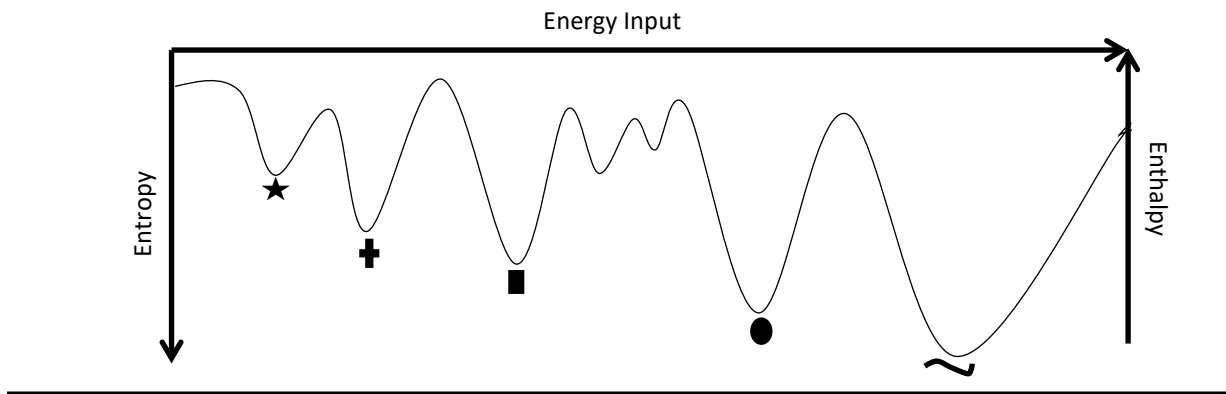
Whilst viscosity is a parameter associated with fluids, it is actually a measure of fluid-like behaviour. All solid materials have fluid like behaviour but it occurs on such long timescales or under such massive stress that they appear perfectly solid. Viscosity is therefore a crucial parameter in the description of all materials.

### 1.4.3 Elasticity and Perfectly Elastic Materials

Elastic behaviour is a thermodynamic equilibrium between entropy, enthalpy, and the energy required to increase one or the other. Any structure, be it a molecule of water or a polymer network, exists at the bottom of an entropic minima, where the enthalpy of the system is sufficient that an energy barrier exists between the current state and one of higher entropy [196]. This means that a structure can be maintained in equilibrium providing no more energy is put into the system. In order to reorder or destroy a structure and increase its entropy sufficient energy must be put into the system to lift it out of its energy well, over the energy barrier, and into the next (usually entropically higher) adjacent energy well. If enough energy

is put into a system it will shift into lower and lower entropic states, until it totally disintegrates and becomes entirely disordered as the equilibrium is irreversibly destroyed. In absolute terms this would be the conversion of all matter into energy according to Einstein's Theory of Special relativity [197]. This relationship between energy, entropy, and enthalpy is schematically represented below In Figure 1.15:

Figure 1.15: Schematic representation of the relationship between system energy input, enthalpy, and entropy.



In order to overcome an energy barrier and shift a structure into a lower entropic state X energy must be input over Y time. Simultaneously the natural equilibrium of the system will push back as the system emits energy. If insufficient energy or too much time is input then the total level of energy in the system will never exceed the energy barrier, and the structure will not be permanently altered [198]. Once excessive energy input ends the structure will re-emit energy (usually as heat or light) as it relaxes back to its minimal entropic state within its energy well as an equilibrium is re-established [199]. This energy input followed by output resulting in no net change to the structure when related to kinetic energy is elasticity. In relation to elastic materials this means that the kinetic energy of stress is sufficient to deform

but not permanently alter the material structure as the kinetic energy is stored as tension within it. Once the stress is removed the material is able to spring back giving out kinetic energy exactly equal to that which was put in. This exact equality between stress input and kinetic relaxation output is the hallmark of a perfectly elastic material. In this way a perfectly elastic material conforms to Hooke's Law: an applied stress is proportional to strain up to the elastic limit, and will return to its initial state once stress is removed. If the elastic limit is exceeded the structure will be permanently distorted [200]. The elastic modulus of a material ( $G'$ ) is a measure of the stiffness of a material, i.e. its degree of resistance to permanent deformation. A perfectly elastic material has no lag time in response; stress causes immediate and proportional strain, and removal of stress is followed by immediate rebounding of the spring [201].

#### 1.4.4 Viscoelastic Materials and Models

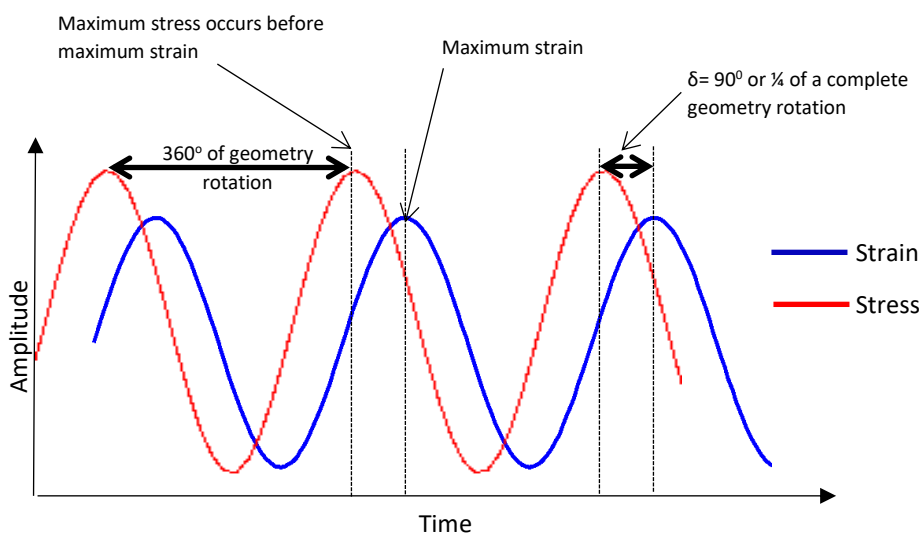
A perfectly elastic material is able to recover its original state after limitless deformation. A perfect fluid has no elastic behaviour and therefore has no ability to resist flow or recover its structure after stress. There are no perfectly elastic materials in the world, and no perfect fluids. This means that every material can be described as viscoelastic. Some are so dominated by fluid or solid-like behaviour that the minor modulus has no real world effect, but it is nonetheless always present. Viscoelastic behaviour is a complex modulus to describe both the viscous and elastic properties of a material, and the degree to which a material's bulk properties are governed by each. A viscoelastic material exhibits a degree of elastic behaviour; they can be deformed and recover, but always imperfectly as a result of their

viscous behaviour (energy dissipation). How far a material sits between these two extremes is described by its phase angle ( $\delta$ ) (Figure 1.16).

Phase angle is a measure of how far from a perfect solid and a perfect fluid a viscoelastic material is [202]. A perfectly elastic material has no lag time between stress and the induction of proportional strain, meaning that maximum stress and strain occur simultaneously. Stress and strain are described therefore as being perfectly in phase; this gives a phase angle of  $0^\circ$ . A perfect fluid with no internal friction has a lag time between maximum stress application and maximum strain being reached. The time difference between maximum stress and strain corresponds to exactly  $\frac{1}{4}$  of a revolution of the circular geometry, meaning that the stress and strain is  $90^\circ$  out of phase. Therefore a perfect fluid has a phase angle of  $90^\circ$ . A viscoelastic material has a phase angle  $>0$  and  $<90^\circ$ , with  $\delta < 45$  indicating a solid material, and a  $\delta > 45$  indicating a fluid [186][5]. A phase angle increasing or declining above or below  $45^\circ$  is indicative of network formation or breakdown.

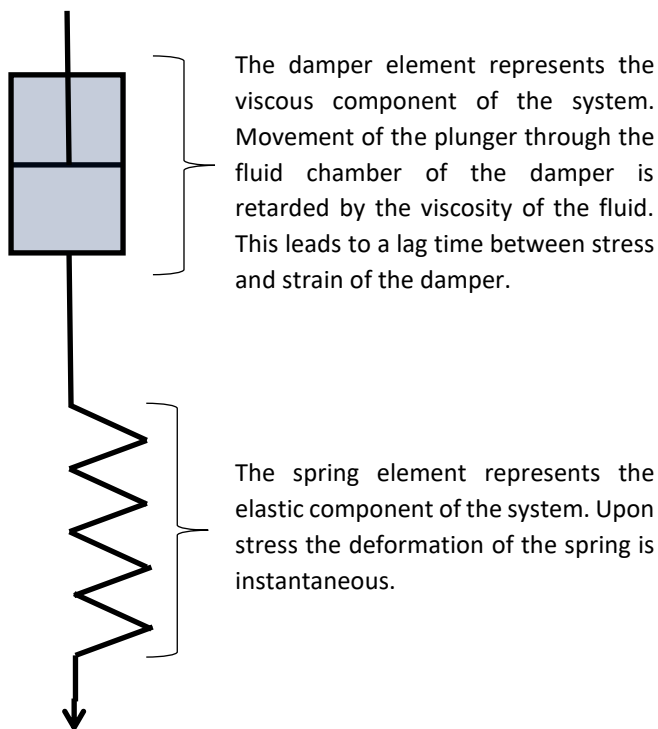
---

Figure 1.16: Stress and strain amplitude versus time for an ideal fluid. Maximum stress occurs before maximum strain, exactly  $90^\circ$  out of phase.

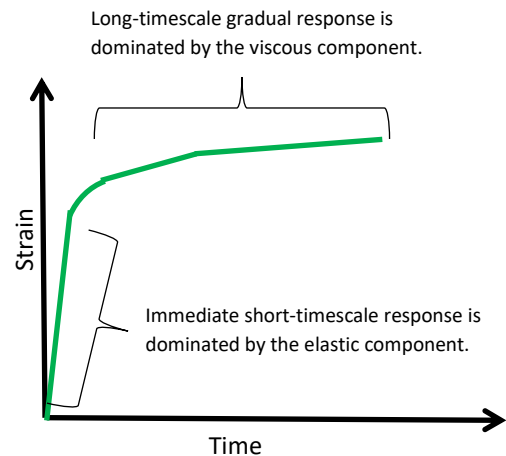


Models of viscoelasticity require the separation of the viscous component and the elastic component into separate halves of a single system. The two components are independent of each other, but are both intrinsic to the final bulk properties of the material. The most common way to visualise these two separate components and build conceptual models of the internal mechanics of a viscoelastic material is by the use of imaginary springs (elastic elements) and dampers (viscous components) as separate elements in a complete system. The simplest model of viscoelasticity is the Maxwell model, in which a single damper and a single spring are joined in series and subjected to extensional stress [203].

#### Maxwell Model:



#### Maxwell Model Deformation Profile Strain versus Time:



The Maxwell model is the simplest model of viscoelastic behaviour, in which stress acts unidirectionally upon both the viscous and elastic components simultaneously. The strain

response is two staged, dominated on short timescales by the elastic component (G) and on longer timescales by the viscous component ( $\eta$ ). Such a simple model however does not accurately model most viscoelastic materials.

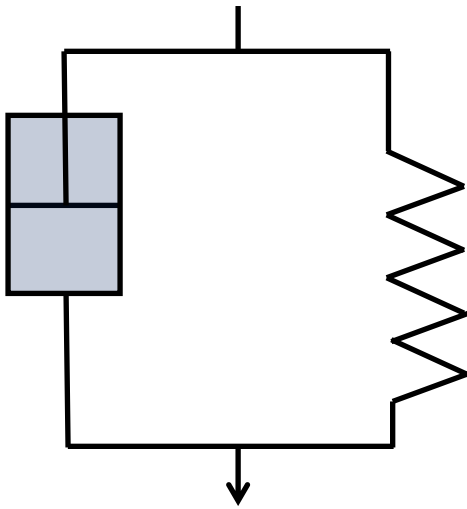
The strain evolution of the Maxwell model is expressed in Equation 1.6:

$$\gamma = \sigma \left( \frac{1}{G} + \frac{\sigma t}{\eta} \right)$$

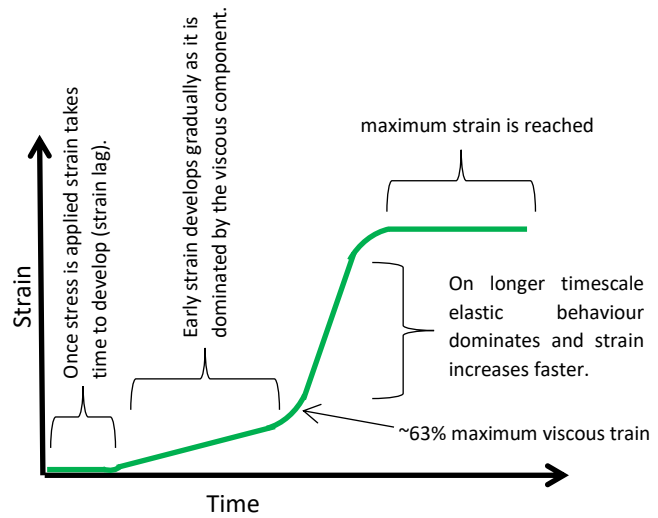
The second simplest model for viscoelasticity is the Kelvin-Voigt model in which the damper and spring are joined in parallel rather than in series. Once again stress acts upon both elements simultaneously, but in this model the spring is unable to undergo instantaneous strain because its response is retarded by the damper. As a result the short to medium timescale response is dominated by the viscous component, before a long-timescale elastic response takes over as the damper reaches close to maximum strain [204]. Interestingly the transition from a viscous-dominated response to an elastic-dominated response occurs consistently at ~63% of the maximum strain of the viscous component [190].



Kelvin-Voigt Model:



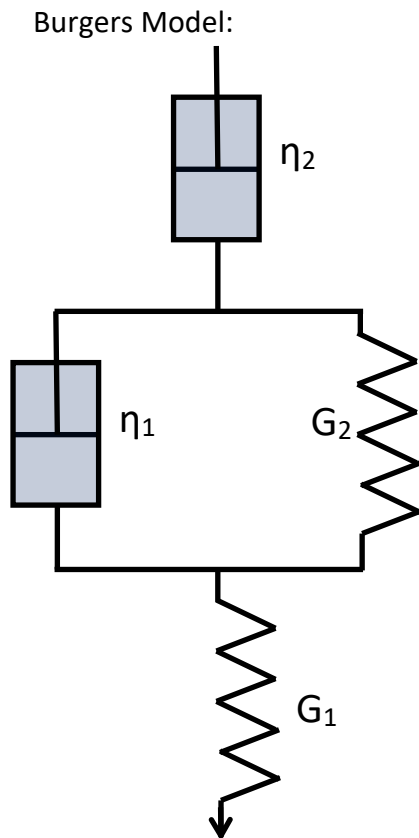
Kelvin-Voigt Deformation Profile Strain versus Time:



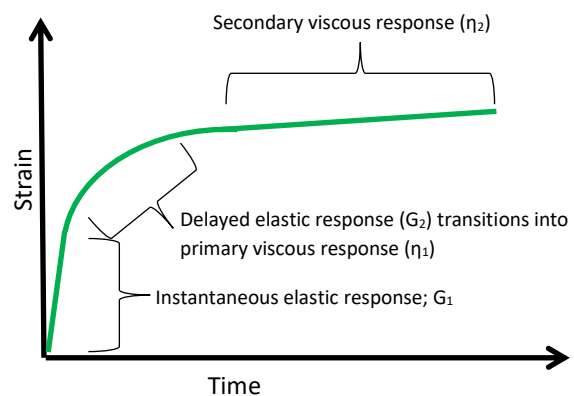
The strain evolution of the Kelvin-Voigt model is expressed in Equation 1.7:

$$\gamma = \frac{\sigma}{G} [1 - e^{-t/\eta}]$$

Real systems are more complex than either the Maxwell or Kelvin-Voigt models can describe alone. Instead a combination model featuring multiple dampers and springs in both parallel and series is generally used to describe the behaviour of real-world viscoelastic materials. This is called the Burgers Model, and yields a complex strain evolution profile. Upon the application of stress an instantaneous elastic response occurs, followed by a dampened elastic response, before a final fully viscous response [205].



Burgers deformation profile strain versus time:



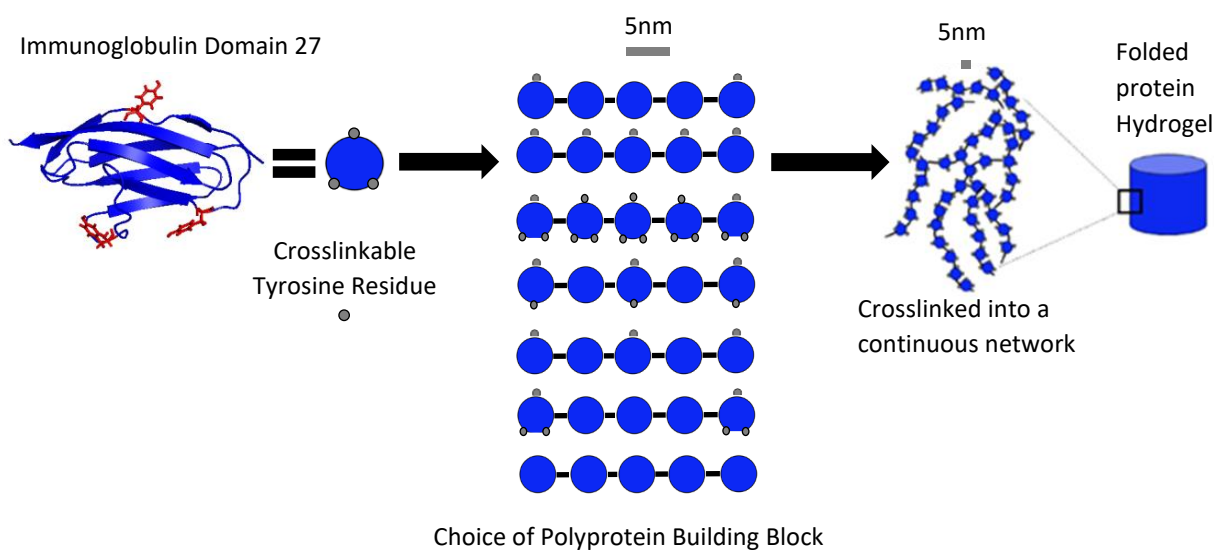
By the addition of more springs and dampers to the system it is possible to develop models to describe the observed behaviour of all viscoelastic materials. In this way strain evolution profiles give a key insight into the relative contribution of the elastic and viscous components to a materials behaviour.

## 1.5 Aims of This Thesis

Hydrogels represent the future of biomaterials. Their design as 3D cell cultures for study of *in vivo* cellular behaviour, and as scaffolds for synthetic tissues and organs promises to revolutionise drug development, disease research, and *in vitro* tissue synthesis. As smart wound dressings capable of accelerating healing hydrogels are already making a significant contribution to improving the clinical outcome of large wounds and burns [206]. As novel

materials capable of filtering solvents, acting as sensor platforms, or catalytic sieves hydrogels may yet come to revolutionise various aspects of many industries. As biomimetic 3-D cell cultures for in vitro studies and synthetic tissues for implantation hydrogels have the potential to drastically accelerate pharmaceutical development and negate the need for organ donation, skin grafts, and non-organic implants. In all examples the development of folded protein hydrogels represents the next step in improving hydrogel technology to meet these challenges. Central to the development of protein hydrogels will be a fundamental understanding of the answers to three questions; firstly what fraction of the protein remains folded post gelation. Secondly how do the micromechanics of the protein translate to the macromechanics of the gel. And thirdly how does the crosslink density of the protein network translate to the macromechanics of the gel [92][93], [207]. Questions one and three are the subject of the investigation reported in this thesis.

Figure 1.17: Schematic workflow undertaken during this thesis for the production of folded protein hydrogels.



### 1.5.1 Measuring the Folded Fraction of Protein in an FGP Hydrogel

A protein is of interest only in its folded state. This is a fundamental truth of biology as a fully disordered protein in isolation can contain no catalytic, responsive, or ligand binding function, although they themselves can act as a ligand for a folded protein [208][209]. Certain proteins are described as intrinsically disordered *in vivo* but this is not really true as apparently disordered proteins are in reality in a holding state as the polypeptide awaits a change in its environment or an interaction with a binding partner which will allow it to become ordered. Once folding is achieved these proteins can perform complex biological functions [210]. No polypeptides in a fully disordered state exhibit dynamic activity. This illustrates the earlier statement; only folded proteins are of interest, without tertiary structure they are little more than semi-flexible polymers with non-specific chemical groups attached.

This need for tertiary structure in order to incorporate biological functionality into a hydrogel means that the protein species from which the lattice is built must be demonstrated as folded post-gelation. A protein solution under “perfect” conditions will always contain a fraction of unfolded protein. This is due to the dynamic nature of protein folding which leads to a dynamic equilibrium between the folded and unfolded state, meaning that over time a protein domain may fold and re-fold stochastically many times, with a constant population in each state, as well as intermediate states [211], [212]. The position of this equilibrium can be shifted towards one state or another by altering the chemical, thermal, or mechanical conditions. All three of these may change during gelation as a result of swelling forces, the chemistry of the crosslinking reaction, and the input of radiation (light). This leads to the hypothesis that the gelation process is likely to shift the equilibrium of protein folding towards the unfolded state. Measuring how far this equilibrium shifts in individual systems is therefore

crucial in understanding the fraction of the protein which becomes unfolded and therefore useless in terms of functionality, and how the unfolded fraction affects the macromechanics of the material. Towards this end the first aim of this thesis is to precisely quantify the (un)folded fraction of protein in a hydrogel post-gelation. This is first measured by the use of circular dichroism adapted to allow measurement of a solid sample, and is described in Section 3 [213]. This technique provides a qualitative measure demonstrating the dominant secondary structure species present within the gel. The folded fraction was then quantitatively measured by use of cysteine shotgun labelling using the fluorescent dye IAEDANS of cysteine residues buried in the core of each protein domain. This is described in Section 4 and allows an accurate quantification of the fraction of protein domains which exist in an unfolded state post-gelation and post network relaxation [100]. The use of CD is suitable for use upon any protein hydrogel system. The use of the described cysteine shotgun labelling method requires a protein lattice species which contains no solvent exposed cysteines, and only one buried cysteine per domain, and is therefore less easily applicable to multiple hydrogel systems. The aim of these experiments is to demonstrate and quantify the folded fraction of protein domains within a true protein hydrogel. The techniques developed to do this will then be suitable for performing unfolded fraction versus macromechanical property studies.

### 1.5.2 Measuring Crosslinking Efficiency and its Correlation to Macromechanical Hydrogel properties

As has been discussed throughout Section 1 of this thesis the crosslink density of a hydrogel network directly affects the bulk mechanical properties of the material [207]. Therefore understanding the relationship between crosslink density and macroscopic mechanical properties is of vital importance in the rational mechanical design of hydrogels. The first step

towards this relies on an accurate quantification of the number of crosslinks in the gel and equating this to a total efficiency. The second step is correlating variations in crosslink density to variations in bulk gel mechanics.

Crosslink density is a complex number. Firstly the geometry of the crosslink sites is important as they will determine the overall shape of the network as geometry will affect not only the degree of crosslinking between monomers but also between clusters during different stages of network formation [214]. Secondly the total number of crosslinks will alter the network properties as more covalent bonds will equate to a higher potential energy storage capacity under stress. These two parameters together will equate to an overall crosslinking efficiency of number available versus number actually formed. Understanding how to regulate crosslinking density by way of alterations in crosslink geometry on the microscale will allow the design of specific network topologies. These different topologies will correlate to changes in bulk mechanical properties. Characterising and elucidating the laws governing this relationship is the second primary aim of this thesis; how does crosslink density on the microscale translate to mechanics on the macroscale.

## Section 2: Materials and Methods

### 2.1 Materials

#### 2.1.1 Centrifuges

Avanti J-26 XP Centrifuge (Beckman Coulter, Brea, CA, USA)

Bench top centrifuge: GenFuge 24D (Progen Scientific, London, UK)

Eppendorf 5810R Centrifuge (Fisher Scientific, Loughborough, UK Beckman Coulter)

#### 2.1.2 Incubators

Innova 43 Shaker Incubator (New Brunswick Scientific, USA)

Innova 44 Shaker Incubator (New Brunswick Scientific, USA)

SI600 orbital incubator (Stuart, Staffordshire, UK)

#### 2.1.3 Protein Purification Equipment

AktaPrime plus (GE Healthcare, Buckinghamshire, UK)

AktaPure (GE Healthcare, Buckinghamshire, UK)

AktaStart 2 (GE Healthcare, Buckinghamshire, UK)

HiLoad Superdex 75 pg preparative size exclusion chromatography column (GE Healthcare, Buckinghamshire, UK)

5ml HisTrap FF chromatography column (GE Healthcare, Buckinghamshire, UK)

5ml HiTrap DEAE Sepharose FF Ion Exchange chromatography column (GE Healthcare, Buckinghamshire, UK)

His SpinTrap protein purification kit (GE Healthcare, Buckinghamshire, UK)

Constant Systems Cell Disrupter CF1 (Constant Systems, Daventry, UK)

Vivaspin Spin Concentrator Molecular Weight Cut Off Columns (Sartorius, Gottingen, Germany)

Heto FD3 Lyophiliser (Thermo Electron, MA, USA)

#### 2.1.4 Spectrophotometers

NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, MA, USA)

Shimadzu UV-1800 (Shimadzu UK Ltd, Buckinghamshire, UK)

##### 2.1.4.1 Cuvettes

0.5mL UV cuvette (Sarstedt)

#### 2.1.5 PCR/Cloning Assembly Thermocycler

T100 thermal cycler (BioRad, CA, USA)

#### 2.1.6 Circular Dichroism

Chirascan™ CD Spectrophotometer (Applied Photophysics, London, UK)

#### 2.1.7 Fluorimeter

Invitrogen Qubit 4 Fluorimeter (Thermo Scientific, MA, USA)

Photon Technology International QM-1 spectrofluorimeter (Photon Technology International, NJ, USA)

#### 2.1.8 Microplate Fluorescence Readers

BMG FLUOStar Optima plate reader (BMG Labtech, Aylesbury, UK)

Corning® 96 Well Half Area Black with Clear Flat Bottom Polystyrene NBS™ Microplate (Corning GmbH, Wiesbaden, Germany)



### 2.1.9 Mass Spectrometer

M-class ACQUITY UPLC (Waters UK, Manchester, UK) interfaced to a Synapt G2S Q-IMT-TOF mass spectrometer (Waters UK, Manchester, UK).

### 2.1.10 Gel Electrophoresis

Slab Gel Electrophoresis Chamber AE-6200 (ATTO, Tokyo, Japan)

Powerpac 3000 (Bio-Rad Lab., Hercules, CA, USA)

### 2.1.11 Gel Ladders and Dyes

Precision plus protein dual colour standards protein ladder (BioRad, CA, USA)

Instant Blue Stain (Expedeon Protein Solutions, UK)

1kb DNA Ladder (NEB, NE, USA)

100kb DNA Ladder (NEB, NE, USA)

Gel Loading Dye, Purple (6X) (NEB, NE, USA)

### 2.1.12 DNA Purification Kits

QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany)

QIAquick PCR Purification Kit (Qiagen, Hilden, Germany)

### 2.1.13 DNA Primers

All PCR primers purchased from Eurofins MWG Operon, Ebersberg, Germany

### 2.1.14 Plasmids

A pGem-T-Easy vector (Promega) encoding a C47S C63S double mutant of human

Immunoglobulin domain 27 was kindly provided by Dr David Brockwell (Astbury Centre for Structural Molecular Biology, University of Leeds, UK).

pET14b vector (Novagen) was used as an expression vector for all 127 polyproteins and monomers.

All polyproteins were assembled according to the protocols described in Section 2.2.1.3.

#### 2.1.15 Bacterial Strains

*E. coli* BLR [DE3] pLysS cells (Merck Millipore, MA, USA)

- *F- ompT hsdSB(rB- mB-) gal dcm (DE3) Δ(srl-recA)306::Tn10 pLysS (CamR, TetR)*

*E. coli* DH5α competent cells (NEB, Hertfordshire, UK)

- *F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ- thi-1 gyrA96 relA1.*

*E. coli* SURE-2 Supercompetent cells (Agilent, CA, USA)

- *e14-(McrA-) Δ(mcrCB-hsdSMR-mrr)171 endA1 gyrA96 thi-1 supE44 relA1 lac recB recJ sbcC umuC::Tn5 (Kanr) uvrC [F' proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr].*

#### 2.1.16 Rheometer

Kinexus Malvern Pro DSR, equipped with 10mm parallel plate geometry (Malvern Panalytical,

#### 2.1.17 Design of Experiment Software

Design-Expert® Software Version 10 by Stat-Ease (Minneapolis, USA)

#### 2.1.18 Refractometer

Celti Refractometer

#### 2.1.19 Buffers, Solutions, and Growth Medias

Experiment specific buffers are described in detail in Section 2.2 as part of detailed protocol descriptions. General buffers used in multiple sections of this thesis are described below.

##### 2.1.19.1 Antibiotic stocks

100x Carbenicillin stock: 100mg/mL in ddH<sub>2</sub>O, 0.22μM filter sterilised.

100x Chloramphenicol stock: 0.34mg/mL in 100% ethanol, 0.22 $\mu$ M filter sterilised.

#### 2.1.19.2 SDS-PAGE Tris-Tricine Gel Solutions

Resolving gel buffer; 15% acrylamide 0.4% bis-acrylamide, 0.38M Tris-HCl pH8.8, 0.1% SDS (w/v) , 0.1% APS (w/v), 0.04% (v/v) TEMED.

Stacking gel buffer; 5% acrylamide 0.13% bis-acrylamide solution, 0.127M Tris-HCl pH6.8, 0.1% SDS (w/v), 0.1% APS (w/v), 0.04% TEMED (v/v).

SDS Gel loading gel buffer: 50mM Tris, 100mM DTT, 2% (w/v) SDS, 0.1% Bromophenol blue (w/v), 10% glycerol (v/v), pH6.8 with HCl.

Cathode Running Buffer (10x concentrated stock); 1 M Tris, 1 M Tricine, 1 % (w/v) SDS.

Anode Running Buffer (10x concentrated stock); 2 M Tris-HCl pH 8.9.

#### 2.1.19.3 Bacterial Growth Medias

LB medium: 25g/L granulated LB

2YT medium: 16g/L bactotryptone, 10g/L Yeast extract, 10g/L NaCl

Terrific broth media: 47.6g/L granulated Terrific broth, glycerol (variable concentration)

Autoinduction medium: 10g/L yeast extract, 20g/L bactotryptone, 50mM NH<sub>4</sub>Cl, 5mM Na<sub>2</sub>SO<sub>4</sub>, 25mM KH<sub>2</sub>PO<sub>4</sub>, 25mM Na<sub>2</sub>HPO<sub>4</sub>. 1.5g/L glycerol, 0.05g/L glucose, 0.2g/L lactose, 2mM MgSO<sub>4</sub>.

#### 2.1.19.4 DoE Protein Purification Buffers

Lysis buffer; 50mM Tris, 500mM NaCl, 20mM Imidazole, 100mL/L glycerol, 2mM MgCl<sub>2</sub>, 0.2mg/mL Lysozyme, pinch of DNase, 1 Piera<sup>TM</sup> protease inhibitor tablet per 50ml, pH 7.4.

Binding buffer; 20mM Tris, 500mM NaCl, 40mM Imidazole, pH 7.4.

Elution buffer; 20mM Tris, 300mM NaCl, 250mM Imidazole, pH8.

#### 2.1.19.5 Large-scale Protein Purification Solutions

Lysis buffer; 20mM Tris, 300mM NaCl, 5mM Imidazole, 150 $\mu$ L/L Triton X-100, 2mM PMSF, 1mM Benzamidine dihydrochloride, pH8.

Nickel Affinity Wash buffer; 20mM Tris, 300mM NaCl, 10mM Imidazole, pH8.

Nickel Affinity Wash buffer for TEV protease; 20mM Tris, 300mM NaCl, 10mM Imidazole, 2mM PMSF, 1mM Benzamidine dihydrochloride, pH8.

Nickel Affinity Elution buffer; 20mM Tris, 300mM NaCl, 250mM Imidazole, pH8.

Ion Exchange Elution Buffer; 25mM Sodium phosphate pH 7.4, 500mM NaCl.

Size Exclusion Running Buffer: 25mM Sodium phosphate pH 7.4; 3.2g/L Na<sub>2</sub>HPO<sub>4</sub>, 0.8g/L NaH<sub>2</sub>PO<sub>4</sub>.

#### 2.1.19.6 DNA Agarose Gel Solutions

50x TAE buffer; 2M Tris, 50mM EDTA, 2.8% Glacial Acetic Acid (v/v), pH 8.

Ethidium Bromide; 10mg/mL in water.

#### 2.1.19.7 Folded Fraction Measurement Solutions

Quenching Buffer; 25mM Sodium phosphate pH 7.4; 3.2g/L Na<sub>2</sub>HPO<sub>4</sub>, 0.8g/L NaH<sub>2</sub>PO<sub>4</sub>, 5mM  $\beta$ -mercaptoethanol

Digestion Buffer; 1M Guanidine Hydrochloride, 25mM Ammonium bicarbonate, 10mM DTT, 10mM CaCl<sub>2</sub>, 5mg/mL Trypsin, pH 8.

#### 2.1.20 Kits and Products

NEB Q5 Site Directed Mutagenesis Kit (NEB, Hertfordshire, UK)

DNA Purple Loading Dye (NEB, Hertfordshire, UK)

NEB<sup>®</sup> Golden Gate Assembly Kit (Bsal-HF<sup>®</sup> v2) (NEB, Hertfordshire, UK)

Vivaspin Protein Concentrator Spin Column (GE Healthcare, Buckinghamshire, UK)

#### 2.1.21 Transilluminators

IN GENIUS gel imaging dock (Syngene Bioimaging, Cambridge, UK).

#### 2.1.22 Gelation Lamp

460nm light emitting diode with intensity calibrated at 452nm, using a PM100D power meter equipped with a S470C thermal sensor.

## 2.1.23 Chemicals

Reagent	Supplier
A	
Acrylamide 30 % (v/v)	Severn Biotech, Kidderminster, UK
Agar	Fisher Scientific, Loughborough, UK
Ammonium persulphate (APS)	Sigma Life Sciences, MO, USA
Ammonium Bicarbonate	Sigma Life Sciences, MO, USA
Ammonium sulfate	Acros Organics, Geel, Belgium
Ammonium Chloride	Sigma Life Sciences, MO, USA
Acetic anhydride	Sigma Life Sciences, MO, USA
Acetonitrile (ACN)	Sigma Life Sciences, MO, USA
B	
Benzamide dihydrochloride	Sigma Life Sciences, MO, USA
Bromophenol Blue	Sigma Life Sciences, MO, USA
Bactotryptone	Sigma Life Sciences, MO, USA
$\beta$ -mercaptoethanol	Fisher Scientific, Loughborough, UK
C	
Carbenicillin	Formedium, Norfolk, UK
Calcium chloride (CaCl <sub>2</sub> )	Melford Laboratories, Suffolk, UK
Chloramphenicol	Sigma Life Sciences, MO, USA
D	
Dithiothreitol (DTT)	Formedium, Norfolk, UK
Dimethylformamide (DMF)	Fisher Scientific, Loughborough, UK
Diisopropylcarbodiimide (DIC)	Sigma Life Sciences, MO, USA
Dichloromethane (DCM)	Sigma Life Sciences, MO, USA
E	
Ethanol	Sigma Life Sciences, MO, USA
Ethylenediaminetetraacetic acid (EDTA)	Fisher Scientific, Loughborough, UK
Ethidium Bromide	Fisher Scientific, Loughborough, UK
2,2'-(Ethylenedioxy)diethanethiol (DODT)	Sigma Life Sciences, MO, USA

## G

Glycerol	Fisher Scientific, Loughborough, UK
Glacial Acetic Acid	Fisher Scientific, Loughborough, UK
Guanadine Hydrochloride	Sigma Life Sciences, MO, USA

## H

Hydrogen peroxide (H <sub>2</sub> O <sub>2</sub> )	Sigma Life Sciences, MO, USA
Hydrochloric acid (HCl)	Fisher Scientific, Loughborough, UK

## I

Imidazole	Sigma Life Sciences, MO, USA
Isopropanol	Fisher Scientific, Loughborough, UK
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Melford Laboratories, Suffolk, UK

## L

LB broth, granulated	Melford Laboratories, Suffolk, UK
----------------------	-----------------------------------

## O

Oxyma Pure	Sigma Life Sciences, MO, USA
------------	------------------------------

## P

Phenylmethanesulfonyl fluoride (PMSF)	Sigma Life Sciences, MO, USA
Potassium Phosphate Monobasic, KH <sub>2</sub> PO <sub>4</sub>	Sigma Life Sciences, MO, USA

## S

Sodium chloride (NaCl)	Fisher Scientific, Loughborough, UK
Sulphuric acid	Fisher Scientific, Loughborough, UK
Sodium hydroxide (NaOH)	Fisher Scientific, Loughborough, UK
Sodium dodecyl sulphate (SDS)	Severn Biotech, Kidderminster, UK
Sodium phosphate dibasic, Na <sub>2</sub> HPO <sub>4</sub>	Thermo Scientific, Surrey, UK
Sodium phosphate monobasic, NaH <sub>2</sub> PO <sub>4</sub>	Sigma Life Sciences, St. Louis, USA
Sodium Sulphate	Sigma Life Sciences, MO, USA

## T

Triton X-100	Sigma Life Sciences, MO, USA
Tris base	Fisher Scientific, Loughborough, UK
Tetramethylethylenediamine (TEMED)	Sigma Life Sciences, MO, USA
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Sigma Life Sciences, MO, USA
Tricene	Sigma Life Sciences, MO, USA
Terrific Broth, granulated	Sigma Life Sciences, MO, USA
Trifluoroacetic acid (TFA)	Sigma Life Sciences, MO, USA
Triisopropylsilane (TIPS)	Sigma Life Sciences, MO, USA
Tris(bipyridine)ruthenium(II) chloride (Ru)	Sigma Life Sciences, MO, USA
Trypsin from Bovine Pancreas (Trypsin)	Sigma Life Sciences, MO, USA

## U

Urea	MP biomedicals, Loughborough , UK
------	-----------------------------------

## Y

Yeast Extract	Sigma Life Sciences, MO, USA
---------------	------------------------------



## 2.2 Methods

### 2.2.1 Molecular Biology

#### 2.2.1.1 Plasmid DNA Preparation

Plasmid DNA was transformed into commercial competent *E.coli* cells according to well established protocol [100]. In brief ~100ng of DNA was added to 50µL competent cells then incubated on ice. The length of the 42°C heat shock was adapted according to manufacturer's instruction for each strain; *E.coli* BLR [DE3] pLysS 30 seconds, *E.coli* DH5α 30 seconds, *E.coli* SURE-2 11 seconds. Transformations were plated out on LB agar plates with appropriate antibiotic selections for plasmid selectivity. Plates were incubated for ~18 hours at 37°C. Single colonies were subsequently scraped from the surface of the agar plate and transferred to 5ml liquid LB medium containing the same antibiotic selection as the plate, and grown for ~18 hours at 37°C shaking at 200rpm. Cells were then harvested by centrifugation at 4000rpm for 30 minutes, and the supernatant discarded. Plasmids were then extracted and purified using QIAprep Spin Miniprep Kit (Section 2.1.12) according to manufacturer's protocol, with the final elution being performed with nuclease-free ddH<sub>2</sub>O. The concentration of DNA obtained was measured using a NanoDrop 2000 UV-Vis Spectrophotometer (Section 2.1.4) according to the optical density of the sample at 260nm ( $A_{260}$ ) in the Equation:

$$\text{Concentration } (\mu\text{g/mL}) = (A_{260} - A_{320}) \times \text{Dilution Factor} \times 50\mu\text{g/mL}$$

The sample was then diluted to ~100ng/µL, and 15µL submitted for sequencing at either Eurofins Genomics or Genewiz. Sequences were then aligned against the designed sequence to confirm identity.

### 2.2.1.2 Site Directed PCR Mutagenesis

All mutagenesis was performed using the NEB Q5 Site Directed Mutagenesis Kit (Section 2.1.20). Primers were designed for optimal  $T_m$  complementarity and efficiency using the online tool <https://nebasechanger.neb.com/> and purchased from Eurofins MWG Operon.

<b>Table 2.1;</b> Mutagenic primers used to create I27 sequence mutants and prepare the pET14b plasmid for use in Golden Gate Assembly:		
Sequence Mutation	Primers	Annealing Temperature (°C)
I27 Plasmid; Y9S	Forward: AAAGCCTCTG <b>TC</b> GGAGTAGAGG Reverse: TCCAATTCTATTAGTAGTTC	57
I27 Plasmid; H31Y	Forward: ACCTGATGTT <b>TAT</b> GGCCAGTGGA Reverse: TCAGAAAGTTCAATTTCAAAGTGG	62
I27 Plasmid; S44Y	Forward: TTTGACAGCT <b>TAT</b> CCTGACTCTGAAATC Reverse: GGCTGTCCTTTCAGCTTC	61
I27 Plasmid; S72C	Forward: CCTTCATAA <b>CTG</b> TAGCTGGGTA Reverse: ATCAGAATATGCTTCTTCCATC	61
I27 Plasmid; TEV-insertion	Forward: TTTTCAAGGTAGCCTAATAGAAGTGGAAAAG Reverse: TAAAGATTTTCCGAGTGATGGTGATGGTG	61
MCS Bsal site removal pET14b plasmid	Forward: ACTATAGGGAt <b>G</b> CCACAACGGTTTC Reverse: GAGTCGTATTAATTTCCGC	56
Ampicillin resistance operon Bsal site removal pET14b	Forward: GTGAGCGTGG <b>A</b> TCTCGCGGTA Reverse: CGGCTCCAGATTTATCAGCAATAAAC	67
Golden Gate cut site insertion pET14b MCS	Forward: GGCTACGGTCTCAGCTCGAGGATCCGG Reverse: GGCTACGGTCTCTTATATCTCCTTCTTAAAGTTAAACAAAATTATTTTC	61

The following reagents were assembled in a 0.2ml thin walled PCR tube:

Q5 Hot start high-fidelity master mix (2X): 12.5 $\mu$ L  
10 $\mu$ M Forward primer: 1.25 $\mu$ L  
10 $\mu$ M Reverse primer: 1.25 $\mu$ L  
Template DNA (25 ng/ $\mu$ l): 1 $\mu$ L  
Nuclease free water: 9 $\mu$ L

All mutagenic PCR reactions were performed according to the following thermal cycle protocol:

Step	Temperature ( $^{\circ}$ C)	Duration (s)	
Initial Duplex Melt	98	30	
Cyclical Duplex Melt	98	10	X30
Primer Annealing	50-72*	30	
Extension	72	30/kilobase	
Final Extension	72	300	
Hold	4	$\infty$	

\*The primer annealing temperature was adjusted according to the calculated optimal annealing temperature of the primer pair as detailed in Table 2.1.

Once the PCR reaction was complete the product was re-circularised by treatment with kinase, ligase, and Dpn1 enzymes. The following reagents were assembled in a 0.2ml thin walled PCR tube:

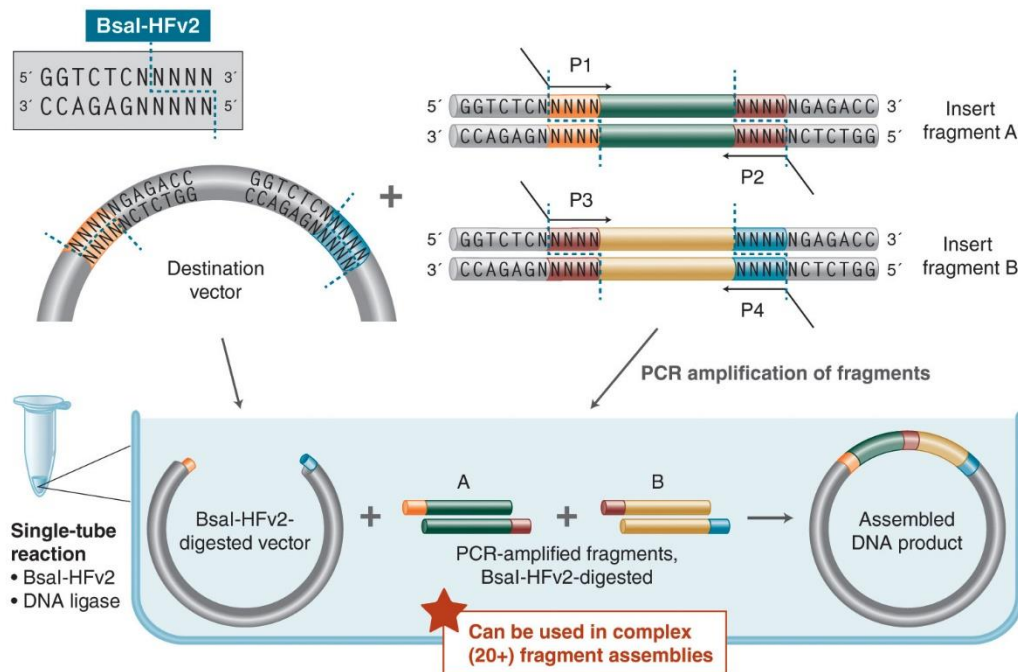
PCR product: 1 $\mu$ L  
NEB KLD reaction buffer (2x): 5 $\mu$ L  
NEB KLD enzyme mix (10x): 1 $\mu$ L  
Nuclease free water: 3 $\mu$ L

The reaction mixture was then incubated at room temperature for 5 minutes, after which 5 $\mu$ L of the reaction was transformed into competent *E.coli* cells (Section 2.1.15).

### 2.2.1.3 Golden Gate Polyprotein Assembly

The assembly of polyproteins was performed using a PCR-based Golden Gate protocol, and modified pET14b as the destination expression vector [215]. 127 mutants were amplified from the pGem-T-easy vector using specific forward and reverse primers to create 5 unique cassettes with different complimentary pairs of Bsal-cleavage induced sticky ends.

Figure 2.1: Summarised mechanism of Golden Gate Assembly cloning. Diagram reproduced from Genetic Engineering & Biotechnology News Magazine, September 1<sup>st</sup> 2018 (Vol. 38, No 15).



#### 2.2.1.3.1 pET14b Vector preparation

pET14b (Section 2.1.14) was subjected to three rounds of mutagenesis. These involved the removal of the native Bsal cleavage sequence in the MCS, the insertion of a double Bsal and flanking sequence in the MCS, and the removal of the Bsal cleavage sequence within the

Ampicillin resistance operon of the plasmid. All mutagenic reactions were carried and verified out according to the protocols described in Section 2.2.1.2.1 and 2.2.1.1.2.

#### 2.2.1.3.2 Cassette PCR Amplification

Individual cassettes were amplified by PCR from I27 monomer sequences contained in pGem-T-easy vectors (Section 2.1.14) to produce short DNA fragments. Below is a summary of the DNA sequence in the Multiple Cloning Site of the pGem-T-easy vector containing a single I27 cassette (start/stop codons in red). In blue are the wild type and mutant alternative sequences second corresponding to “Y9S”, “H31Y”, and “S44Y”:

```
TATATTAAGAAGGAGAATGCATCACCATCACCATCACTCGAGCCTAATAGAAGTGGAAAAGCCTCTG
(TAC/TCG)GGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGA
TGTT(CAC/TAT)GGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCT(TCC/TAT)CCTGACTCTGA
AATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAG
GTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAAATG
```

The following primers were manually designed to yield unique cassettes from a single DNA template. This is illustrated in Table 2.2:

Cassette	Forward primer	Reverse primer	Annealing Temperature (°C)
1	GGCTACGGTCTCATATTAAGAAGGAGATATACCATG	GTAGCCGGTCTCTAATGGTCGGCGCCACGCTCAGCAATTCTTTCAC TTCAGATTGG	59
2	GGCTACGGTCTCACATTCTAATAGAAGTGGAAAAGC	GTAGCCGGTCTCTGCTCGCCAGACCAATAACGGTCAATTCTTTCAC TTCAGA	54
3	GGCTACGGTCTCAGAGCCTAATAGAAGTGGAAAAGC	GTAGCCGGTCTCTACAATGGTGCCGCTCAGCGCCAATTCTTTCAC TTCAGA	54
4	GGCTACGGTCTCATGTGCTAATAGAAGTGGAAAAGC	GTAGCCGGTCTCTCGCCAGGCTACCGTAATAACCAATTCTTTCAC TTCAGA	54
5	GGCTACGGTCTCAGGCGCTAATAGAAGTGGAAAAGC	GTAGCCGGTCTCTGAGCTACAATTCTTTCAC TTCA	55

Sequences in black denote I27 complementarity

Sequences in purple denote Golden Gate annealing sticky ends

Sequences in red denote linkers

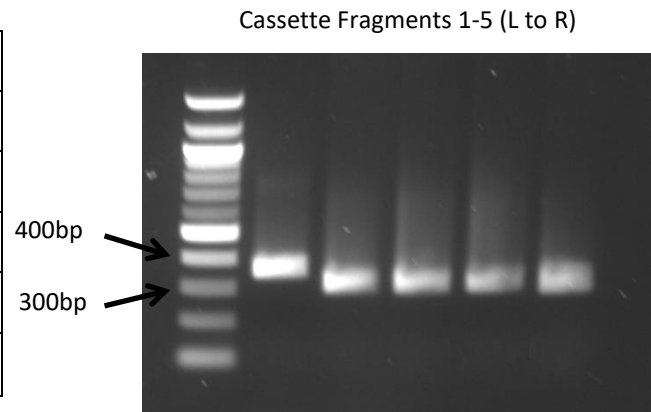
Sequences in green denote BsaI cleavage sites

Sequences in gold denotes the new stop codon

All PCR amplifications were carried out according to the protocol described in Section 2.2.1.2. The cyclical extension time was constant at 30 seconds.

Once the PCR reactions were complete the size of the products was analysed by agarose gel electrophoresis (Section 2.2.4.2). 5µL PCR product was added to 5µL Purple Loading Dye (NEB).

Fragment	Expected Length (bp)
Cassette 1	363
Cassette 2	318
Cassette 3	318
Cassette 4	318
Cassette 5	304



The remaining 20µL of PCR product was then cleaned up using a QIAquick PCR Purification Kit (Section 2.1.12). 1µL of the purified product was then used in a Qubit fluorimetric assay to precisely quantify the DNA fragment concentration according to the manufacturers protocol (Section 2.1.7).

### 2.2.1.3.3 Assembly

The DNA fragments, target pET14b vector, and Golden Gate Assembly Kit reagents were assembled in a 0.2ml PCR tube according to the following recipe in order to create a 2:1 molar ratio between each fragment and the target vector:

Reagent	Mass/Volume
Modified pET14b Vector	75ng
Cassette 1	11.8ng
Cassette 2	10.3ng
Cassette 3	10.3ng
Cassette 4	10.3ng
Cassette 5	9.9ng
10x T4 DNA Ligase Reaction Buffer	2 $\mu$ L
Golden Gate Assembly 20x Master Mix	1 $\mu$ L
Nuclease free water	Up to 20 $\mu$ L

The reaction mixture was then placed in a thermocycler and subjected to the following thermal program:

Stage	Temperature ( $^{\circ}$ C)	Duration (minutes)	
Bsal Cleavage	37	1	X30
Ligation	16	1	
Final Bsal Cleavage	60	5	

Once the assembly program was complete 5 $\mu$ L of the assembly reaction was transformed into SURE-2 supercompetent cells (Section 2.1.15) and plated out on Carbenicillin LB agar plates.



#### 2.2.1.3.4 Colony Screening

Single colonies from the transformation plate were streaked out within numbered grid squares on a second identical LB agar plate and incubated overnight at 37°C. PCR reaction mixtures were prepared in 0.2ml PCR tubes in the following volumes:

Vent DNA Polymerase:	1µL
10x ThermoPol Reaction Buffer:	5µL
10mM dNTP Mix:	2µL
100mM MgSO <sub>4</sub> :	3µL
10µM Forward primer:	2.5µL
10µM Reverse primer:	2.5µL
Nuclease free water:	34µL

Colony Screen primers (T<sub>m</sub>= 58°C):

Forward: AACTTTAAGAAGGAGATATACCATGCATCACCA

Reverse: GGATCCTCGACATATTCAAC ACACAATTCTTT

A P200 pipette tip was then used to scrape a portion of a bacterial colony off the streaked agar plate and was dunked into an aliquot of the PCR reaction mixture twice. The reaction mixture was then placed in a thermocycler and subjected to the following PCR program:

Step	Temperature (°C)	Duration (s)	
Initial Cell Lysis and Duplex Melt	95	120	
Cyclical Duplex Melt	95	30	X35
Primer Annealing	57	30	
Extension	72	60	
Final Extension	72	300	
Hold	4	∞	

Once the PCR program had finished 10µl 6x purple DNA loading dye was added to each reaction and mixed thoroughly by pipetting action. 15µL of each reaction was then loaded and run on an agarose DNA electrophoresis gel (Section 2.1.19.6). The expected band indicating a successful product was ~1500bp. If such a band was present in a PCR product, the corresponding colony on the streaked agar plate was prepared for plasmid sequencing as described in Section 2.2.1.1.

## 2.2.2 Polyprotein Expression Optimisation

### 2.2.2.1 Design of Experiment Methodology

Three media were selected for assessment; 2YT, Terrific Broth (TB), and Autoinduction. The input variables selected were temperature of growth, length of incubation after induction (where applicable), and mass per litre of glucose (2YT) and glycerol (TB). High, low, and intermediate values of each variable were selected by analysis of the wider literature, and are summarised in Table 2.3.

<b>Table 2.3: DoE Input Variables</b>			
<b>Input Variable</b>	<b>+1 Value</b>	<b>0 Value</b>	<b>-1 Value</b>
Temperature of growth (°C)	37	28	19
Time of incubation (post-induction where applicable) (Hours)	48	36	24
% Glucose/Glycerol (w/v or v/v)	1	0.6	0.2

The output variables analysed were the OD600 prior to centrifugation, the total dry weight pellet mass after harvesting, and the relative expression of purified (I27)<sub>5</sub> polyprotein.

### 2.2.2.2: Test Culture Growth:

(127)<sub>5</sub> [99] was transformed into BLR DE3 pLysS cells and plated out on a carbenicillin/chloramphenicol LB agar plate. A monoclonal glycerol stock of BLR DE3 pLysS (127)<sub>5</sub> was then prepared by scraping a single colony from the transformation plate into 5ml LB media containing identical antibiotics. This was then incubated overnight at 37°C 200rpm shaking. 0.5ml of this culture was then mixed with 0.5ml sterile 100% glycerol, snap frozen, and stored at -80°C. All test growth media (Section 2.1.19.3) were prepared to 50ml final volume in 200ml baffled flasks and autoclaved. Starter cultures were prepared by scraping a P200 pipette tip across the surface of the glycerol stock and dropping the tip into 5ml sterile LB media with carbenicillin and chloramphenicol, and incubated overnight at 37°C 200rpm shaking. 200µL of the starter culture was then used to inoculate each of the 50ml test cultures, in combination with antibiotics. The test cultures were then incubated at the appropriate temperature 200rpm shaking. The OD<sub>600</sub> of 2YT and TB cultures was measured every 30 minutes until an OD<sub>600</sub> of 0.6-0.8 was reached. Protein expression was then induced by the addition of IPTG to a final concentration of 1mM, and expression was allowed to proceed for the appropriate length of time.

### 2.2.2.3: Protein Purification and Response measurement

Once the pre-prescribed incubation time of each culture was complete 1ml of culture was removed and the OD<sub>600</sub> measured as described in Section 2.2.4.4. The remaining 49ml of culture was centrifuged at 25000rpm (Avanti J-26 XP, JLA 25.50 rotor) for 30 minutes in centrifuge tubes of known mass. The supernatant was then carefully poured and pipetted off, and the tubes re-weighed allowing the calculation of the wet bacterial pellet mass. The pellet was then re-suspended in 2ml lysis buffer (Section 2.1.19.4) and left on rollers for 60 mins at

room temperature. The cell lysate was then spun down at 13000rpm (Avanti J-26 XP, JLA 25.50 rotor) for 25 mins. 600 $\mu$ L of the protein-containing supernatant was run through a His SpinTrap column at a time until all had been run through according to the manufacturer's protocol (Section 2.1.3). Protein was then eluted into 400 $\mu$ L elution buffer (Section 2.1.19.4). 5 $\mu$ L protein-containing eluent was mixed with 5 $\mu$ L 5mg/mL  $\beta_2$ M protein standard, and 10 $\mu$ L SDS-PAGE gel loading buffer (Section 2.1.19.2), and run on a 15% (w/v) SDS-PAGE polyacrylamide gel (Section 2.2.4.1). The relative quantity of (I27)<sub>5</sub> produced from each culture was measured using densitometry (Section 2.2.2.4).

#### 2.2.2.4: Densitometry:

SDS-PAGE gels were stained with Expedeon Instant blue stain (Section 2.1.19), and images taken using an IN GENIUS gel imaging dock. All subsequent densitometry measurements were performed using ImageJ software. The (I27)<sub>5</sub> band peak (52219.15Da) pixel density was divided by the  $\beta_2$ M standard peak (11kDa) pixel density in each lane to give a ratio between the two. This ratio was used as a measure of relative protein expression between samples.

#### 2.2.2.5 Subsequent Modelling of Expression Landscape

All output results were correlated with input variables and modelled using Design-Expert<sup>®</sup> Software Version 10.

### 2.2.3 Large Scale Expression and Purification of All Protein/Polyproteins

#### 2.2.3.1 Starter Cultures and Media

Autoinduction medium (AI) (Section 2.2.19.3) was prepared to a final volume of 500ml in 2L baffled flasks and autoclaved. A starter culture was prepared by scraping a single colony from

a transformation plate or the surface of a glycerol stock using a P200 pipette tip, and dropped into 5ml sterile LB medium containing appropriate antibiotics. Cultures were incubated overnight 37°C 200rpm shaking.

#### 2.2.3.2 Inoculation and Growth

Growth media were pre-warmed overnight at 28°C. 500ml AI medium was inoculated with 2ml starter culture and appropriate antibiotics. Cultures were then placed in a shaking incubator (Section 2.1.2) at 200rpm and 28°C for 24 hours.

#### 2.2.3.3 Cell Harvesting and Lysis

Cultures were centrifuged at 8000rpm (Avanti J-26 XP, JA8.1 rotor) for 30 minutes and chilled to 4°C. The pellet was re-suspended in 100ml lysis buffer (Section 2.1.19.5) using a Digital IKA S2 Ultra-TURRAX homogeniser, and incubated on rollers at room temperature for 1 hour. The cells were lysed further by use of a Constant Systems Cell Disrupter CF1 (Section 2.1.3) at a running pressure of 40kpsi according to manufacturer's protocol. The flow through was then centrifuged at 25000rpm (Avanti J-26 XP, JLA 25.50 rotor) for 30 minutes and the pellet discarded.

#### 2.2.3.4 Chromatography

Cell lysate was applied to 5ml HisTrap FF chromatography columns (Section 2.1.3) for Ni-NTA purification. A wash step was performed using 5 column volumes of wash buffer (Section 2.1.19.5). Protein was eluted stepwise using high imidazole concentration elution buffer (Section 2.1.19.5). 10ml fractions were collected continuously. All elution peaks were analysed by SDS-PAGE gel (Section 2.2.4.1). Fractions containing the protein of interest were pooled and diluted 10-fold with 25mM sodium phosphate buffer pH 7.4 (NaPi) (Section 2.1.19.5). This was then applied to a 5ml HiTrap DEAE Sepharose FF Ion Exchange

chromatography column (Section 2.1.3) for Ion Exchange (IEX) purification. A gradient elution from 0-100% ion exchange elution buffer (Section 2.1.19.5) was performed over a 300ml volume. All elution peaks were analysed by SDS-PAGE gel.

#### 2.2.3.4.1 TEV-Cleavage of Proteins Destined for Gel Formation

2.5mg of TEV protease (Section 2.2.3.7) was added to the pooled protein fraction of interest after IEX. This mixture was incubated overnight at 8°C on rollers.

#### 2.2.3.5 Size Exclusion Chromatography

Cleavage/IEX product was concentrated to ~45ml using 30kDa MWCO spin concentrators (Section 2.1.3). This was then run applied to a HiLoad Superdex 75 pg preparative size exclusion chromatography (SEC) column (Section 2.1.3) in multiple 5ml injections. Individual peaks from multiple runs were pooled and analysed on SDS-PAGE gel. Proteins of interest were submitted for mass spectrometry analysis at >5µM concentration (Section 2.2.5.6) for identity and purity assessment.

#### 2.2.3.6 Lyophilisation

SEC product was dialysed into purified water over ~8 hours with dialysis buffer refreshment every hour at room temperature. The solution was then concentrated to ~100ml using 30kDa MWCO spin concentrators, then centrifuged at 25000rpm (Avanti J-26 XP, JLA 25.50 rotor) for 30 minutes. The supernatant was then transferred to a 1L round bottom flask and snap frozen in a dry ice-ethanol bath. The flask was then attached to a Heto FD3 Lyophiliser (Section 2.1.3) and freeze-drying was allowed to proceed for ~4 days. Final lyophilised product was stored at -80°C.

### 2.2.3.7 TEV Protease Production

TEV protease starter culture was prepared, expressed, and purified as described in Sections 2.2.3.1-2.2.3.3. No cleavage step was performed and all chromatography up to but not including SEC chromatography was performed in the presence of 2mM PMSF and 1mM Benzamidine dihydrochloride. The final product was not lyophilised, and was instead concentrated to ~1.25mg/mL in NaPi and snap frozen.

## 2.2.4 Biochemistry techniques

### 2.2.4.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS resolving and stacking gel solutions were prepared as described in Section 2.1.19.3. The resolving gel was prepared and poured first into a sealed casting chamber. A layer of 100% ethanol was poured on top of the gel solution to prevent evaporation. Once the resolving gel had set the stacking gel was prepared and the ethanol washed off the resolving gel. The stacking gel was poured to fill the remaining volume of the casting chamber and a 14-well comb inserted. Once the stacking gel had set the comb and rubber sealing gasket were removed before use. The gel within its casting chamber was placed in the cathode chamber of a Slab Gel Electrophoresis Chamber AE-6200. The anode chamber was filled with 1x Anode Running Buffer, and the cathode chamber with 1 x Cathode running buffer (Section 2.1.19.2). Unless otherwise stated 10 $\mu$ L of sample was mixed with 10 $\mu$ L 2x SDS loading buffer (Section 2.1.19.2) and 10 $\mu$ L was loaded into a gel well. 10 $\mu$ L of Precision plus protein dual colour standards protein ladder (section 2.1.11) was loaded into one or more wells for molecular weight determination. A 30mA/gel current was then used to initiate the migration of the samples into and through the stacking gel. Once the sample reached the resolving gel the

current was increased to 60mA/gel until the blue dye front reached the bottom of the gel. The current was then switched off and the gel removed from the casting chamber. The gel was then transferred into a vessel containing Instant Blue Stain and was allowed to stain for ~1 hour before transfer to purified water. Gels were subsequently imaged and analysed using an IN GENIUS gel imaging dock (Section 2.1.21).

#### 2.2.4.1.1 In-gel Crosslinking versus Irradiation Time

SDS-PAGE gels were prepared as described in Section 2.2.4.1. Protein was re-suspended in 25mM NaPi pH 7.4 to ~5 $\mu$ M. Ammonium persulphate and Tris(bipyridine)ruthenium(II) chloride were added to yield a final concentration of 10mM and 100 $\mu$ M respectively, in a final volume of 20 $\mu$ L. Samples were kept in the dark before and after irradiation. Samples were irradiated as described in Section 2.2.5.1 for increasing lengths of time. 20 $\mu$ L of SDS loading dye (Section 2.1.19.2) was added to each sample and rapidly mixed. The SDS-PAGE gel was then loaded, run, and analysed as described in Section 2.2.4.1.

#### 2.2.4.2 DNA Agarose Gel Electrophoresis

Agarose was added to 1x TAE buffer (Section 2.1.19.6) for a final 1.5% w/v concentration. The mixture was microwaved for 90 seconds then stirred until the agarose was fully dissolved. Ethidium Bromide (Section 2.1.19.6) was added at 1 $\mu$ L/mL of agarose solution and mixed by swirling in a fume hood. Comb(s) were inserted into a running tank and the gel solution poured into the tank in a fume hood. The gel was allowed to set for ~30 minutes. DNA samples were combined in a 6:1 ratio with 6x Purple Loading Dye (Section 2.1.23) and 15-20 $\mu$ L sample was loaded per well. Each end well was loaded with 5 $\mu$ L either 1kb or 100bp DNA ladder (Section 2.1.11) pre-mixed with loading dye. Initial current was 100V until all samples had



entered the gel, then increased to 150V until the dye front reached the end of the gel. Once complete the gel was transferred directly into the IN GENIUS gel imaging dock and imaged.

### 2.2.4.3 Protein Concentration Determination

#### 2.2.4.1 Low Concentrations

Samples were placed in UV cuvettes (Section 2.1.4.1) and their optical density at 280nm was recorded using a Shimadzu UV-1800 spectrophotometer (Section 2.1.4) blanked with appropriate buffer. This absorbance was incorporated into the Beer-Lambert equation to calculate the concentration of protein. Theoretical molecular extinction coefficients were calculated using the Expassy ProtParam online tool. If the initial A<sub>280</sub> reading was >1 the sample was diluted until it was <1 and used to calculate the original concentration.

#### 2.2.4.2 High Concentrations

5µL of sample was pipette mixed with 95µL diluent buffer. A stepwise dilution was performed to yield a final sample diluted 200x from the original. This was then analysed as described in Section 2.2.4.1.

### 2.2.4.4 Optical Density 600 Measurements

1ml aliquots of cell culture were transferred to UV cuvettes (Section 2.1.4.1) and placed in a Shimadzu UV-1800 spectrophotometer (Section 2.1.4) blanked with equivalent fresh medium. Absorbance at 600nm was then measured. In instances where the optical density exceeded 1 the sample was diluted until a reading <1 was achieved, then multiplied to give the original absorbance.

## 2.2.5 Biophysical Techniques

### 2.2.5.1 Hydrogel Preparation

4-5 0.5ml Eppendorf tubes were filled ~1/4 full of lyophilised protein. 150 $\mu$ L 25mM NaPi pH 7.4 was then added to the first tube and gently flicked to force all the protein and liquid to the bottom of the tube. The tube was then placed on a rotator until the protein powder had dissolved. The solution was gently tapped to the bottom of the tube then transferred to the second tube of protein and the process repeated. This was repeated until all tubes of protein had been dissolved. The empty tubes were spun down at 5000rpm for 5 minutes and the residual liquid transferred to the final solution. The final solution was then sonicated in a sonicated water bath for 3 minutes at room temperature, then centrifuged at 5000rpm for 5 minutes. The supernatant was transferred to a clean tube. The protein concentration was measured as described in Section 2.2.4.2. Typical concentration of stock ranged from 30-40mg/mL. The protein stock solution was diluted to 25mg/mL using 10% final volume of 1mM Ru for a final concentration of 100 $\mu$ M, and the remaining diluent volume of an APS solution of sufficient concentration to yield a final APS concentration of 10mM. The Ru and APS were placed in a new tube first and the requisite volume of protein added afterwards and mixed thoroughly and rapidly. The gel solution was immediately transferred to whichever location was required for gel formation (rheometer, 2ml Eppendorf, CD cuvette) and in whatever volume was required. Gelation was initiated by irradiation for a given length of time using a 460nm light emitting diode lamp calibrated at 452nm ( $\lambda_{\text{max}}$  Ru(BiPy)<sub>3</sub>), using a PM100D power meter equipped with a S470C thermal sensor.

### 2.2.5.2 Thermodynamic Stability Measurements of Polyproteins by Urea Denaturation

Protein was re-suspended in 25mM NaPi pH 7.4 to ~50 $\mu$ M. Urea was dissolved in 25mM NaPi pH 7.4 to >8.8M. The concentration of the urea stock was measured using a Ceti Refractometer and calculated using Equation 2.1 [216]:

$$C = 117.66(\Delta_{\eta}) + 29.753(\Delta_{\eta})^2 + 185.56(\Delta_{\eta})^3$$

Where  $\Delta_{\eta}$  is the difference in the measured refractive index between the buffer and the urea-containing buffer. The urea stock concentration was then adjusted to 8.8M. The 8.8M urea stock and an aliquot of NaPi were then each diluted 9/1 with the protein stock. The 8M and 0M urea protein stocks were then mixed in appropriate volumes to yield 2, 4, and 6M urea protein stocks, which were then mixed to create samples of 0.2M urea increments between 0.2 and 7.8M. All samples were incubated for ~18 hours at room temperature. All measurements were performed using a Photon Technology International QM-1 spectrofluorimeter (Section 2.1.7) at room temperature. Excitation was set to 280nm and emission was measured continuously for 30 seconds at 317nm for each sample and averaged.

### 2.2.5.3 Hydrogel Folded Fraction Measurement

#### 2.2.5.3.1 Labelling

Hydrogels solution was prepared as described in Section 2.2.5.1. 35 $\mu$ L gel solution was placed in the bottom of a 2ml Eppendorf tube, and gelled as described in Section 2.2.5.1. IAEDANS was dissolved in NaPi 25mM pH 7.4, and appropriate urea concentration where required. Concentration of IAEDANS was measured using a 1/400 dilution with NaPi,  $\lambda = 336\text{nm}$ ,  $\epsilon = 5700$ . The concentration of IAEDANS was then adjusted to create a 131:1 mols excess between 1ml

of IAEDANS solution and the mols of cysteine present in a 35 $\mu$ L hydrogel. 1ml of IAEDANS solution was then added to the gel-containing Eppendorf and incubated at room temperature in the dark on a rotator for 2 hours. The IAEDANS solution was pipetted off, and replaced with 2ml Quenching Buffer (Section 2.1.19.7). Samples were rotated in the dark at room temperature for 6 hours with the quenching buffer being refreshed every hour. The quenching buffer aliquot was removed and replaced with 1ml digestion buffer (Section 2.1.19.7). Samples were incubated for ~18 hours in the dark at room temperature on a rotator. Once the gel had fully dissolved the solution was transferred to a fresh pre-weighed Eppendorf tube, then reweighed allowing the volume of solution to be calculated (assuming a fluid density of 1g/cm<sup>3</sup>).

#### 2.2.5.3.2 Fluorescence Measurement

A standard curve of IAEDANS concentration versus fluorescence intensity at 490nm (excited at 336nm) was constructed using IAEDANS dissolved in digestion buffer and concentration measured by A336 measurement (Section 2.2.5.3.1). Each standard's 490nm emission was measured continuously for 30 seconds and averaged. Hydrogel sample fluorescence was then measured at 490nm for 30 seconds and the concentration of IAEDANS determined by use of the equation  $y=mx+c$  applied to the standard curve. The number of moles of IAEDANS-labelled cysteine was then back calculated and divided by the total moles present in the original gel to give a % labelled fraction equivalent to the unfolded fraction in the gel.

These measurements were performed in tandem with dityrosine quantification (Section 2.2.5.4) and so a dityrosine standard (Section 2.2.5.4) was excited at 336nm and its emission at 490nm was measured continuously for 30 seconds for subtraction from the Hydrogel sample measurement.

## 2.2.5.4 Measurement of Crosslinking Efficiency

### 2.2.5.4.1 Gel Preparation and Degradation

Hydrogel solution was prepared, gelled, labelled with IAEDANS, and degraded as described in Section 2.2.5.3.

### 2.2.5.4.2 Dityrosine Quantification

A standard curve of dityrosine concentration versus fluorescence intensity at 410nm (315nm excitation) was constructed in digestion buffer (Section 2.1.19.7). Each standard's 410nm emission was measured continuously for 30 seconds and averaged. After proteolytic cleavage (Section 2.2.5.3) each hydrogel sample fluorescence was then measured at 410nm for 30 seconds and the concentration of dityrosine determined by use of the equation  $y=mx+c$  applied to the standard curve. The number of moles of dityrosine was then back calculated and divided by the theoretical maximum number of moles (moles of tyrosine in the gel/2) to give a % crosslinking efficiency. This could then be used to give an average number of crosslinks per protein monomer.

These measurements were performed in tandem with folded fraction measurement (Section 2.2.5.3) and so an IAEDANS standard (Section 2.2.5.3) was excited at 315nm and its emission at 410nm was measured continuously for 30 seconds for subtraction from the Hydrogel sample measurement.

### 2.2.5.5 Circular dichroism

All measurements were performed using a Chirascan™ CD Spectrophotometer (Section 2.1.6) regulated at 25°C unless otherwise stated.

### 2.2.5.5.1 Low Concentration Spectra

Protein was prepared to 0.2mg/mL in 25mM NaPi pH 7.4 as described in Section 2.2.4.1. 200 $\mu$ L of sample was placed in a 1mm x 350ul 110-QS cuvette (Hellma). The CD spectrophotometer set up used was:

Wavelength (nm)	280
Bandwidth (nm)	2
Low Scan Wavelength (nm)	180
High Scan Wavelength (nm)	260
Time-per-point (s)	1
Step (nm)	1
Measurement Type	Absorbance
Output Units	Millidegrees

A full spectrum was then automatically obtained in triplicate and averaged. Spectra were converted from ellipticity ( $\theta$ ) to Mean Residue Ellipticity (MRE) by use of the following Equations where  $c$  is protein concentration,  $l$  is the path length, and  $R$  is number of amino acid residues per protein molecule:

$$2.2: [\theta] = \frac{\theta}{10 \times c \times l} \quad 2.3: [\theta]_{MRE} = \frac{[\theta]}{R-1}$$

All subsequent secondary structure analysis was performed using DiChroWeb online analysis software [217].

### 2.2.5.5.2 Thermal Melt

Samples were prepared and the spectrophotometer set up as described in Section 2.2.5.5.1. An initial spectrum was obtained at 25°C, and subsequently every degree up to 90°C. Samples were incubated at each temperature for 180 seconds with a temperature tolerance of  $\pm 0.2^\circ\text{C}$ .

The sample was then cooled back to 25°C and a final spectrum obtained. Data was analysed by Photophysics Global3 software and manual fitting.

#### 2.2.5.5.3 Solid State and High Concentration Spectra

Hydrogel solution was prepared as described in Section 2.2.5.1. 5µL was placed in a 106-QS, 0.01 mm path length demountable cell cuvette (Hellma) and sealed with parafilm. A solution spectrum was obtained using the following set up:

Wavelength (nm)	280
Bandwidth (nm)	1-2
Low Scan Wavelength (nm)	180
High Scan Wavelength (nm)	260
Time-per-point (s)	1
Step (nm)	1
Measurement Type	Absorbance
Output Units	Millidegrees

The sample was then removed from the instrument and irradiated for 2 minutes as described in Section 2.2.5.1. The sample was then returned to the instrument and a spectrum accumulated every 15 minutes for 1 hour.

#### 2.2.5.6 Mass spectrometry

All mass spectrometry experiments presented in this thesis were performed by Rachel George on a M-class ACQUITY UPLC interfaced to a Synapt G2S Q-IMT-TOF mass spectrometer.

#### 2.2.5.7 Rheology

Hydrogel solutions were prepared as described in Section 2.2.5.1. 35-100µL of gel solution was loaded onto the transparent acrylic bottom parallel plate of the rheometer and the

geometry lowered to give a gap height of 0.3-1mm. The sample was trimmed and surrounded by low viscosity silicone oil. A timesweep measurement was then initiated; 1% strain at 1Hz frequency. After one minute the under-bottom plate fitted lamp (Section 2.2.5.1) was switched on for 5 minutes. The timesweep then continued measurement for 1 hour. All subsequent experiments are described in Section 3.



## **Section 3: Optimisation of Polyprotein Cloning, Design, and Expression**

### **3.1 Introduction**

The primary aim of the study reported in this thesis is to investigate the relationship between protein hydrogel crosslink density/geometry and the macromechanical properties of the hydrogels they form. This has been investigated by the rational modulation of the number, distribution, and geometric pattern of crosslink sites (tyrosine residues) in the monomeric polyprotein building block. It is therefore necessary to be able to precisely regulate the positions and number of tyrosine residues in each individual I27 domain of each polyprotein construct. The DNA sequence of each I27 domain of the polyprotein is identical meaning that a mutagenic PCR primer designed to mutate a single domain will likely mutate every other as well. This makes precise crosslink density modulation via PCR mutagenesis of the complete homo-pentamer DNA construct impossible. Instead monomer sequences containing every combination of desired mutations were generated to build a mutant domain library, before repeatable re-assembly into pentamers of any mutant domain order/composition. Traditional cloning methods involving the stochastic restriction digestion and ligation of single DNA fragments into a digested vector were too slow and inefficient to be useful for the production of the number of mutants envisaged for this study. The more appropriate technique of Gibson Assembly was trialled but proved inappropriate; due to the non-restricted nature of the exonuclease activity during assembly the identical I27 sequences were exposed leading to preferential ligation of I27 sequences into a new monomer rather than ligation of the order-determining unique linker sequences [218]. The molecular biology technique which ultimately proved the most suitable was Golden Gate assembly. Unique

linker regions and flanking type II restriction endonuclease sites were designed and optimised to utilise a PCR-fragment/intact vector workflow Golden Gate method. This resulted in a repeatable assembly efficiency of  $23.5 \pm 9.1\%$ .

Bulk rheological characterisation of soft materials requires significant sample volumes, usually in the scale of millilitres, with a general correlation between larger volumes and lower experimental errors [219]. Material requirements can be minimised by the use of smaller radius geometries and smaller gap heights, but both of these correlate to increased error [219]. Smaller geometries correspond to greater ratios of inertia to sample contact friction meaning that the contribution of inertia to the measurement becomes more difficult to correct for. Smaller gap heights reduce the depth of laminar flow in the sample leading to inaccuracies in the measurement of viscosity, and an increase in the particle size to sample height ratio can cause grinding of particles between the plates and lead to oddities in moduli measurements. These facts mean the larger sample volume one is willing to use the lower the total measurement error tends to be, with the minimal sample volume possible at the beginning of this project being 0.5mL. The formulation of protein hydrogels has historically been set at  $\sim 100\text{mg/mL}$  as this is analogous to the total protein concentration in the extracellular matrix of mammalian tissues [93], [97]. These twin requirements; large sample volumes and high protein concentration formulations, makes protein hydrogel characterisation extremely expensive in terms of the mass of protein required to perform an in-depth and reliable study. A bottleneck was therefore identified in the planned workflow of this project as the historic expressability of I27 polyproteins (10-15mg/L) was insufficient to meet the protein mass requirements of this study. The industry-standard technique Design of Experiments was utilised to rationally optimise the expression of (I27)<sub>5</sub> polyproteins by statistically exploring a large design space based on minimal wet-lab test expression data. This

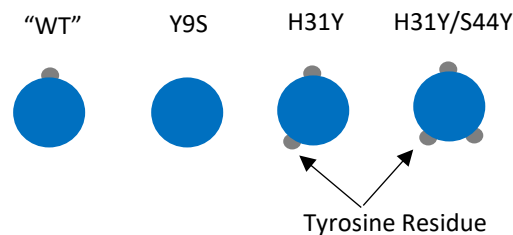
ultimately resulted in an increase in typical protein yield from 10-15mg/L to 100-150mg/L of bacterial culture [179].

### 3.3 Polyprotein Cloning

In order to make possible the rapid production of multiple pentamer polyprotein DNA sequences with specific patterns and orders of tyrosine residues a library of mutant domains was first created via PCR mutagenesis (Section 2.2.1.2, Table 1). Four domain types were created; “wild type” (no mutagenesis of the I27 sequence), Y9S (native tyrosine residue substituted), H31Y, and H31Y/S44Y double mutant (Figure 3.1).

---

Figure 3.1: Schematic Representation of I27 Mutant Domain Library

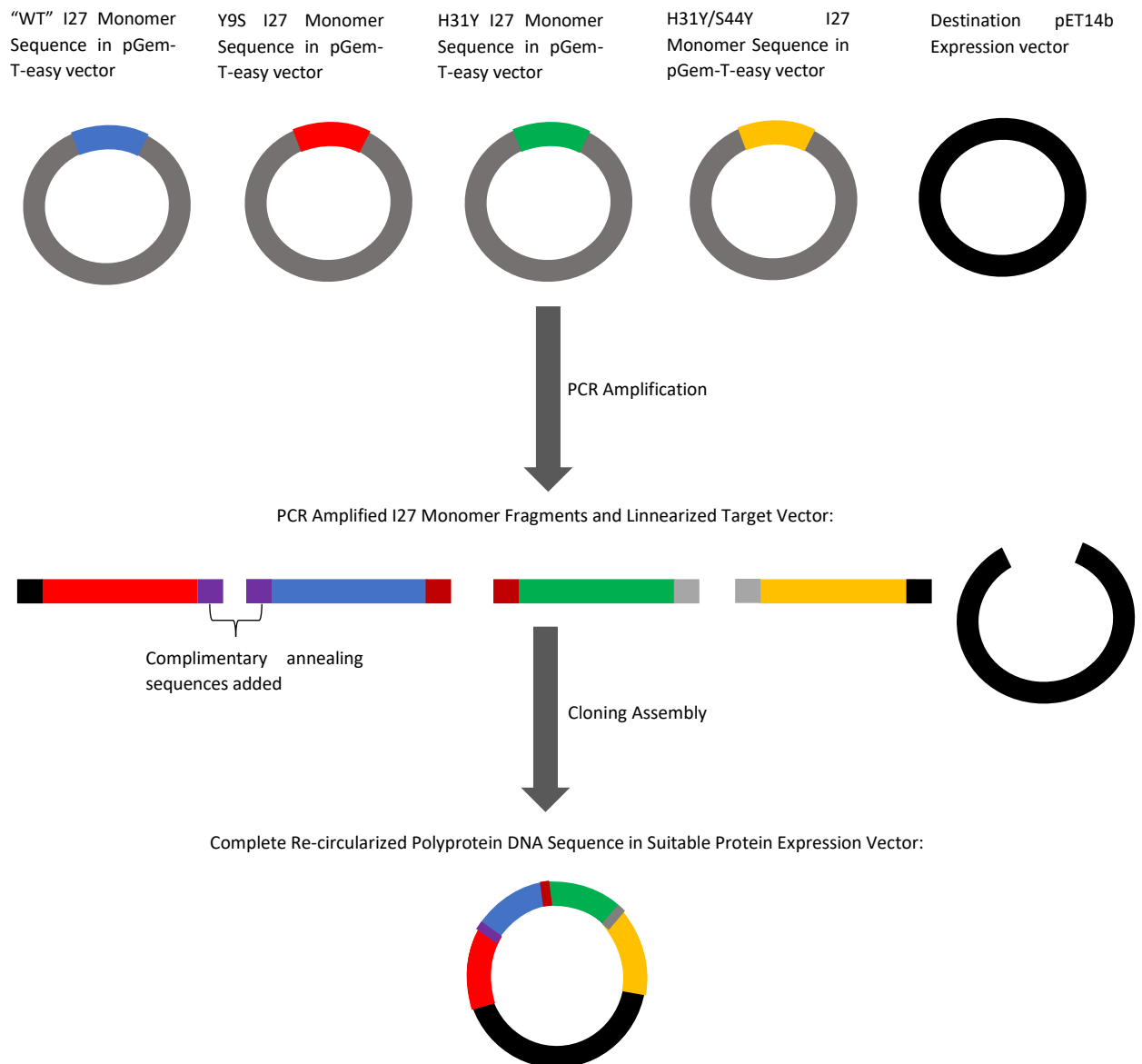


---

Other than these specific mutations all domain sequences are identical, and were generated as isolated monomer sequences in separate pGEM-T-easy vectors. A TEV protease cleavage site sequence was inserted between the hexa-His tag and the start of the I27 sequence. A PCR-based method was then designed by which each individual domain sequence could be PCR amplified out of the vector and a variety of specific flanking sequences added in order to create short double stranded DNA fragments with a central I27 region containing the desired mutations, and any combination of 5' and 3' flanking linker sequences. These linker sequences were initially designed for use in a Gibson Assembly method, but were later adapted for use in Golden Gate assembly. These linkers were designed to allow the order-

specific assembly of multiple monomer DNA fragments into single polyprotein sequences in a suitable expression vector (pET14b). This workflow is summarised in Figure 3.2.

Figure 3.2: Summarized schematic polyprotein cloning workflow for both Gibson and Golden Gate Assembly.

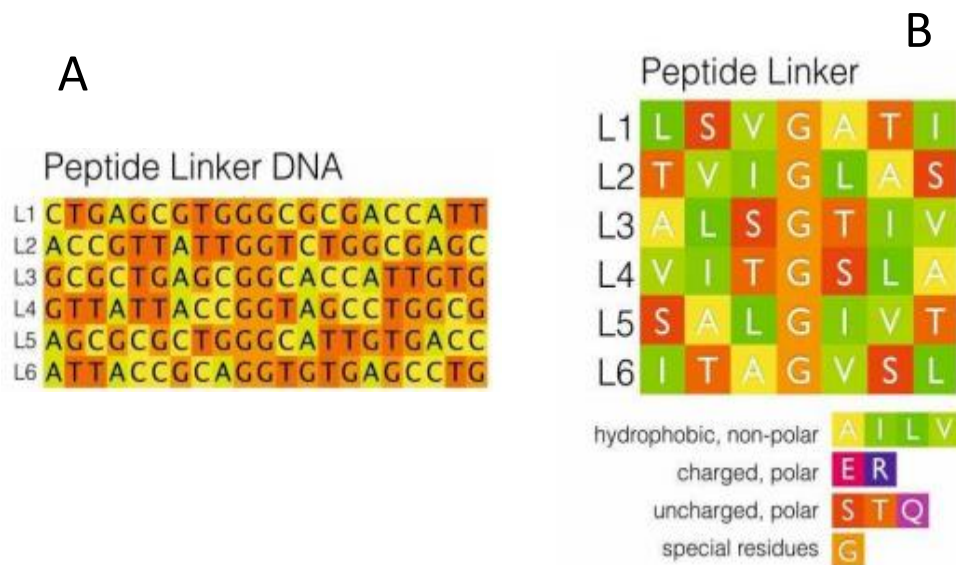


### 3.3.1 Linker Region Design

Earlier work within the lab group developed a protocol for the assembly of up to 7 DNA fragments into a heteropolyprotein gene via Gibson Assembly [182]. These linker regions

were designed around two central principles; first to make them chemically identical at the protein level, and sufficiently diverse at the DNA level to allow specific annealing of linker pairs. A secondary consideration at the protein level was to minimise disruption to the folding of the globular I27 protein domains using intrinsically disordered linker sequences. Towards this end 21 nucleotide (7 residue) linkers were designed as this is an optimal complimentary overhang for Gibson Assembly fragment annealing [220]. A central glycine residue was maintained in each linker, with symmetrical combinations of single residues of serine and threonine (polar), and alanine, isoleucine, leucine, and valine (non-polar) either side. The order of these 6 residues was shuffled and the codon for each altered to create six unique linkers suitable for Gibson assembly. These linkers are summarised in Figure 3.3.

Figure 3.3: A; nucleotide linker sequences designed for Gibson assembly compatibility and unique annealing, B; the amino acid sequences designed for chemical physio-chemical properties [183].



The first four of these linkers were selected for use in the assembly of polyproteins for this study. These are summarised in Table 3.1.

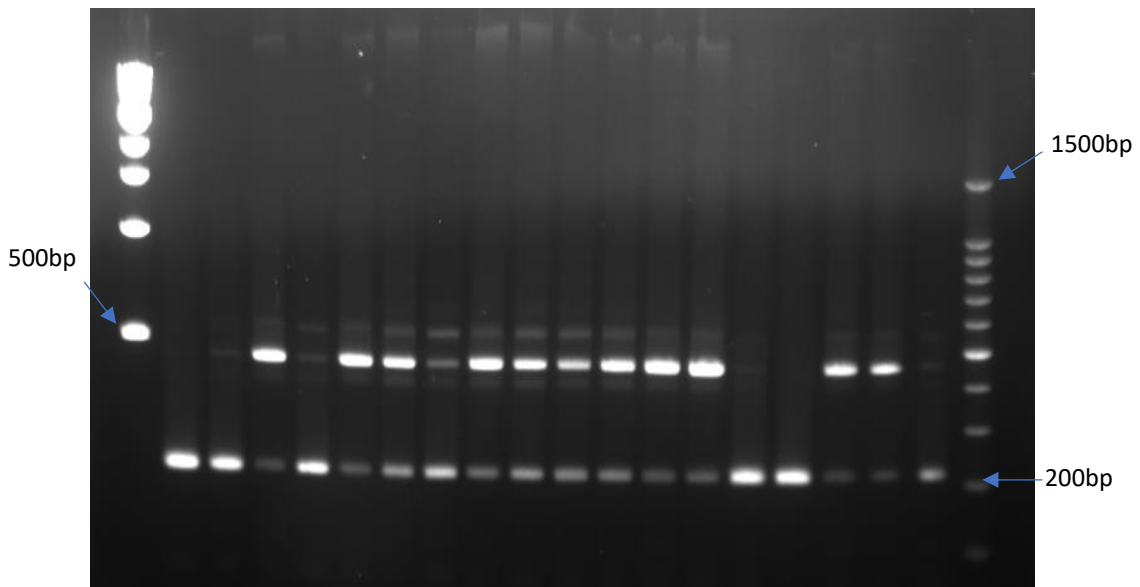
### 3.3.2 Gibson Assembly Method Limitations

Joined Cassettes	Peptide Linker	Amino Acid Sequence
C1>C2	CTGAGCGTGGGCGCGACCATT	L S V G A T I
C2>C3	ACCGTTATTGGTCTGGCGAGC	T V I G L A S
C3>C4	GCGCTGAGCGGCACCATTGTG	A L S G T I V
C4>C5	GTTATTACCGGTAGCCTGGCG	V I T G S L A

Initially replication of the Gibson assembly method used previously was attempted but yielded no full assemblies. The most common assembly product was a single I27 domain correctly annealed at both the 5' and 3' end to the pET14b vector backbone, but lacking any internal linker sequences and expressing only a protein monomer. Despite extensive troubleshooting and effort correct assembly was never achieved reproducibly (Figure 3.4). Ultimately it was concluded that the most likely cause of the continued production of monomer inserts was due to the nature of the exonuclease activity inherent to the Gibson assembly reaction exposing more nucleotides than just the 21 nucleotide linker sequence. In this case the longer identical I27 sequence was exposed leading to preferential annealing of I27 sequences to one another. This caused cassettes 1 and 5 to correctly anneal to the exposed vector backbone but then to each other via their I27 regions instead of cassettes 2 or 4 via their linker regions.

---

Figure 3.4: Electrophoresis agarose gel image of the colony screen PCR products (Section 2.2.1.3.3) generated from *E.coli* transformed with an attempted 5-fragment polyprotein Gibson assembly. A negative result is indicated by an ~200bp band, and a positive full assembly a band at ~1500bp. Bands at ~500bp indicate a single cassette insertion.



---

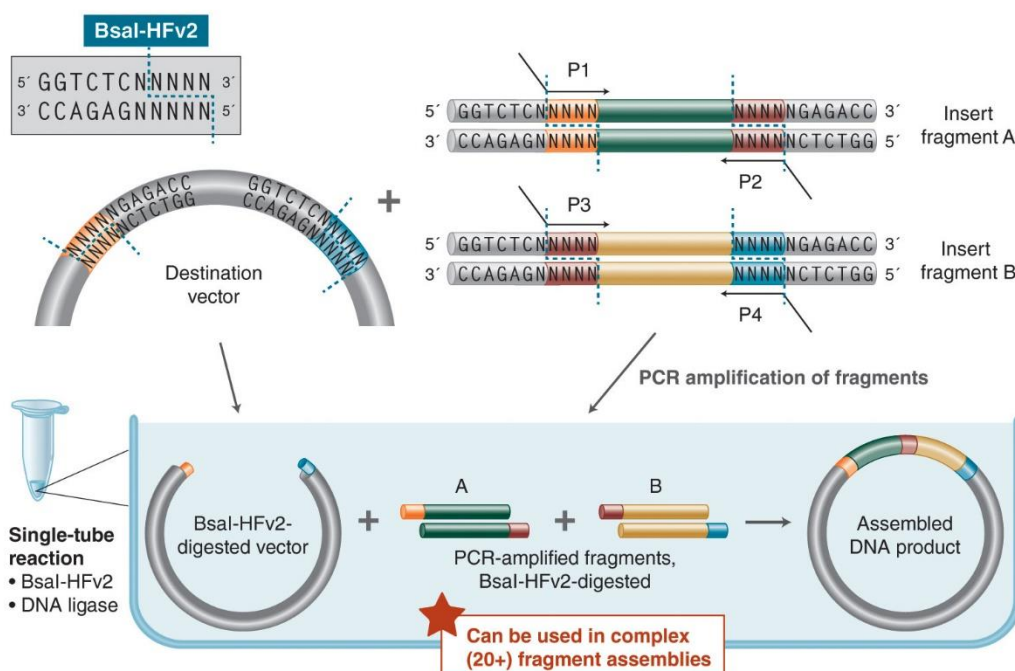
Despite exhaustive attempts to prevent this off-target monomer assembly it was concluded that Gibson assembly is inappropriate for the assembly of homopolyproteins composed of adjacent identical repeating sequences. The alternative but less widely used technique of Golden Gate assembly was therefore adopted instead.

### 3.3.3 Golden Gate Assembly Design

The linker regions previously designed for use in Gibson assembly remain appropriate for use in Golden Gate. However the PCR primers must be redesigned and the target vector extensively modified. Golden Gate is based on type-II restriction endonuclease activity whereby the restriction enzyme binds to a specific sequence but then cuts the DNA at a defined number of bases upstream/downstream of this site in a sequence independent

manner. This allows the generation of any desired sticky end sequence for ligation. The restriction enzyme selected for use in this study was Bsal which binds to the non-palindromic sequence 5'-GGTCTC-3' and cuts downstream at (N)<sub>1</sub> on the coding and (N)<sub>5</sub> on the complementary strand to create any 4 base 5' overhang. In addition to the binding sequence an upstream flanking sequence of any 6 nucleotides is optimal for efficient cleavage. According to these principles new PCR amplification primers were designed to add half of a linker region, the Bsal binding site, and a 6 base flanking sequence onto each end of each I27 cassette. These are described in Table 2.2. The pET14b target vector was modified by PCR mutagenesis (Section 2.2.1.3.1) to remove its two native Bsal binding sites, and insert a bidirectional cut site within the multiple cloning site. Assembly and colony screening was then performed according to the protocols described in Sections 2.2.1.3.2 and 2.2.1.3.3. This process is summarised in Figure 3.5.

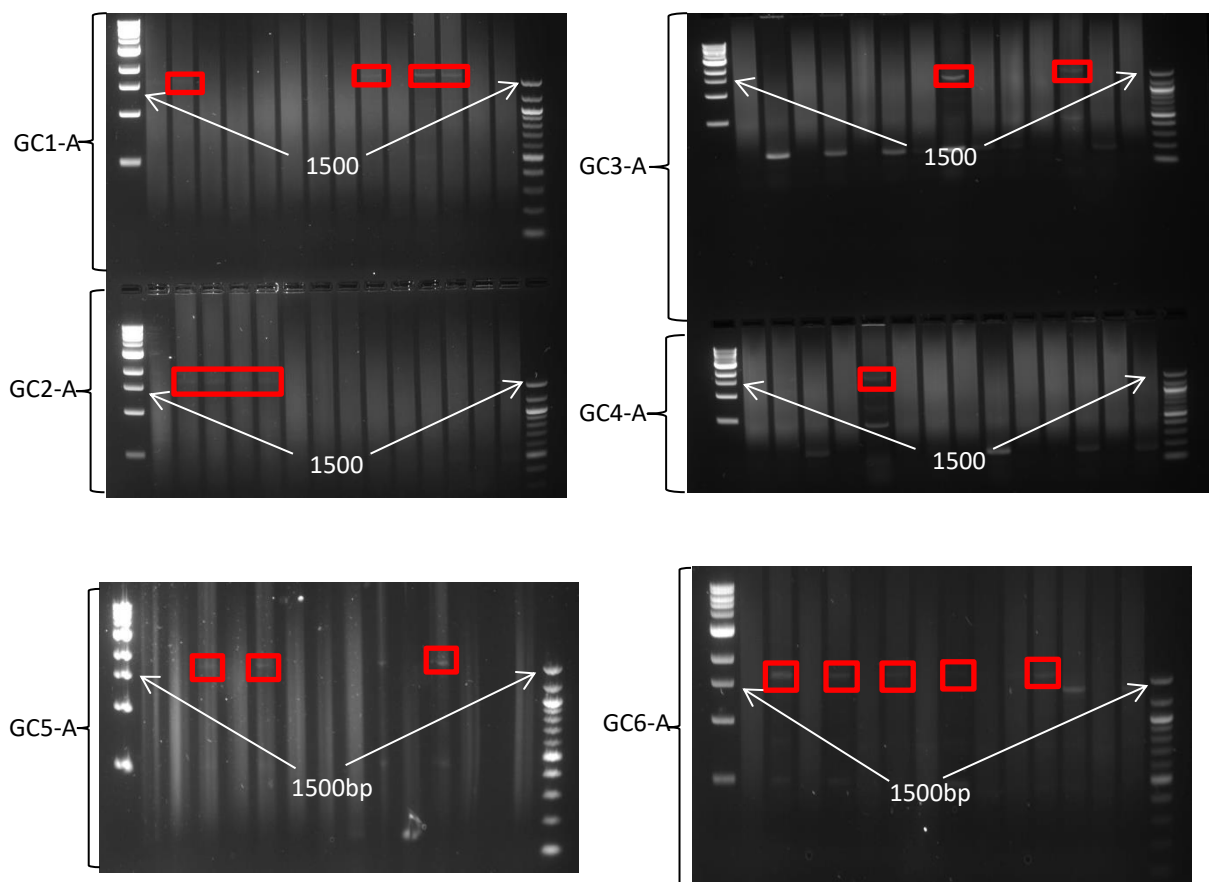
Figure 3.5: Summarised mechanism of Golden Gate Assembly cloning. Diagram reproduced from Genetic Engineering & Biotechnology News Magazine, September 1<sup>st</sup> 2018 (Vol. 38, No 15).





Colony screening and subsequent plasmid DNA sequencing of candidate colonies was used to identify and confirm the correct assembly all polyprotein DNA constructs. Sequencing primers were designed to anneal ~10bp upstream and downstream of the start and stop codons, meaning that a positive result band was expected to be ~1500bp. Successful 5-cassette assembly was calculated to have been achieved with an efficiency of  $26.5 \pm 9.9\%$ . Once a single successful assembly reaction had been performed each subsequent assembly yielded a minimum efficiency of 14.3% (3 correct assemblies from 14 screened colonies). This is summarised in Figure 3.6.

Figure 3.6: Agarose electrophoresis gel images showing the PCR colony screen products of each polyprotein DNA construct. The expected negative band is ~200bp. The expected positive band is ~1500bp. Left side ladder is Bio-Rad 1kb DNA ladder, right side ladder is Bio-Rad 100bp DNA ladder.



Plasmid DNA purified for sequencing was then transformed into BLR DE3 pLysS *E.coli* cells for protein expression (Section 2.2.1). Test expressions of all proteins were then performed to ensure that the vector expression system remained intact.

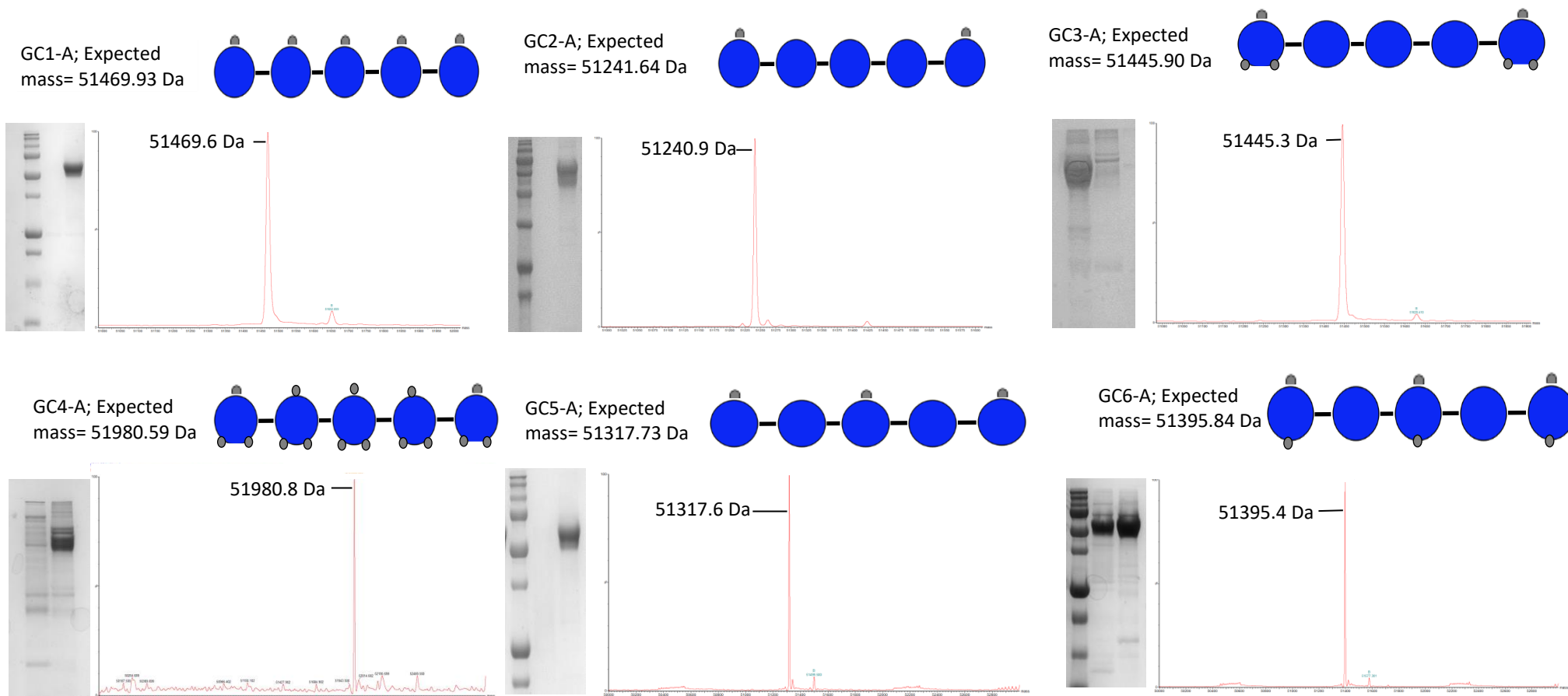
### 3.3.4 Test Expression and Mass Spectrometry Analysis

Each polyprotein construct was expressed, purified, and His-tag removed as described in Section 2.2.3 on a 0.5L scale. All purified samples were then analysed by intact mass spectrometry to ensure that the correct product had been expressed, and subsequently by circular dichroism to ensure they maintained a correctly folded state (Section 2.2.5.5.1).

#### 3.3.4.1 Protein Test Expression and ID Conformation

The tyrosine distribution, final protein purity, and mass spectrometry analysis of each polyprotein is summarised in Figure 3.7. All polyproteins were expressed, purified, and cleaved successfully. Sufficient purity of each was achieved for their use in hydrogel studies.

Figure 3.7. Schematic illustrations showing the number and distribution of tyrosine residues throughout the various polyprotein constructs. In addition the final purified protein as seen on an SDS polyacrylamide gel, and the final sample intact mass spectrum are also shown.



### 3.4.3.2 Circular Dichroism of Novel Constructs

The native secondary structure of each polyprotein was assessed by circular dichroism (Section 2.2.5.5.1). All raw output data (Ellipticity;  $\theta$ ) was converted to Mean Residue Ellipticity (MRE) (Equations 2.2 and 2.3) in order to correct for differences in concentration. This allows direct comparison of spectra to observe differences in peak wavelength and amplitude. Analysis showed that all polyprotein constructs exhibit identical spectra, indicating that they all retain native secondary structure. In addition analysis of each spectra using the DiChroWeb CONTIN and CDSSTR algorithms calculated that the relative composition of each constructs secondary structure is identical [221][222]. The combined CD spectra are summarised in Figure 3.8 and the structural analysis performed using the CDSSTR algorithm on DiChroWeb is presented in Table 3.2 [217], [223].

Figure 3.8: Overlaid CD spectra of all polyprotein constructs designed for hydrogel formation. In addition a reference spectra of (I27)<sub>5</sub> was obtained to compare the novel construct folds against a known folded I27 spectra.

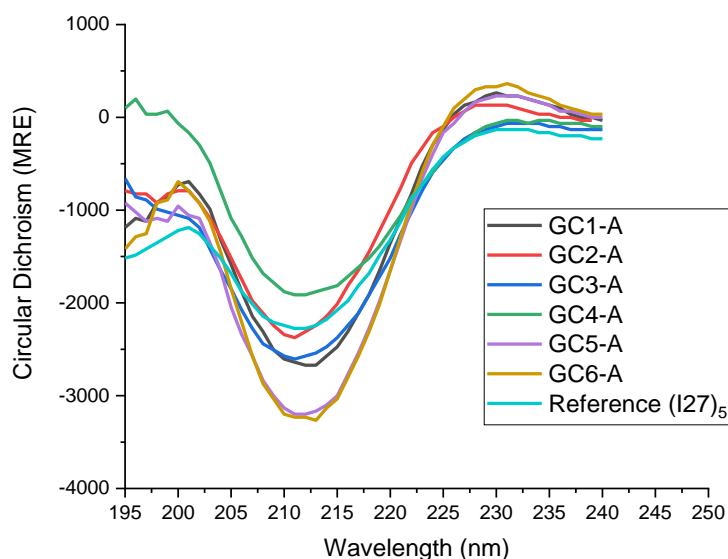


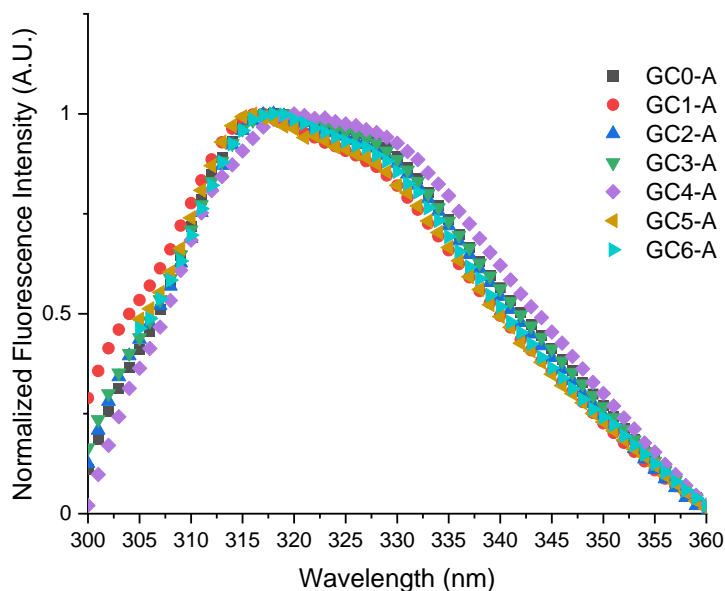
Table 3.2: All constructs secondary structure composition calculated by CDSSTR algorithm [224]. Also included is the Normalised RMSD of each fit and subsequent structure calculation.

Polyprotein Construct	$\alpha$ - Helix 1	$\alpha$ - Helix 2	$\beta$ -Strand 1	B-Strand 2	Turns	Unordered	NRMSD of Fitted Reference Data	Helix segments per 100 residues	Average Helix length per Segment	$\beta$ -strand segments per 100 residues	Average $\beta$ -strand length per segment
GC1-A	0.01	0.05	0.27	0.14	0.22	0.32	0.05	1.20	3.38	6.89	5.91
GC2-A	0.00	0.05	0.24	0.13	0.24	0.34	0.04	1.34	3.73	6.57	5.67
GC3-A	0.00	0.05	0.26	0.13	0.22	0.33	0.04	1.15	3.94	6.45	6.00
GC4-A	0.02	0.03	0.23	0.10	0.18	0.44	0.06	0.68	1.33	5.00	6.62
GC5-A	0.01	0.05	0.28	0.14	0.21	0.31	0.05	1.16	3.28	7.11	5.95
GC6-A	0.01	0.05	0.28	0.15	0.21	0.30	0.07	1.35	3.47	7.33	5.77
(I27) <sub>s</sub> Reference	0.00	0.06	0.23	0.13	0.24	0.34	0.04	1.47	3.68	6.66	5.41
Average	0.01	0.05	0.26	0.13	0.22	0.34	N/A	1.19	3.26	6.57	5.90
stdev	0.01	0.01	0.02	0.01	0.02	0.04	N/A	0.23	0.81	0.70	0.35

All except GC4-A exhibit near-identical spectra indicating that their secondary structure is the same. Differences in amplitude are likely partly due to differences in concentration. GC4-A subsequently failed to form a gel, and exhibited poor thermal and chemical stability, indicating that it may be inherently destabilised by the 10 mutations it contains. This is discussed in detail in Section 4. Tryptophan fluorescence spectra of each construct were also obtained in native conditions (Figure 3.9). Each construct contains 12 tryptophan's (1 per domain).

---

Figure 3.9: Overlaid native fluorescence spectra of all constructs. All raw fluorescence intensities have been normalized to between 1 and 0 to allow comparison.



Once again the spectra shown in Figure 3.9 indicate that all constructs exhibit identical folded fluorescence spectra and therefore have the same degree of tryptophan solvent accessibility and secondary structure. The exception to this once again is GC4-A which displays a clear red shift towards 350nm emission indicating a greater degree of tryptophan solvent accessibility and therefore a higher degree of unfoldedness in native conditions. These results together demonstrate that all seven constructs are correctly folded and exhibit a secondary structure composition identical to that measured from previously published I27 polyprotein constructs. The one potential exception to this is GC4-A which appears to contain a significantly greater unordered fraction than all the others (0.44 versus  $0.32 \pm 0.014$ , Table 3.2). This destabilisation of GC4-A is further discussed in Section 4 in relation to its reduced thermodynamic stability and its inability to form a hydrogel. With the successful assembly, expression, and confirmed

folded state of all pentamers it was the necessary to optimise their expression yield to provide sufficient material for further characterisation and gel formation.

### **3.4 Optimisation of Polyprotein Expression**

#### **3.4.1 Design of Experiments Methodology**

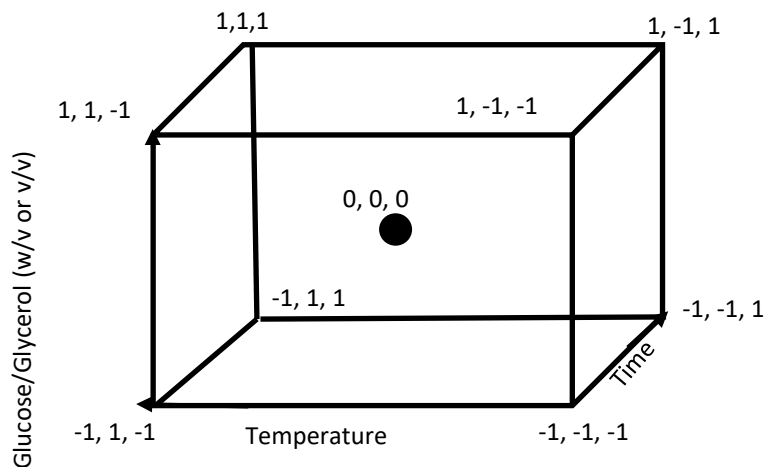
In order to fully optimise the expression of 127 polyproteins a technique commonly used in an industrial setting called Design of Experiments (DoE) was utilised [225]. DoE is based upon principles of statistical analysis to extrapolate a detailed 3-dimensional model of the relationship between multiple independent variables and multiple measured outputs of a process. This is achieved by obtaining a carefully selected subset of 2-dimensional real-world experimental data, and predicting the likely intermediate data points to predict the optimal conditions to maximise any output from the process without the need to perform every conceivable experiment and obtain the real-world data. Experimental design is carried out based on three criteria; the input factors, the output factors, and the distribution of experimental extremes to construct a statistically robust 3D model [226]. The input variables and measured outputs used in this study are summarised in Table 3.3. The distribution of experimental extremes refers to the extreme values of each input variable, namely high (arbitrary value 1), medium (arbitrary value 0), and low (arbitrary value -1), within an imaginary 3D cube in order to pin the landscape sheet to each corner and to intermediate points within the cube. This produces a contoured landscape which is computationally fitted to a statistical model (e.g. 2 factor interaction, quadratic, or linear) in order to assess the effect of all/any input variables on all/any output variables. Transformation of the data is sometimes performed in order to obtain a better statistical fit, whereby the input data is transformed,

but the output predicted data is not, meaning that a better model is produced, but the predicted output values are expressed to the same order of magnitude as the original data.

Table 3.3: DoE input variables and output measurements	
Input Variables:	Output Measurements:
Temperature of Incubation: - 19°C (-1) - 28°C (0) - 37°C (1)	Dry cell culture pellet biomass
Time of Incubation (before induction where appropriate): - 24 hours (-1) - 36 hours (0) - 48 hours (1)	OD <sub>600</sub> of culture
% Glucose/Glycerol (w/v or v/v) - 0.2% (-1) - 0.6% (0) - 1% (1)	Relative Polyprotein Expression

An industrially standard distribution of input values was used to produce a statistically strong landscape. This pattern of inputs is summarised in Figure 3.10.

Figure 3.10: Distribution of test expression condition values within the 3-dimensional expression landscape cube.





This distribution of experimental conditions yields a 3D landscape containing sufficient data points for robust analysis of variance and identification of interactive factors [226][225]. To improve this analysis further data points were added for this study using mid-points for each variable. The full set of data point distributions is summarised in Table 3.4.

Table 3.4: Experiment class designations with corresponding variable values			
Experiment Class Designation	Factor		
	1: Glucose % (w/v) (2YT medium) / Glycerol % (v/v) (TB medium)	2: Time (Hours)	3: Temperature (°C)
1, 1, 1	1	48	37
1, 1, -1	1	48	19
1, -1, 1	1	24	37
1, -1, -1	1	24	19
-1, 1, 1	0.2	48	37
-1, 1, -1	0.2	48	19
-1, -1, 1	0.2	24	37
-1, -1, -1	0.2	24	19
0, 0, 0	0.6	36	28
0, 0, -1	0.6	36	19
0, 0, 1	0.6	36	37
0, -1, 0	0.6	24	28
0, 1, 0	0.6	48	28
-1, -1, 0	0.2	24	28
1, 1, 0	1	48	28

A 3D landscape is generated by plotting output data and fitting it to a predefined interactive model, with the accuracy of the model being gauged by a p-value derived by assessing how far each data point differs from a “perfect” statistical model. The interactions between the input factors is then assessed by an F-value test to determine if the null distribution is affected

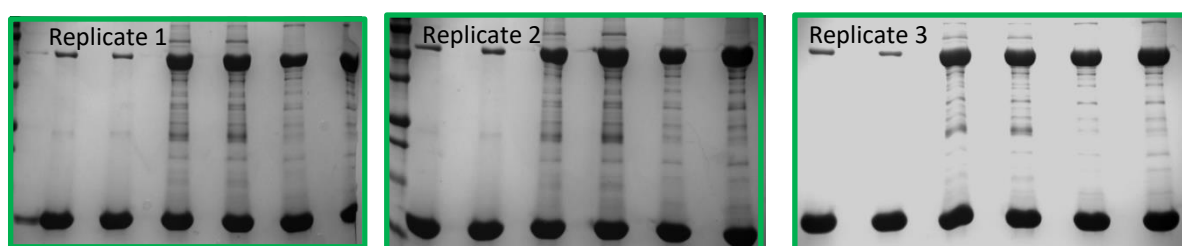
by the changing of one or more input factors. This method allows the identification of input factors which affect the desired outputs, and identifies potential design spaces for further optimisation. This process can be used iteratively with each round identifying a smaller region of design space, leading to greater degrees of optimisation. By carefully designing experiments using Design of Experiment principles, datasets for 2YT, TB, and autoinduction medium were generated and fitted to 3-dimensional models to analyse the interactive factors which affect the outputs:  $OD_{600}$ , cell biomass, and relative protein expression.

Each output variable (biomass,  $OD_{600}$ , protein expression) is analysed separately, in order to identify the optimal conditions to maximise each variable. Each output is plotted against two input variables on the X and Z axes, and then positioned on the Y-axis by the value of the third variable (1, 0, or -1). The landscape can then be moved through all the values on the Y-axis not limited to those for which real data exists, allowing the prediction of conditions of greater output. For example, the greatest region of output as determined by the real data may be 28°C for 24 hours with 1% glycerol, but then by varying the theoretical temperature to 23°C the model may predict that this is a better temperature than 28°C and will estimate an output. The p-value by which the statistical strength of the model is measured is the average distance each real data point sits from the best fit line/curve for the data. A high p-value therefore indicates that the data does not conform well to the best fit the model can achieve, whilst a low p-value indicates that the real data sits well on the best fit model, making it more likely that the predicted intermediate data points are correct. This p-value is not as discerning as a standard T-test, as p-values  $>0.05$  do not necessarily render the model useless, merely more requisite of rationalisation. All test condition experiments and measurements were performed according to the protocols described in Section 2.2.2.

### 3.4.2 Validation of Repeatability of (I27)<sub>5</sub> Expression Measurement

(I27)<sub>5</sub> protein expression was measured by densitometry as a ratio between the (I27)<sub>5</sub> band pixel density divided by the pixel density area of a  $\beta_2$ M reference band in an SDS polyacrylamide gel (Section 2.2.2.4). This process required the manual definition of the band areas and so it was necessary to assess the reproducibility of this process to ensure that all data was directly comparable. This was done by preparing and running 6 samples 3 times each on 3 separate SDS-PAGE gels and measuring each individually, with the aim being to observe minimal variation in peak expression ratio between the same samples measured on different gels. This allowed the observation of any variance in expression ratio result caused by human error in sample preparation or image processing. It was determined that variance was extremely low and that the manual image processing method was highly robust. Therefore we can be confident that the values measured from densitometry analysis are reliable and not significantly variable subject to human error. Results are summarised in Figure 3.11.

Figure 3.11: Summary of variance of protein expression ratio measured from multiple gels.



Lane	Ratio replicate 1	Ratio replicate 2	Ratio replicate 3	Average	Standard Deviation
1	0.154	0.216	0.142	0.171	0.04
2	0.111	0.151	0.087	0.116	0.03
3	0.564	0.515	0.691	0.590	0.06
4	0.692	0.779	0.670	0.714	0.06
5	0.481	0.544	0.608	0.544	0.06
6	N/A	0.958	0.725	0.841	0.16

The average standard deviation across these three gels for these six samples is 0.07. This demonstrates the robustness of the peak area designation protocol, and that protein expression ratios can be reliably interpreted as comparable measures of protein expression. Having concluded that the data obtained for protein expression was robust it was used to construct DoE cubes.

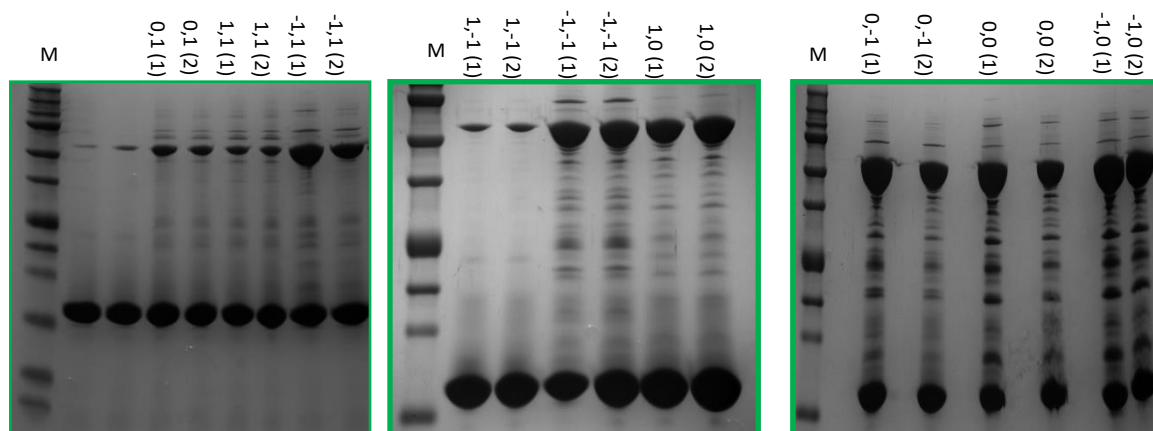
### 3.4.3 Polyprotein Expression Response Measurement

The wet cell pellet biomass, culture OD600, and relative expression levels of polyprotein were measured according to the protocols described in Section 2.2.2.3 and 2.2.2.4. The raw data and SDS gel images of protein expression is presented in Figures 3.12-3.16. Where standard deviations are provided the condition has been performed in triplicate, where no error is provided only the average of duplicate measurements has been reported.

#### 3.4.3.1 Autoinduction Medium

Example SDS-PAGE gel images of autoinduction medium trials are presented in Figure 3.12. Top digits refer to experiment class designation (Table 3.4), and in brackets replicate number. Table 3.5 summarises all response measurements averaged across three independent repeat expressions.

Figure 3.12: Autoinduction Raw Response Data and protein expression SDS gel images for replicates 1 and 2.



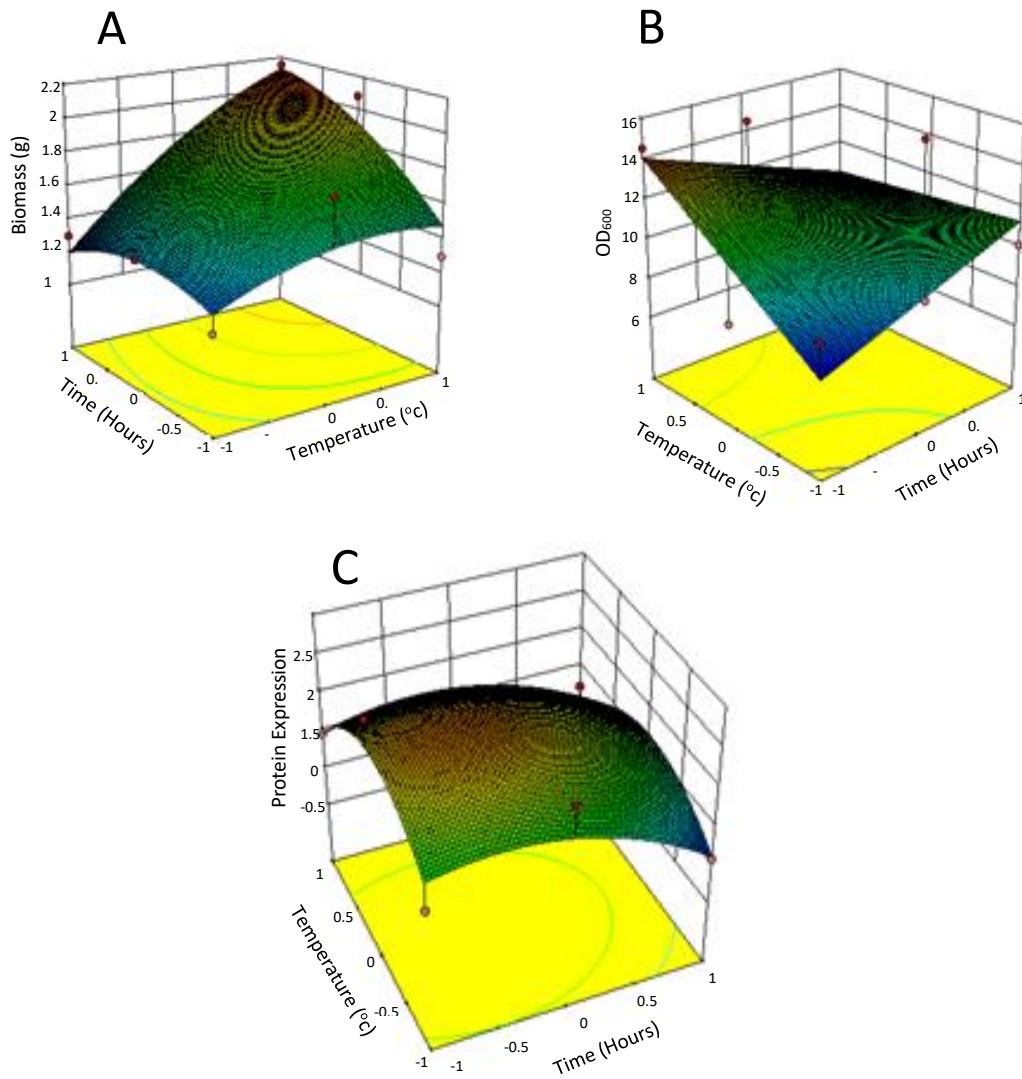
The raw data presented in Figure 3.12 and Table 3.5 immediately illustrates the significant variation in polyprotein expression resulting from minor changes in growth conditions. The data summarised in Table 3.5 was then used to perform DoE statistical analysis to identify the interacting factors and predict the optimal expression conditions. Example resultant DoE output prediction cubes are shown in Figure 3.13. Figure 3.13A predicts biomass output using a linear model with no data transform. Model P-value= 0.0583, F-value= 4.74. The highest predicted area of biomass concurs with that observed experimentally; 19°C for 48 hours.

Experiment Class	Response		
	1: Wet Cell Biomass (grams)	2: OD600 (A.U.)	3: Protein Expression relative to 5mg/mL $\beta_2$ M reference
1,1	1.2±0.1	9.15±0.05	0.221±0.014
1, -1	2.25±0.15	14.51±0.01	0.089±0.004
-1, 1	1.15±0.05	10.05±0.35	0.475±0.120
-1, -1	1.25±0.05	7.95±0.55	0.533±0.042
0,0	1.6±0.1	10.25±0.35	1.316±0.281
0,-1	2.05±0.05	14.45±0.15	1.294±0.319
0,1	1.45±0.15	9.15±0.05	0.272±0.018
-1,0	1.85±0.15	13.70±0.3	1.704±0.190
1,0	1.45±0.15	7.80±0.5	0.403±0.022

Figure 3.13B predicts OD<sub>600</sub> output. The model used is linear with no data transform. Model p-value= 0.3724, F-value= 1.30. Once again the highest predicted area for OD600 is 19°C for 48 hours. However this contradicts the experimental data which indicates that 36 hours is significantly superior to 48. This highlights the limitations of the model as the data does not conform well to any available model, hence the high p-value. Figure 3.13C predicts protein expression output. The model used is quadratic with no data transform. Model p-value= 0.3449, F-value= 1.73. A large design space for optimal protein expression is centred around 24-28°C and 24-30 hours. Whilst this model is statistically weak it does suggest a design area which could be rationally justified for further enquiry.

---

Figure 3.13: A; Autoinduction DoE cube predicting biomass output. B; Autoinduction DoE cube predicting OD<sub>600</sub> output. C; Autoinduction DoE cube predicting protein expression output.

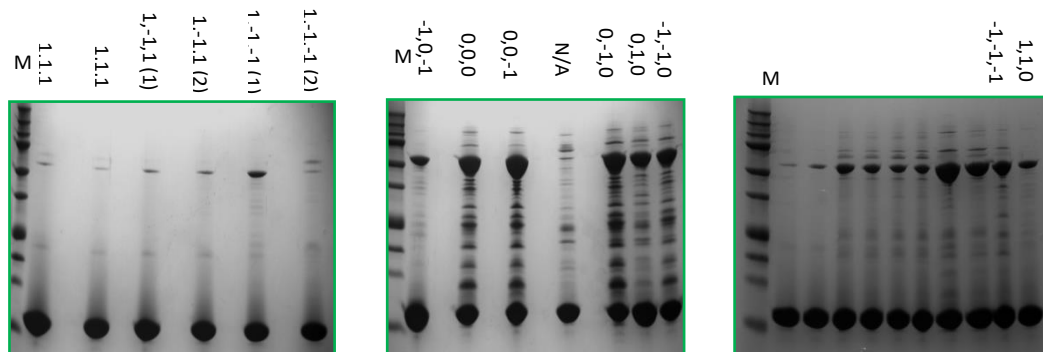


### 3.4.3.2 Terrific Broth Medium

During investigation of terrific broth medium three input variables were changed; Glycerol %, time of incubation, and temperature of incubation. In Figure 3.14 top digits refer to experiment class designation (Table 3.4) with the replicate number in brackets. Only one example of each experiment class is shown in the SDS-PAGE gel images displayed in Figure 3.14. Table 3.6 summarises the three output variable results with standard deviation shown when performed in triplicate with all others performed in duplicate.

---

Figure 3.14: Terrific Broth Media (TB) SDS-PAGE gels for Densitometric Analysis.



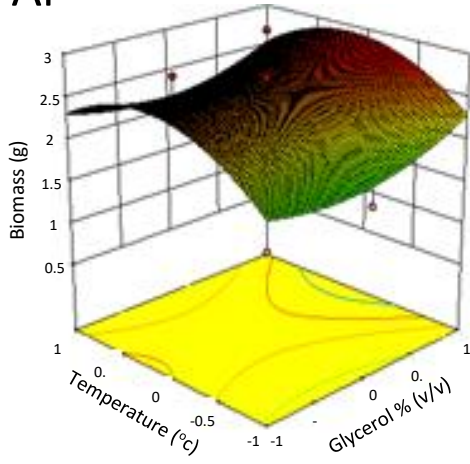


Experiment Class	Response		
	Wet Cell Biomass (g)	OD600 (A.U)	Polyprotein Expression ratio
1,1,1	1.9±0.3	9.1±1.85	0.031±0.002
1,1,-1	0.9	1.8	0.048
1,-1,1	2	2.5	0.057
1,-1,-1	0.7	1.6	0.184
-1,1,1	1.1	4.7	0.034
-1,1,-1	1.3	1.8	0.238±0.014
-1,-1,1	1.4	5.5	0.093±0.001
-1,-1,-1	0.8	1.5	0.239±0.036
0,0,0	2.7	14.6	0.712±0.034
0,0,-1	0.9	1.6	0.826
0,-1,0	1.6	2.2	0.678±0.111
0,1,0	2.4	13.2	0.463±0.011
-1,-1,0	1.5	2.2	0.447±0.095
1,1,0	2.6	16.1	0.116

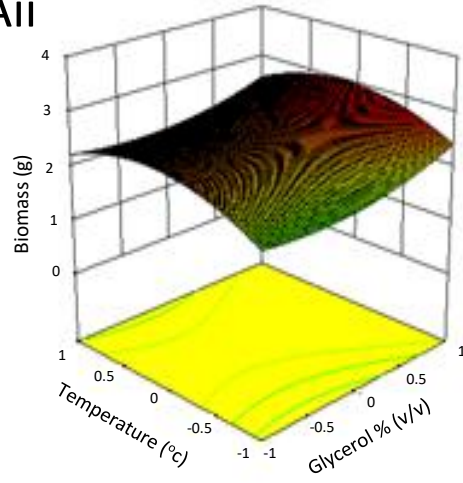
DoE output prediction cubes for terrific broth medium are shown in Figure 3.15. Figures 3.15Ai and Aii predict biomass output. The model used is Quadratic with no data transform. The model p-value= 0.0333, and the F-value= 5.83. (Ai) is built from “real” data, and predicts the area of highest biomass output to be 1% Glycerol, 28°C, and ~42 hours. (Aii) is a purely theoretical landscape at a temperature of 0.3 (~30.7°C), and predicts that greater biomass output can be achieved at this temperature, of ~3g. The highest biomass data point obtained was 2.7g at 0.6% Glycerol, 28°C, 36 hours. Figure 3.15B predicts OD<sub>600</sub> output. The model is Linear with a p-value= 0.0778, and F-value= 3.07. No improvements upon trialled conditions are suggested by the model. Figure 3.15C predicts protein expression output. The model is quadratic with a Log transform of the input data. Model p-value= 0.0206, F-value= 9.92. The greatest area of output is predicted at 0.6% Glycerol, 28-32.5°C, for ~30 hours.

Figure 3.15: DoE cubes for Terrific Broth Medium.

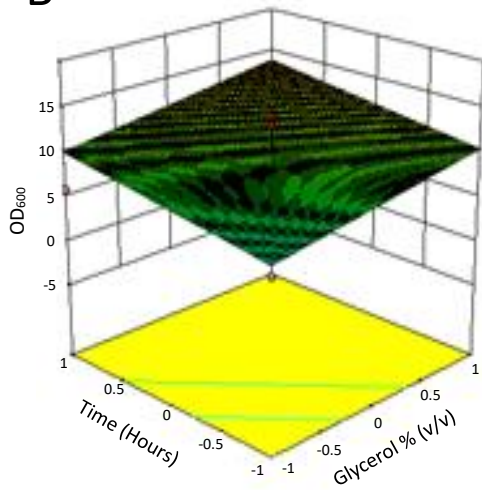
Ai



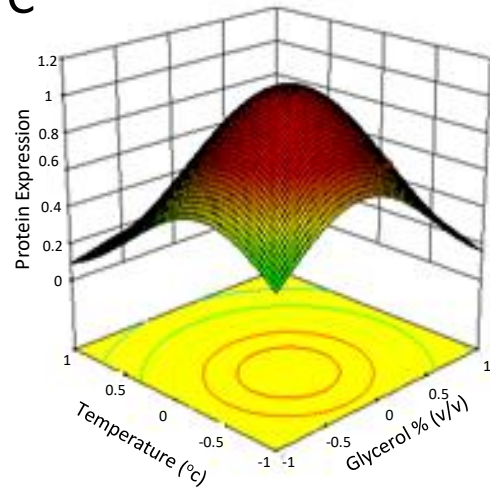
Aii



B



C



### 3.4.3.3 2YT Medium

During investigation of 2YT medium three input variables were changed; Glucose %, time of incubation, and temperature of incubation. In Figure 3.16 top digits refer to experiment class designation (Table 3.4). Only one example of each experiment class is shown in the SDS-PAGE gel images displayed in Figure 3.16. Table 3.7 summarises the three output variable results. No standard deviation is shown because all experiments were performed only in duplicate.

Figure 3.16: 2YT Medium SDS-PAGE gels for densitometric analysis.

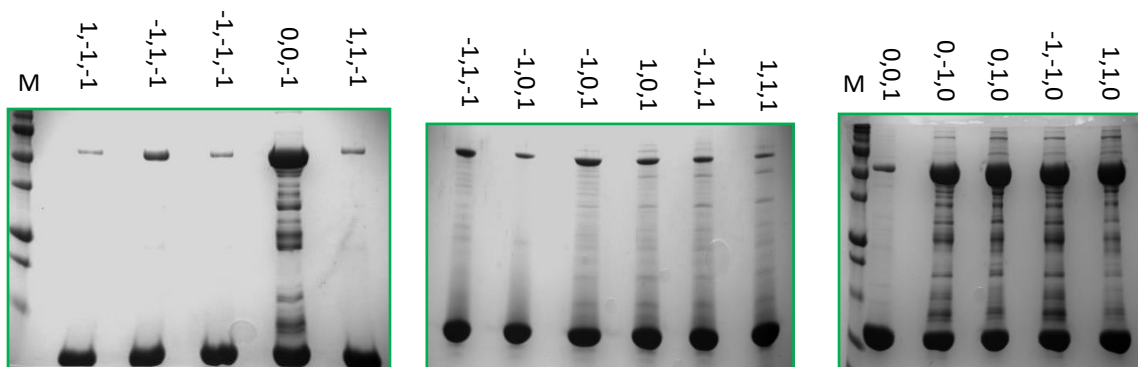
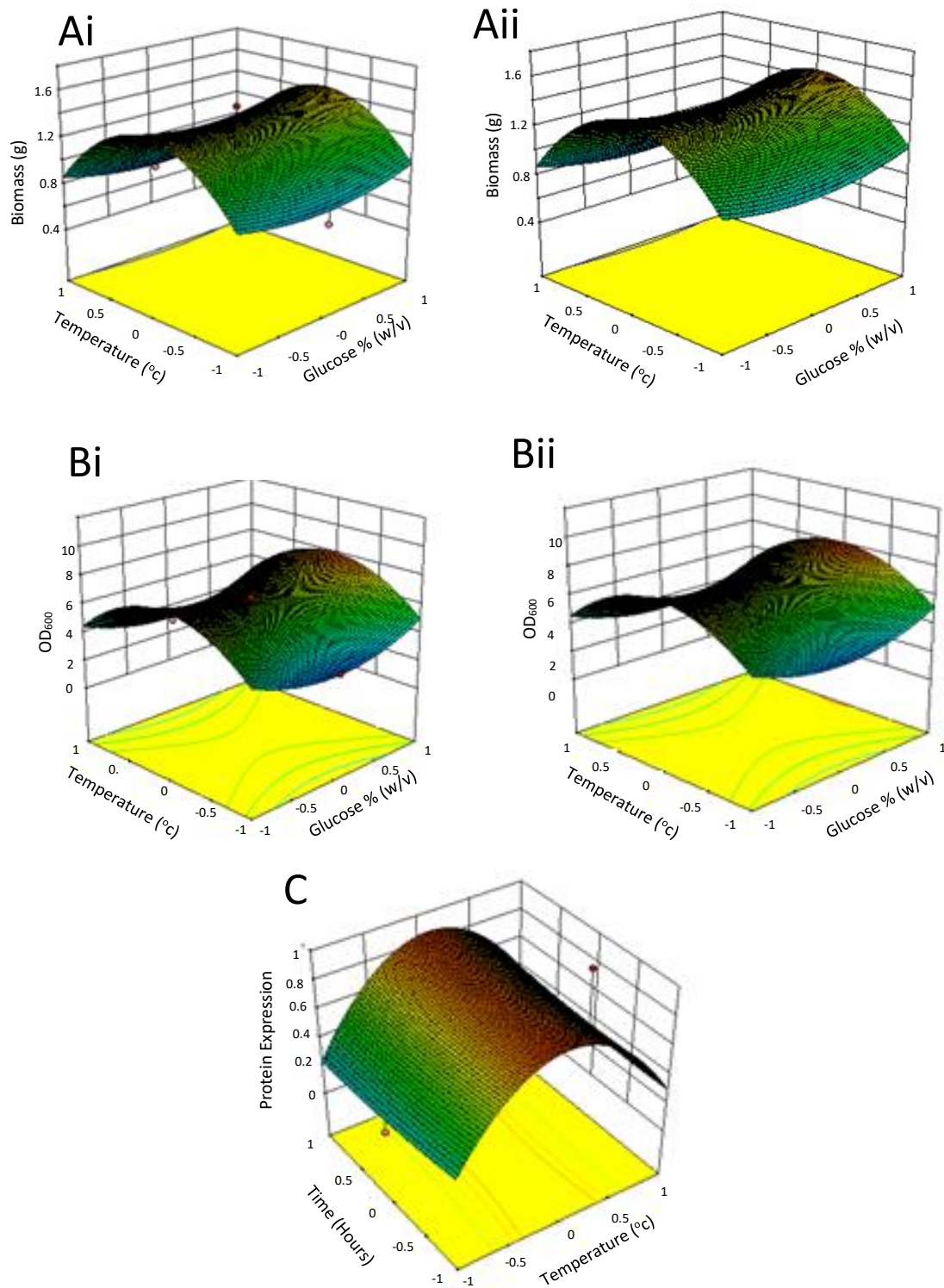


Table 3.7: Summarised terrific broth medium output results			
Experiment Class	Response		
	Wet Cell Biomass (g)	OD600 (A.U)	Polyprotein Expression ratio
1,1,1	0.63	3.39	0.087
1,1,-1	0.79	3.75	0.086
1,-1,1	0.63	3.31	0.157
1,-1,-1	0.75	2.66	0.039
-1,1,1	0.63	3.50	0.143
-1,1,-1	0.92	4.09	0.186
-1,-1,1	0.87	3.24	0.243
-1,-1,-1	0.72	2.76	0.049
0,0,-1	0.71	3.28	0.745
0,0,1	0.71	2.89	0.167
0,-1,0	0.67	2.81	0.908
0,1,0	1.63	6.63	0.825
-1,-1,0	0.80	3.69	0.915
1,1,0	1.56	8.41	0.927

DoE output prediction cubes for terrific broth medium are shown in Figure 3.16. Figures 3.16Ai and Aii predict biomass output at time= 0 and time= 0.4 respectively. Figure 3.16Ai predicts the area of highest biomass output to be 1% Glucose (w/v), 28°C, and ~36 hours. Figure 3.16Aii is a theoretical landscape and suggests that greater output can be achieved at time 0.4 (~42 hours) with a predicted biomass of ~1.55g. Figures 3.16Bi and Bii predict OD<sub>600</sub> output. Figure 3.16Bi predicts the area of highest OD<sub>600</sub> output to be 1% Glucose (w/v), ~30°C, and 36 hours. Figure 3.16Bii is a theoretical landscape and suggests that greater output can be achieved at time 0.24 (~41 hours) with a predicted OD<sub>600</sub> ~8. The highest recorded OD<sub>600</sub> was 8.41 (1,1,0). Figure 3.16C predicts protein expression output. The model used is Quadratic with no data transform. The model p-value= 0.2732, and F-value= 1.78. Figure 3.15C predicts the area of highest protein expression output to be 0.6% Glucose (w/v), at 28°C. Interestingly it appears that time does not affect protein expression, indicating that optimal expression may be lower than 24 hours.

Figure 3.16: DoE Cubes for 2YT medium; Ai real landscape predicting biomass output, Aii theoretical landscape predicting biomass output at time= 0.4. Bi real landscape predicting OD<sub>600</sub> output, Bii theoretical landscape predicting OD<sub>600</sub> output at time= 0.24. C protein expression output landscape.



#### 3.4.4 Design of Experiment Analysis

An initial interpretation of the raw trial data indicates that 2YT is inherently inferior to TB and autoinduction as no conditions in this medium rank at the top of any output. TB 0,0,0 exhibits the highest biomass and OD<sub>600</sub> (2.72g, 14.6), but a protein expression ratio of only 0.68, almost 3 times lower than that of autoinduction -1,0 (expression ratio 1.7), which yields the highest protein expression. Furthermore, all of the top three expression ratios are in autoinduction medium. DoE analysis of each output individually indicates that the optimum conditions for all three outputs are rarely the same, with higher OD<sub>600</sub> and biomass generally correlating better with a lower growth temperature than protein expression, and a higher glycerol/glucose percentage. This indicates that high OD<sub>600</sub> and biomass do not necessarily correspond to higher protein expression, a theory further confirmed by the best 2YT expression condition (-1,-1,0); biomass is low (0.801g) as is OD<sub>600</sub> (3.69) but the protein expression is the second highest achieved in 2YT medium (0.915), indicating that the efficiency of protein production per cell is far higher than other conditions. As a result protein expression becomes the only output of interest in further optimisation.

DoE statistical analysis indicates that in all three media all three input variables effect protein expression, as shown by F-values in excess of 1 for all protein expression cubes, and landscape alterations upon the changing of any single variable. One notable exception to this rule is protein expression in 2YT, which appears to be unaffected by time, which may indicate expression ceases after a time period below 24 hours. These three-way relationships of conditions offer an explanation as to why some of the models are weaker than others; if only one factor affects the output, the model is free to fit a line optimally through this one set of variables. By adding two or more interacting factors the degrees of freedom with which the

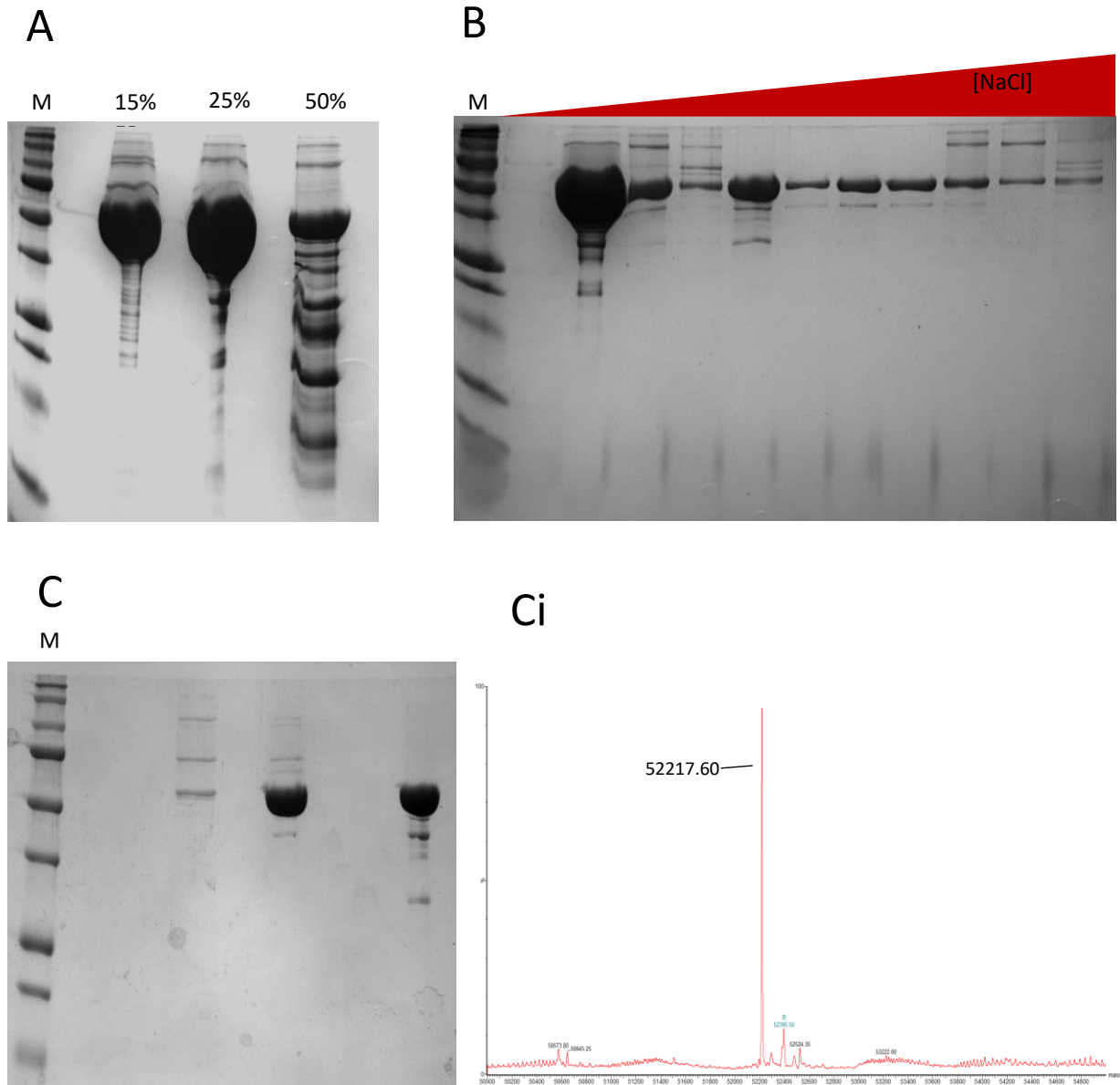
program can fit each set of variables is reduced as they interact with each other and are moved off their best fit line. Furthermore each data set may adhere to a different model e.g. time may be linear, temperature may be quadratic. This further reduces the strength of the model as it tries to reconcile these two datasets to one model.

Despite a weak protein expression DoE model, it is possible to conclude from the raw data that autoinduction medium is the optimal medium for expression of (I27)<sub>5</sub>, and that the optimal conditions are 28°C for 24 hours. The DoE model also indicates that a temperature between 19 and 28°C may improve expression. The model also shows that expression is highest at low time independent of all other variables, suggesting that a shorter time may improve expression.

### 3.4.5 Large-Scale Protein Expression and Yield

A large-scale test culture was performed using three technical repeats of 0.5L of autoinduction medium using the optimised conditions identified in Section 3.4.3.1 (28°C for 24 hours). Each growth culture was purified separately according to the protocol described in Section 2.2.3. Final protein yield was calculated for each culture after the final size exclusion chromatography purification step prior to dialysis and lyophilisation (Section 2.2.3.4). The average yield of protein obtained was calculated by absorbance at 280nm (Section 2.2.4.1) in conjunction with the final purified sample volume. The average protein mass yield was 125±25mg/L of bacterial culture. The purification process is summarised in Figure 3.17. The expected mass of the (I27)<sub>5</sub> polyprotein is 52219.15 Da.

Figure 3.17: A; Nickel-ion exchange Chromatography peak fractions visualized in an SDS-PAGE gel. The 15% elution buffer and 25% elution buffer peaks were retained for subsequent purification stages, B; Anion Exchange Chromatography peak fractions visualized in an SDS-PAGE gel collected from 0-99% 1M NaCl elution gradient. The fractions in lanes 2,3, and 5 were pooled and retained for final stage purification, C; Size exclusion chromatography peak fractions (1mL/min flow rate) visualized in an SDS-PAGE gel. The fractions in lanes 5 and 8 were pooled, analysed by mass spectrometry (Figure 3.18Ci), and lyophilised.



Prior to lyophilisation a protein yield of  $125 \pm 25$  mg/L of culture was calculated. In addition the final purity of the sample as assessed by mass spectrometry was estimated to be consistently >90% as the dominant mass peak of 52217.6Da (expected mass 52219.15Da) was significantly



larger than any other peak above the noise of the spectrum. This purity was judged to be sufficient for use in the experimental investigation of hydrogels, and the purification workflow fast enough to make purification of sufficient protein mass for this study to be feasible. In addition, the mass yield obtained was calculated to be sufficient that 10L of purified growth culture would be sufficient to provide ample protein for the full investigation of a single polyprotein construct, rendering further DoE optimisation unnecessary.

### **3.5 Discussion**

During the planning stages for this study it was obvious that two significant technical hurdles had first to be overcome before any hydrogel research could begin. The first of these technical hurdles was the reproducible and efficient production of the mutant I27 pentamer DNA constructs designed for hydrogel crosslink density modulation. The first method of streamlined multiple fragment cloning investigated was Gibson assembly. This method proved unsuitable for the assembly of homopolyproteins presumably due to difficulties in precisely controlling the length of the 3' overhangs generated by exonuclease activity exposing the identical I27 DNA sequences. This resulted in off-target annealing outside the unique linker regions designed for the purpose. Subsequently Golden Gate Assembly was adopted and was optimised to produce a 5 fragment assembly efficiency of  $23.5\pm 9.1\%$ . This allowed the rapid and efficient assembly of all the polyprotein constructs designed for this study. The robust expression of these novel mutant I27 polyproteins was then confirmed, and their folded state and secondary structure relative to the historical reference construct measured. This showed that all constructs were robustly expressed, and that all exhibited

approximately identical secondary structure to the historical reference, with the exception of GC4-A.

The second major technical hurdle was the production and purification of sufficient recombinant protein mass to make the characterisation of multiple hydrogel constructs feasible. This bottleneck was solved by the rational use of the statistical Design of Experiments approach. This approach allowed minimal experiments to be carried out for maximum gain in understanding in the expression landscape of the protein. The optimisation test culture methods used illustrated immediately the variation and relative increase in expression it was possible to achieve by altering growth conditions. Once analysed these results indicated the interacting factors affecting protein expression, and allowed clear visualisation of the optimal conditions for recombinant protein yield. This was revealed to be autoinduction medium with protein expression favouring short periods of incubation at medium temperatures, not correlating to the longer incubation times at higher temperatures favoured by biomass and OD<sub>600</sub> optimisation. The optimisation test cultures were performed at 10% of the scale envisaged for large-scale expression, and were grown in conditions as perfectly scaled down as possible; 50mL culture in 200mL baffled flasks versus 500mL culture in 2L baffled flasks, etc. As a result large-scale test grows yielded relative increases in protein yield commensurate with those predicted by the optimisation expressions. The final result of this was the increase in I27 polyprotein yield from 10-15mg/L of culture to 100-150mg/L of culture when using autoinduction medium grown at 28°C for 24 hours [227], [228]. This increase in protein yield underpinned the feasibility of all subsequent work.

These two optimisation and validation processes allowed the rapid and efficient production of multiple novel polyproteins. This availability of design space and recombinant protein raw

material allowed the subsequent biochemical/biophysical characterisation of each construct and the hydrogels they formed to be assessed. The work described in this thesis would not have been possible without success in these first stages.

## Section 4: Thermodynamic Characterization of I27 Polyproteins

### 4.1 Introduction

Proteins are the product of billions of years of evolution. Each and every one has evolved over time to perform a specific task, and integral to this is the primary amino acid sequence. This sequence alone is sufficient to allow a protein to fold and gain secondary structure, with the laws of evolution dictating that once a sequence evolves which programs appropriate function little further evolution will occur [229], [230]. The structure, stability, and function of proteins is therefore highly sensitive to changes in their primary amino acid sequence. Single residue mutations can destabilize the folded state of a protein rendering it unable to fold or become more highly sensitive to environmental insult [231]. The nature of this study requires the rational substitution of residues in I27 in order to remove or insert tyrosine residues as a strategy for controlling the crosslink density of folded protein hydrogels. It is therefore necessary to measure any alterations in thermodynamic stability that these mutations may cause in order to correlate them with alterations in hydrogel mechanics. This was achieved by investigation of both the thermal and chemical denaturation behavior of each pentamer construct to extract their thermodynamic properties.

All proteins can exist in a variety of states across a range of environments. These states range from the highly ordered states of most functional globular proteins, pass through a series of intermediate folds which may have altered/reduced activity, until finally all structure is lost and the amino acid chain becomes a semi-flexible polymer chain [232], [233]. Each state can be stable in a certain environment, existing at the bottom of an energy well. At the bottom of an energy well and in the absence of sufficient energy input to lift it out, a structure can persist infinitely; the structure is in thermodynamic equilibrium. Thermodynamic stability is a

measure of how stable any structure is and describes how much energy is required to shift a structure from its native folded state into a different state. In the case of protein thermodynamics this is usually expressed as a transition from a folded state into an unfolded state [234], [235] [236][237][238].

## 4.2 Thermodynamic Characterization of Polyproteins

### 4.3.1 Chemical Denaturation

Measurement of tryptophan fluorescence intensity at 317nm versus denaturant concentration was performed according to the protocol described in Section 2.2.5.2. [239]. Homopolyproteins containing only 1 species of I27 domain were fitted to a two-state model of unfolding using Equation 4.1:

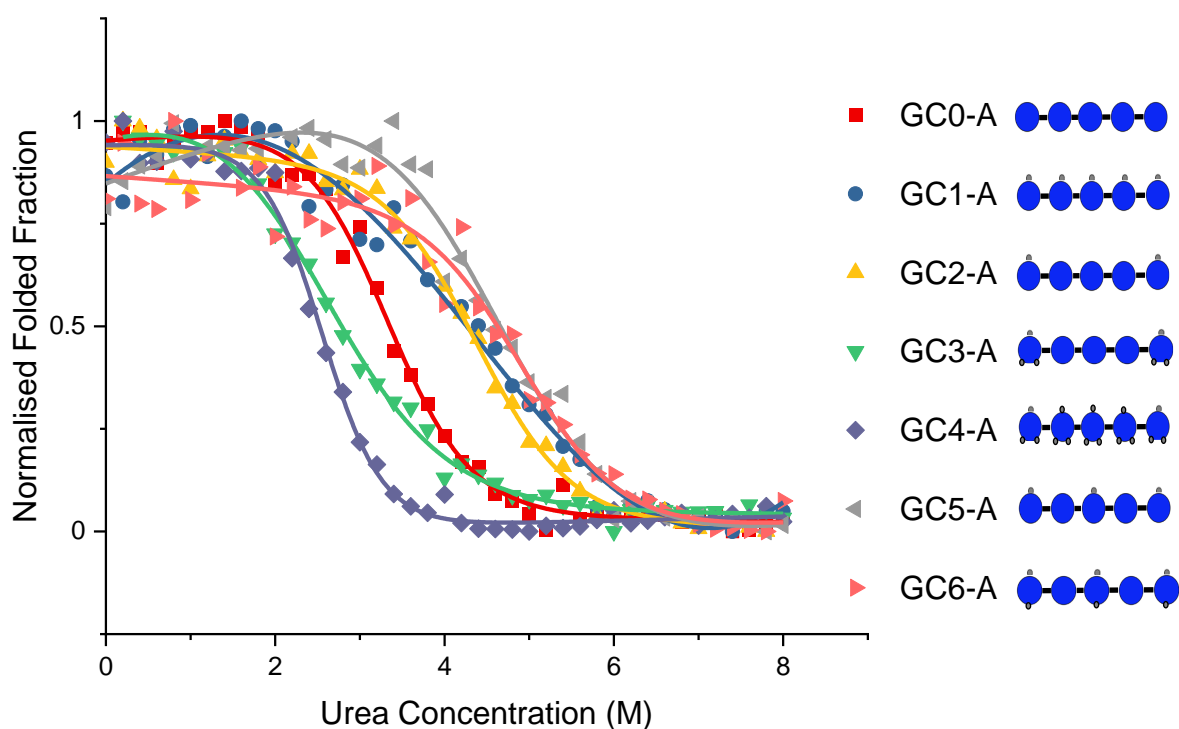
$$obs\ signal = \frac{\left( (A * x + B) + \left( \exp\left(\frac{(-G + m * x)}{(R * T)}\right) \right) * (C * x + D) \right)}{\left( 1 + \exp\left(\frac{(-G + m * x)}{(R * T)}\right) \right)}$$

Where  $-G$  ( $\text{kJmol}^{-1}$ ) is the equilibrium stability,  $m$  is the m-value,  $A$  is the denaturant dependence of the folded signal and  $C$  the unfolded, and  $B$  and  $D$  are the signal intensities of the folded and unfolded states in the absence of denaturant.  $R$  is the gas constant and  $T$  is temperature. This fitting function was used to calculate the  $\Delta G$  (Gibbs free energy) and m-value (unfolding slope gradient) for each polyprotein construct by iterative fitting of the unfolding curve. Heteropolyproteins containing >1 species of I27 domain mutant were fitted to a three-state unfolding model to extract both domain mutants sets of thermodynamic parameters according to Equation 4.2:

$$obs\ signal = \left( \frac{\left( (A*x+B) + \left( \exp\left(\frac{(G1+m1*x)}{(R*T)}\right) \right) * (C*x+D) \right)}{\left( 1 + \exp\left(\frac{(G1+m1*x)}{(R*T)}\right) \right)} \right) + \left( \frac{\left( (C*x+D) + \left( \exp\left(\frac{(G2+m2*x)}{(R*T)}\right) \right) * (E*x+F) \right)}{\left( 1 + \exp\left(\frac{(G2+m2*x)}{(R*T)}\right) \right)} \right)$$

Each half of Equation 4.2 was multiplied by the fraction of the pentamer corresponding to the number of each mutant domain species, for example GC2-A contains 2 “wild type” domains and 3 Y9S domains, so the first half of Equation 4.2 was multiplied by 0.4 and initial  $\Delta G$  and m-values were taken from the two-state fitting of GC1-A to model the wild type domains unfolding. The second half of the equation was multiplied by 0.6 with  $\Delta G$  and m-value initialization values taken from the two-state fitting of GC0-A to model the Y9S domains unfolding. After initialization all parameters were free fitted to final reported values (Table 4.1). Individual curves were normalized to between 1 (maximum folded fraction) and 0 (minimum folded fraction) for direct comparison on a single graph (Figure 4.1). The urea concentration at the normalized unfolded fraction value of 0.5 was then calculated to yield the urea concentration at which 50% of the protein domains became unfolded (UF50). This data is summarized in Figure 4.1 and Table 4.1.

Figure 4.1: Overlaid chemical denaturation curves of I27 polyprotein constructs with schematic diagrams on the right illustrating the mutant domains present in each construct.



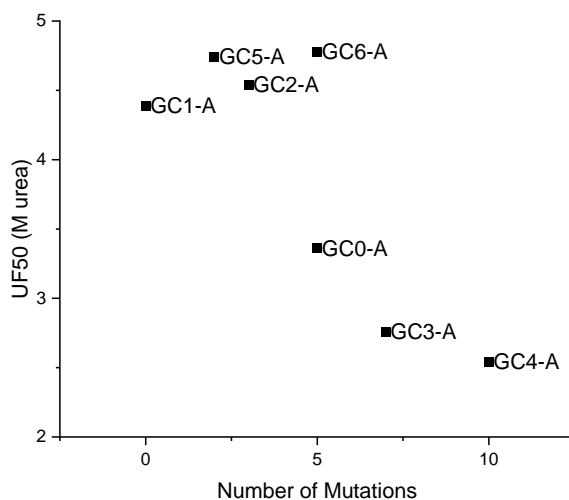
Construct	$\Delta G1$ (kJ.mol <sup>-1</sup> )	M1-value (kJ.mol <sup>-1</sup> )	$\Delta G2$ (kJ.mol <sup>-1</sup> )	M2-value (kJ.mol <sup>-1</sup> )	G1/M1 Domains Analysed	G2/M2 Domains Analysed	UF50 (M urea)
GC0-A	16.3	5	N/A	N/A	Y9S	N/A	3.36
GC1-A	14.1	3.2	N/A	N/A	WT	N/A	4.4
GC2-A	15.3	2.7	12	3.6	WT	Y9S	4.54
GC3-A	16.4	4.4	12.3	5.4	H31Y/S44Y	Y9S	2.77
GC4-A	17.9	7	N/A	N/A	H31Y/S44Y	N/A	2.55
GC5-A	20.0	4.9	16.4	3	Y9S	WT	4.75
GC6-A	15.1	3.9	17.1	3.2	Y9S	H31Y	4.79

Globally there appears to be a group of four constructs (GC1-A, GC2-A, GC5-A, and GC6-A) which exhibit similar denaturation profiles (Figure 4.1) and minimal differences in UF50 concentration (Table 4.1). In contrast a group of three constructs (GC0-A, GC3-A, and GC4-A)

with clear shifts in denaturation profiles towards a lower UF50 concentration indicating a decline in overall thermodynamic stability. This appears to be the result of a general correlation between the total number of mutations present in each polyprotein as shown in Figure 4.2.

---

Figure 4.2: Total number of mutations in a polyprotein construct versus the UF50 concentration.



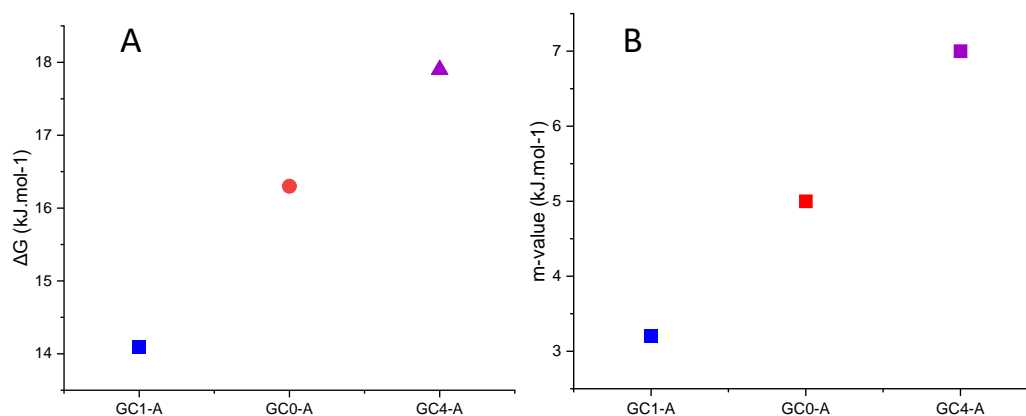
---

The other thermodynamic parameters estimated during this fitting process ( $\Delta G$  and m-value) are complex to interpret. This is due to the convolution of more than one I27 domain mutant unfolding curve in most of the constructs. However the homopolyproteins GC0-A, GC1-A, and GC4-A can be used to show a general order of chemical stability decline from WT>Y9S>H31Y/S44Y (Figure 4.3). This is the same order of decline seen in Figure 4.2 in UF50 concentration.



---

Figure 4.3: A;  $\Delta G$  of homopolyproteins GC1-, 0-, and 4-A, B; m-value of homopolyproteins GC1-, 0-, and 4-A



---

This reveals that all mutation species have had a negative effect on the chemical stability of the I27 domain. However only GC0-A, GC4-A and GC3-A exhibit significant global decreases in UF50 concentration, whilst all others show relatively little change relative to the WT.

#### 4.3.2 Thermal Denaturation

Thermal denaturation of all polyprotein constructs was performed according to the protocol described in Section 2.2.5.5.2. Enthalpy of denaturation ( $\Delta H$ ), change in heat capacity ( $\Delta C_p$ ), and melting temperature ( $T_m$ ) were calculated using a two-state fitting function [240] of circular dichroism change at 222nm versus temperature according to Equation 4.2:

$$Obs\ signal = \frac{(a * x + b) * \exp\left(\left(\frac{DH * \left(1 - \left(\frac{x}{Tm}\right)\right) - DC * \left((Tm - x) + x * \ln\left(\frac{x}{Tm}\right)\right)}{R * x}\right)\right) + (c * x + d)}{1 + \exp\left(\left(\frac{DH * \left(1 - \left(\frac{x}{Tm}\right)\right) - DC * \left((Tm - x) + x * \ln\left(\frac{x}{Tm}\right)\right)}{R * x}\right)\right)}$$

Where DH is  $\Delta H$ , DC is  $\Delta C_p$ , and R is the gas constant (8.314). Initial versus unfolded CD spectra and 222nm versus temperature graphs with fit lines in red for each polyprotein construct are summarised in Figure 4.3. Due to the sheer number of parameters to be calculated (6) it was not possible to fit these thermal denaturation curves to a three-state model. Instead global parameters have been fitted for the heteropolyproteins and used to interpret the effect of the mutant domains they contain in reference to the homopolyproteins. The thermal denaturation of all polyproteins was irreversible therefore it is inappropriate to discuss or analyse the  $\Delta H$  or  $\Delta C_p$  estimated by the fitting function.

Figure 4.4: Folded versus unfolded full CD spectra and individual denaturation curves at 222nm for all constructs.

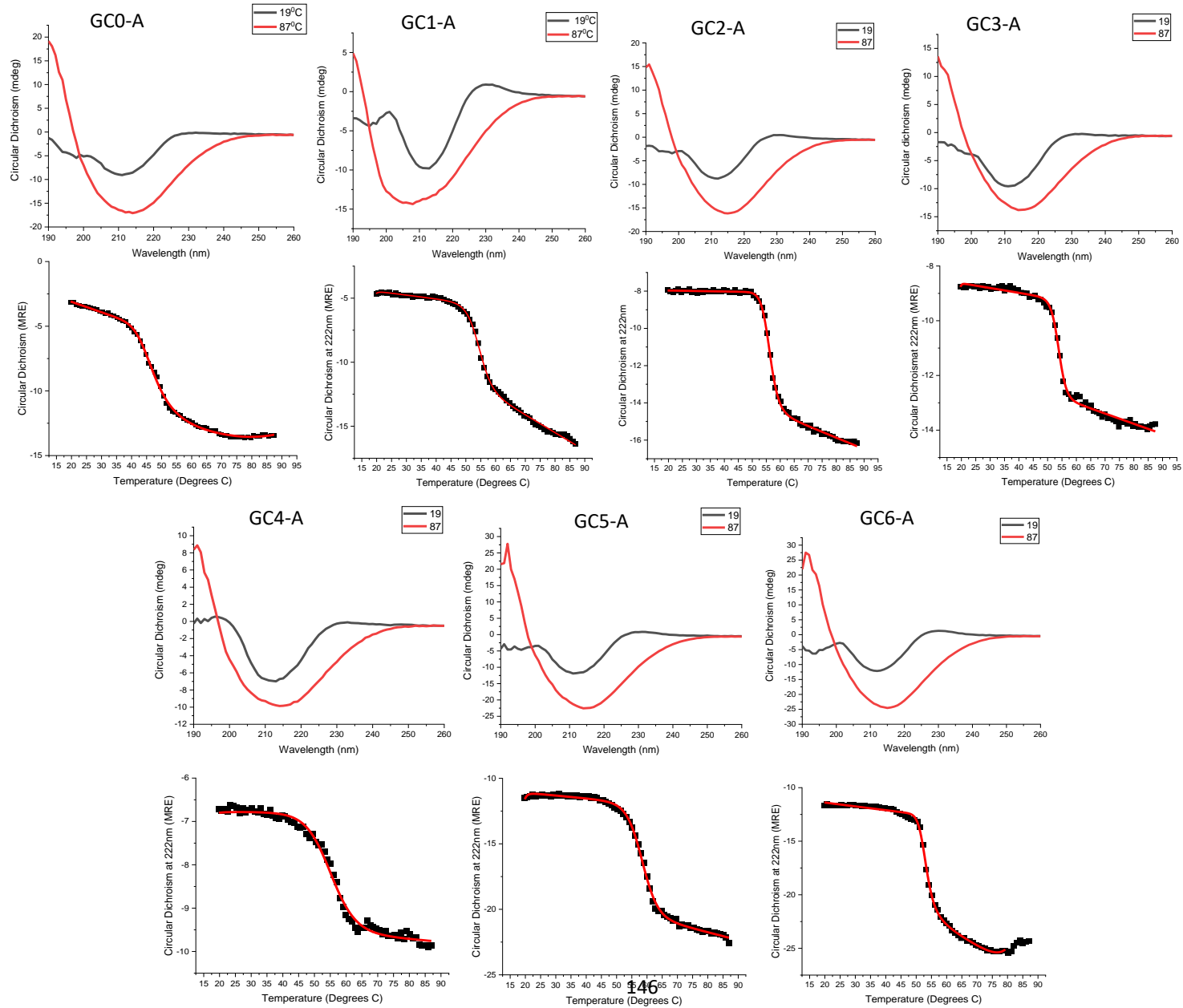


Table 4.2: Summarised Melting Temperatures of Polyprotein Constructs Calculated from Thermal Denaturation CD Data	
Construct	T <sub>m</sub> (°C)
GC0-A	50.4
GC1-A	54.2
GC2-A	56.1
GC3-A	53.8
GC4-A	54.8
GC5-A	58.2
GC6-A	53.2

T<sub>m</sub> or melting temperature is defined as the temperature at which the protein has become >50% unfolded. This can be defined on a Y-axis normalised between 1 (maximum folded fraction) and 0 (minimal folded fraction) or as the midpoint of the exponential sigmoidal decay during the fitting process of raw data. An increase/decrease in T<sub>m</sub> is indicative of an increase/decrease in global thermal stability though not the differential sampling of partially unfolded/folded intermediate states.

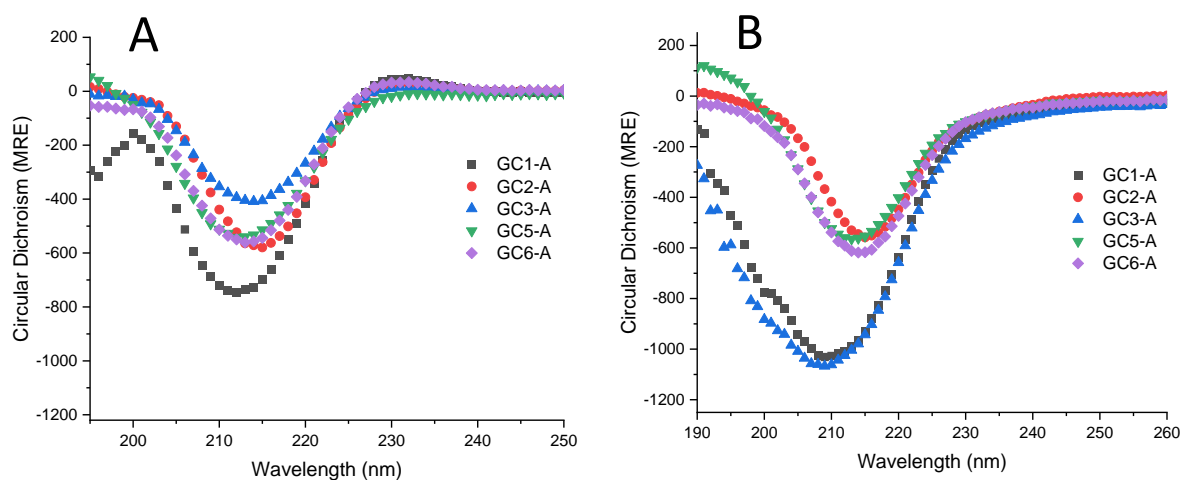
Five of the seven constructs exhibit extremely close T<sub>m</sub>'s; GC1-, 2-, 3-, 4-, and 6-A collectively have an average T<sub>m</sub> of 55±1.8°C which in real terms is within experimental and fitting error indicating that these are in reality the same. The two constructs which exhibit changes in T<sub>m</sub> relative to the wild type GC1-A are GC0- and 5-A. GC0-A shows a significant decline in T<sub>m</sub> of >3°C. This strongly indicates that the Y9S domains from which GC0-A is exclusively composed are less thermally stable than the wild type. The native tyrosine residue of I27 is involved in the hydrogen bond clamp between the A' and G strands of the protein [178] and so weakening of this H-bond dense zipper may result in this reduction in thermal stability as any reduction in the intra-molecular bonding network of a protein reduces its capacity to absorb energy

prior to unfolding. GC5-A shows a 4°C increase in  $T_m$ . This is unexpected due to the presence of two Y9S domains in GC5-A, and so the failure of these domains to reduce the thermal stability of the protein suggests a cooperative mechanism between domains [241]. This same effect appears to manifest in GC2-A which despite three Y9S domains has near identical  $T_m$  to GC1-A. GC4-A is the second mutant homopolyprotein containing five H31Y/S44Y domains, and shows no significant decline in  $T_m$  indicating that the H31Y/S44Y double mutation has a minimal effect on thermal stability. GC4-A was subsequently discovered to be highly insoluble, likely as a result of the addition of 10 hydrophobic tyrosine residues. GC4-A could not be formulated to greater than ~5mg/mL, and was therefore unable to form a hydrogel.

### 4.3 Solid-State Circular Dichroism

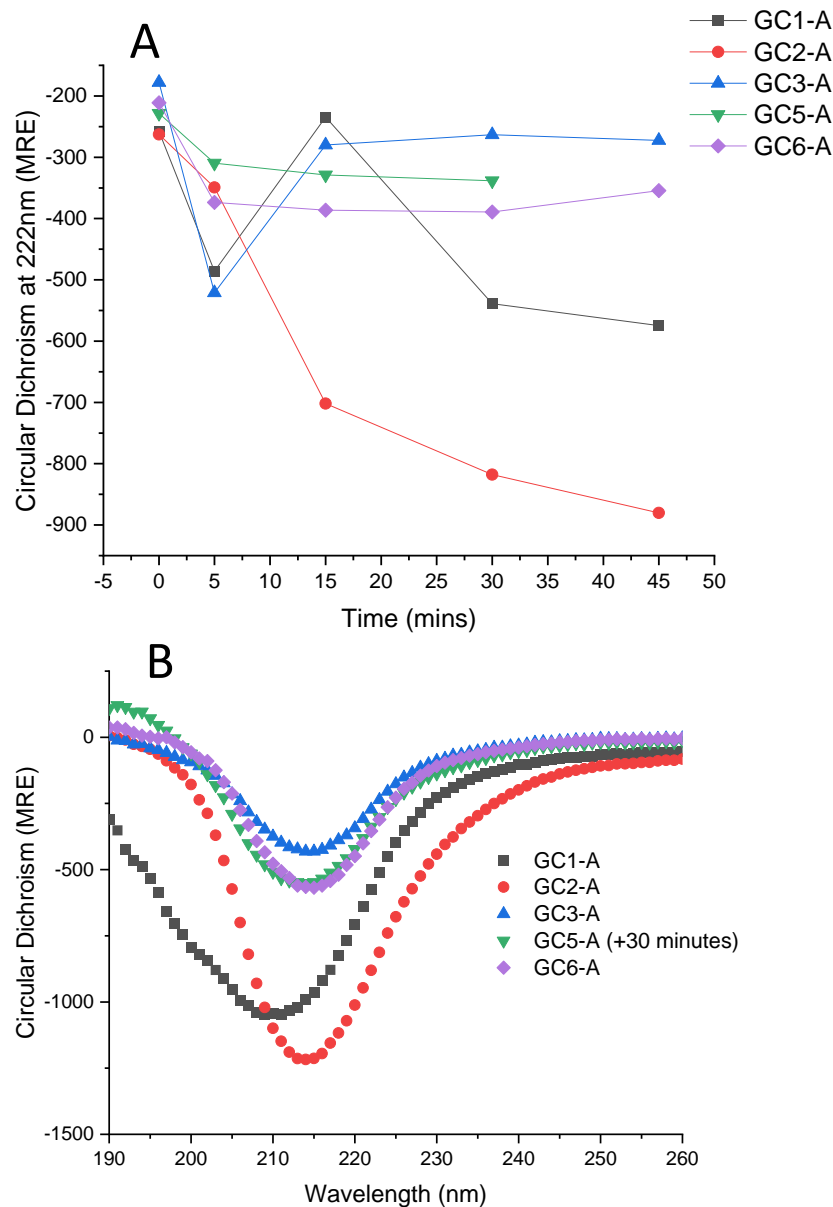
Once each construct had been characterized at the single molecule level solid-state CD measurements were performed on hydrogels formed from each in order to confirm that they retained a degree of native structure post-gelation according to the protocol described in Section 2.2.5.5.3. The gelled nature of the sample was confirmed post-measurement by an abrasion test of the material. GC4-A was not measured due to its high degree of insolubility making it impossible to re-suspend at a gel-forming concentration (25mg/mL), and GC0-A was not measured because it is unable to form a hydrogel (Section 5). Displayed in Figure 4.5A are the overlaid solution spectra of each construct prior to gelation, and in Figure 4.5B the overlaid spectra of each construct immediately post gelation (~5 minutes) (IPG).

Figure 4.5: A; Overlaid solution CD spectra of each construct prior to gelation, B; Overlaid CD spectra of each construct hydrogel immediately post gelation.



All solution spectra show that all constructs exhibit an approximately equal degree of foldedness prior to gelation and that all are primarily  $\beta$ -sheets (Figure 4.5A). The minor variations of GC1-A and GC3-A are likely the result of different qualities of resuspension (Section 6). Figure 4.5B shows that GC2-, 5-, and 6-A have undergone very little spectral change during gelation. The GC1-A and GC3-A spectra have undergone a significant degree of change during gelation and exhibit both  $\beta$ -sheet and random coil characteristics. A degree of continued spectral change occurs in all hydrogel species during gelation and continues across a 45 minute timecourse as demonstrated in Figure 4.6A, but at varying rates.

Figure 4.6: A; Circular dichroism at 222nm for each construct hydrogel at discrete timepoints (0=solution, 5=IPG), B; Overlaid CD spectra of all hydrogel species at 45 minutes post-gelation.



All species exhibit a gradual decline in  $\beta$ -sheet spectral characteristics after 15 minutes of approximately the same rate with the exception of GC2-A which shows a significantly greater rate of change. The mechanism for this may be linked to the molecular entanglements which are anticipated to be the gel-forming interaction species in these gels (Section 6). Figure 4.6B shows that all hydrogel constructs retain a mostly  $\beta$ -sheet-like spectrum 45 minutes post gelation, although GC1-A shows a significant shift towards a more random-coil dominated

spectrum, and the decline of signal at 222nm of GC2-A suggests that a significant shift towards a random coil signal may have occurred.

#### 4.4 Discussion

##### 4.4.1 Thermodynamics

The chemical stability of I27 domains are clearly deleteriously affected by the introduction of mutations. Neither of the homopolyproteins showing the most significantly reduced chemical stability built from mutant domains (GC0-A and GC4-A) are suitable for hydrogel formation, and are therefore not of concern for the rest of this study. The constructs which were used for hydrogel formation and characterization all exhibited minimal changes in chemical stability and in optimal conditions retain near-identical secondary structures with the exception of GC3-A which does exhibit a reduction in chemical stability which must be kept in mind when reviewing hydrogel data.

All three mutations reduce the thermal stability of individual domains. Y9S likely disrupts the hydrogen bonding network of the domain causing a more significant reduction in thermal stability than the H31Y/S44Y double mutant [184]. The H31Y mutation appears to have a less deleterious effect on stability than the other two mutation species possibly because in terms of hydrophobicity, charge, and side chain packing tyrosine and histidine are more similar giving this substitution a low penalty score in amino acid substitution matrices [242][243][244]. The inability of GC0-A to form a hydrogel (Section 5) and the inability of GC4-A to be formulated to gel forming concentrations means that neither is of concern in characterizing hydrogel mechanics and so their reduced thermal stability is not a concern. The constructs which will be investigated going forward (GC1-, 2-, 3-, 5-, and 6-A) all exhibit very minor differences in overall thermal stability.



#### 4.4.2 Protein Secondary Structure in Hydrogels

These results demonstrate that all five hydrogel-forming constructs retain varying degrees of  $\beta$ -sheet-like secondary structure up to 45 minutes post-gelation. This suggests that some I27 domains remain folded post-gelation with the exception of GC1-A which appears to have undergone a more significant shift towards a random coil spectrum. However it cannot be ruled out that the source of the  $\beta$ -sheet spectrum post-gelation is not an amyloid fibril-like signal generated by a form of fibrillogenesis [245], [246]. It is possible that the spectrum measured is a result of amyloid-like structures formed by the crosslinking and subsequent cross- $\beta$ -sheet packing of the domains, and therefore we cannot conclusively state from this data that these hydrogels contain folded protein domains, only that it is possible but requires further verification.

## Section 5: Quantification of Hydrogel Crosslinking Efficiency and Post-Gelation Folded Fraction

### 5.1 Introduction

As degree of interconnectedness within a network (encompassing both physical and chemical interactions) defines its mechanical characteristics it is important to attempt to quantify the frequency of these interactions. Physical interactions are difficult to quantify as by nature they do not cause permanent and identifiable alterations to the network-forming molecule. This is in stark contrast to chemical crosslinks which result in chemical alterations to the molecules involved.

The macroscale mechanics of a material are definable using rheometric methods as described in Section 6. However rational design of these characteristics requires an understanding of how microscale network topology translates to the macroscale; a bridge of understanding must exist between the two length scales [247]. In order to bridge this gap two key pieces of knowledge are generally required; the volume fraction which accounts for physical interactions and packing, and the number of crosslinks per material unit [214][248]. In the case of FGP hydrogels a third major component influencing macroscale mechanics is added; the folded/unfolded state of the FGP domain building blocks from which the network is formed [249][93][250]. This is due to the behavior of FGP's as mechanical springs when under sub-rupture strain, with this ability to deform and then recover in an elastic fashion having a major effect on macroscale mechanics. The degree of unfoldedness post-gelation is therefore an important measurement to make in order to characterize the mechanics of FGP hydrogels. The measurement of crosslinking efficiency and unfolded fraction were performed simultaneously upon each gel during this study allowing direct correlations to be identified.

Throughout this work the term “monomer” is defined as a single pentamer polyprotein molecule composed of 5 concatemerized I27 domains.

## 5.2 Quantifying Crosslinking Post-Gelation

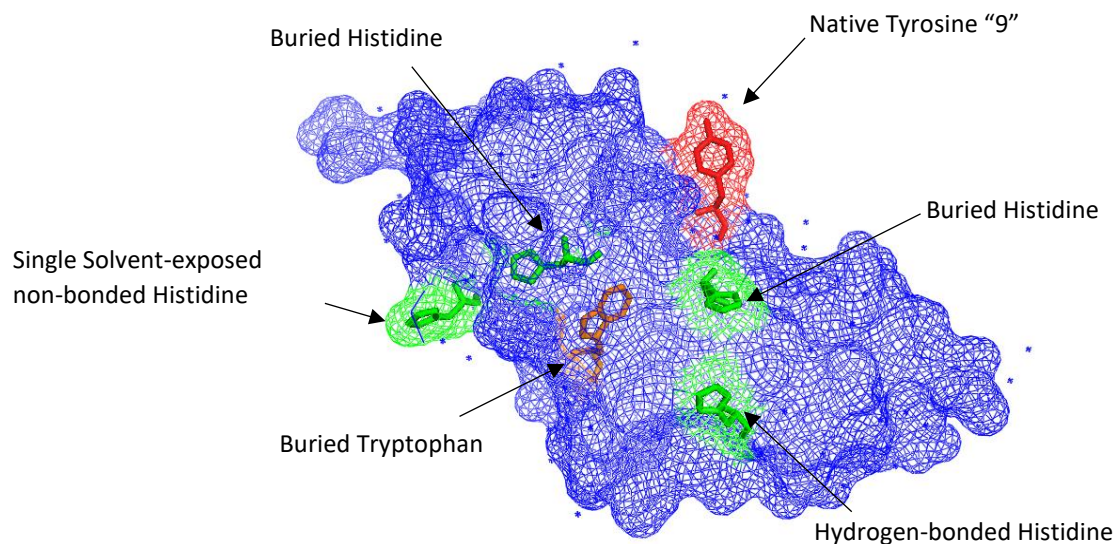
All construct polyprotein monomers were expressed and purified as described in Section 2 and their specific identity confirmed by intact mass spectrometry analysis. All gels were formulated to a final concentration of 487.9 $\mu$ M corresponding to 1.7<sup>-8</sup> mols of protein in a 35 $\mu$ L volume, and crosslinked as described in Section 2.2.5.1.

### 5.2.1: Investigating Crosslinking Reaction Specificity

In order to accurately determine the density of chemical crosslinking within a hydrogel network it must first be demonstrated that the only mechanism of chemical crosslinking present is that which was designed and is to be quantified. In order to do this it must be shown that in the absence of the molecules involved in the designed crosslinking mechanism no crosslinking occurs. The specificity of the crosslinking reaction used during this study is discussed in detail in Section 1.2.3.2 [139]. In summary tyrosine-tyrosine crosslinking is the dominant product of the reaction, but cysteine-tyrosine crosslinks are also likely to form, with tyrosine-histidine crosslinks an even lower abundance side reaction. I27 domains contain a single buried tryptophan and phenylalanine residue in addition to a buried cysteine residue, meaning that none of these are able to crosslink. In addition I27 contains 4 histidine residues; two are buried, one forms a hydrogen bond with a glutamic acid, leaving one solvent exposed and potentially available to crosslink in an off-target manner. This histidine residue was mutated to tyrosine in domains designed to contain 2 and 3 tyrosine crosslink sites and is designated the “H31Y” mutant.

---

Figure 5.1: I27 Domain with Potential Crosslinking Residues Highlighted



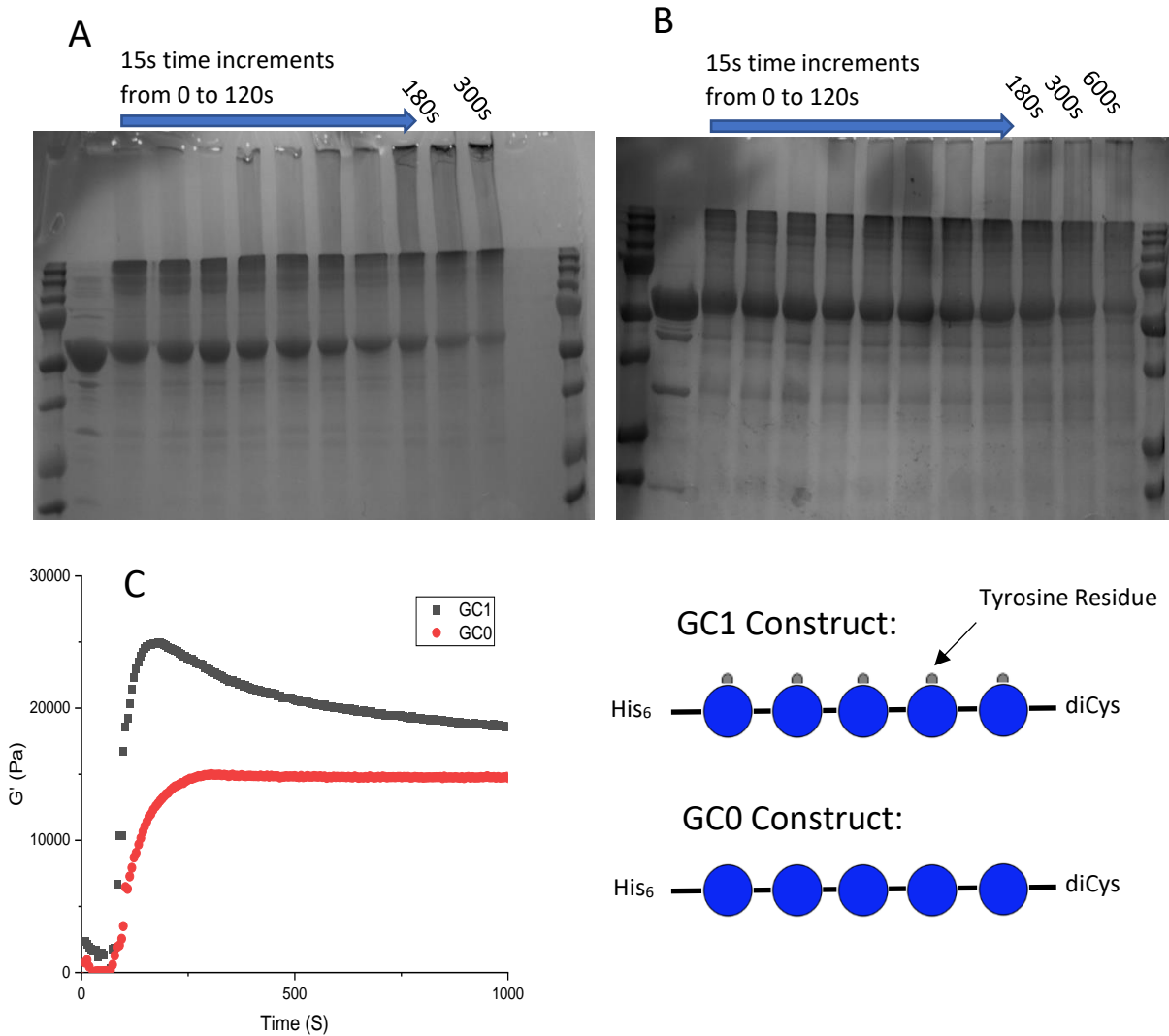
---

For purification purposes the polyproteins were expressed with a hexa-His tag, and because they were originally designed for dynamic force spectroscopy studies contained a C-terminal dicysteine motif [179]. To assess whether these residues were able to undergo “off-target” crosslinking a control construct containing 5 “Y9S” I27 domains (i.e. no tyrosine residues) designated GC0 (Figure 5.2) was expressed. A crosslinking reaction of GC0 and GC1 (an I27 pentamer containing 5 tyrosine residues) was performed at low concentration preventing hydrogel formation.

Low-concentration (45 $\mu$ M) crosslinking of GC1 yielded the rapid evolution of large oligomer species whose large molecular mass precluded entry into the resolving gel of an SDS-PAGE gel (Figure 5.2A). The same behaviour was observed of GC0 indicating that significant off-target crosslinking occurs in the absence of tyrosines. In addition crosslinking of GC0 at a concentration which might allow in gelation (100mg/ml) resulted in a sol-gel transition measured rheologically (Figure 5.2C), indicating that sufficient crosslinking had occurred for

a network to form. This is intolerable for a study characterising the relationship between crosslink density and mechanical characteristics.

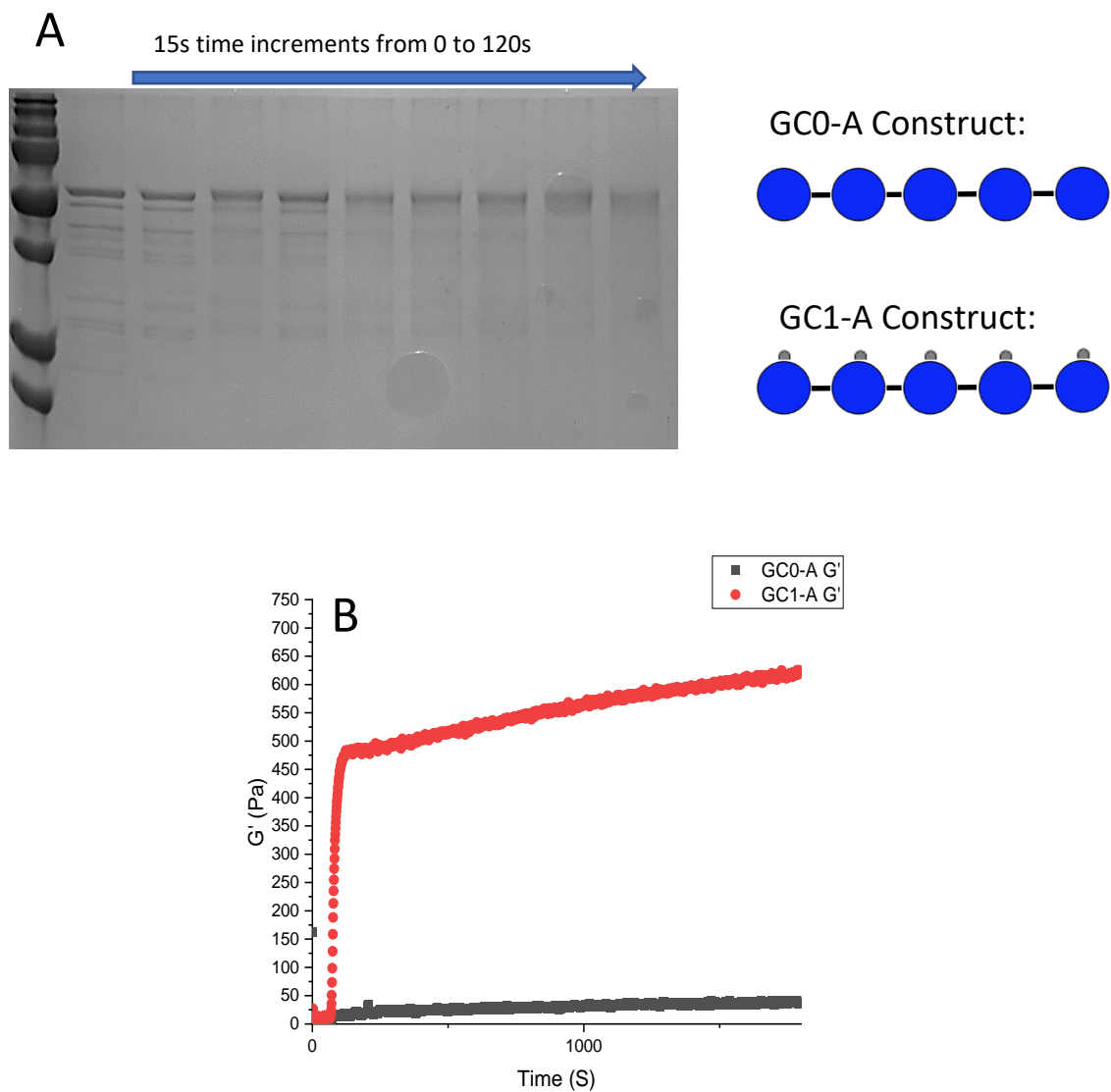
Figure 5.2: A; 45 $\mu$ M GC1 crosslinked with a sample taken every 15 seconds, B; 45 $\mu$ M GC0 crosslinked with a sample taken every 15 seconds, C; Timesweep rheological measurement of the G' evolution of GC1 and GC0 during crosslinking at 100mg/ml with irradiation proceeding for 5 minutes.



The obvious candidates for this undesired off-pathway crosslinking were the N-terminal His-tag and the C-terminal dicysteine motif, and so another control construct was created without a dicysteine motif and with a TEV-cleavage sequence to allow the post-purification removal of the Hexa-His tag. This new construct is designated GC0-A (Figure 5.4). Crosslinking of GC0-A at both low and high concentration as described above yielded no higher order oligomers

when visualised by SDS-PAGE gel, and undergoes no sol-gel transition when measured rheologically, even when irradiated for 15 minutes (Figure 5.3).

Figure 5.3: A; 20 $\mu$ M GC0-A crosslinked with a sample taken every 15 seconds B; Timesweep rheological measurement of the  $G'$  evolution of GC0-A and GC1-A formulated at 40mg/ml during constant irradiation.



The absence of oligomerisation or network formation by GC0-A shows conclusively that in the absence of tyrosine residues constructs lacking a His-tag or dicysteine motif cannot form

sufficient crosslinks to allow network formation, meaning that any off-target crosslinking which does occur will make little to no contribution to the macroscale mechanical characteristics of the hydrogel. For this reason we can be confident that all crosslinking leading to network formation is due to tyrosine-tyrosine crosslinking, and that the number of dityrosine adducts corresponds to the total number of chemical crosslinks present in a gel sample.

### 5.2.2: Measuring the Photometric Equivalency of Monomeric Dityrosine and Dityrosine Adducts within Protein Hydrogels

In order to quantify the crosslinks present in a network the crosslinked moiety must be isolated from the mixture of molecules within the gel by analysis of some unique property which allows its identification from the noise generated by all the other molecules present. Throughout this study the crosslinking mechanism utilized to form protein hydrogels was the photoactivated tris-bipyridylruthenium (II) ( $\text{Ru(II)bpy}_3^{2+}$ )-mediated tyrosine crosslinking reaction described in Section 1.2.3.2 [251]. This reaction yields dityrosine adducts which exhibit specific and unique photometric properties. Dityrosine's absorbance  $\lambda_{\text{max}}$  is 315nm with a fluorescence emission peak at 410nm, making it distinct from any other fluorescent molecules present in the hydrogel [252]. In this study hydrogels were formed using each I27 polyprotein construct described in Section 3, before trypsinisation to degrade the gel into small peptide fragments including, when present, those containing crosslinked dityrosine adducts. Subsequent fluorescence analysis allowed the quantification of the number of moles of dityrosine present in the sample followed by back-calculation to determine the crosslinking efficiency achieved globally and the average number of crosslinks formed per monomer (I27 pentamer) of starting material.

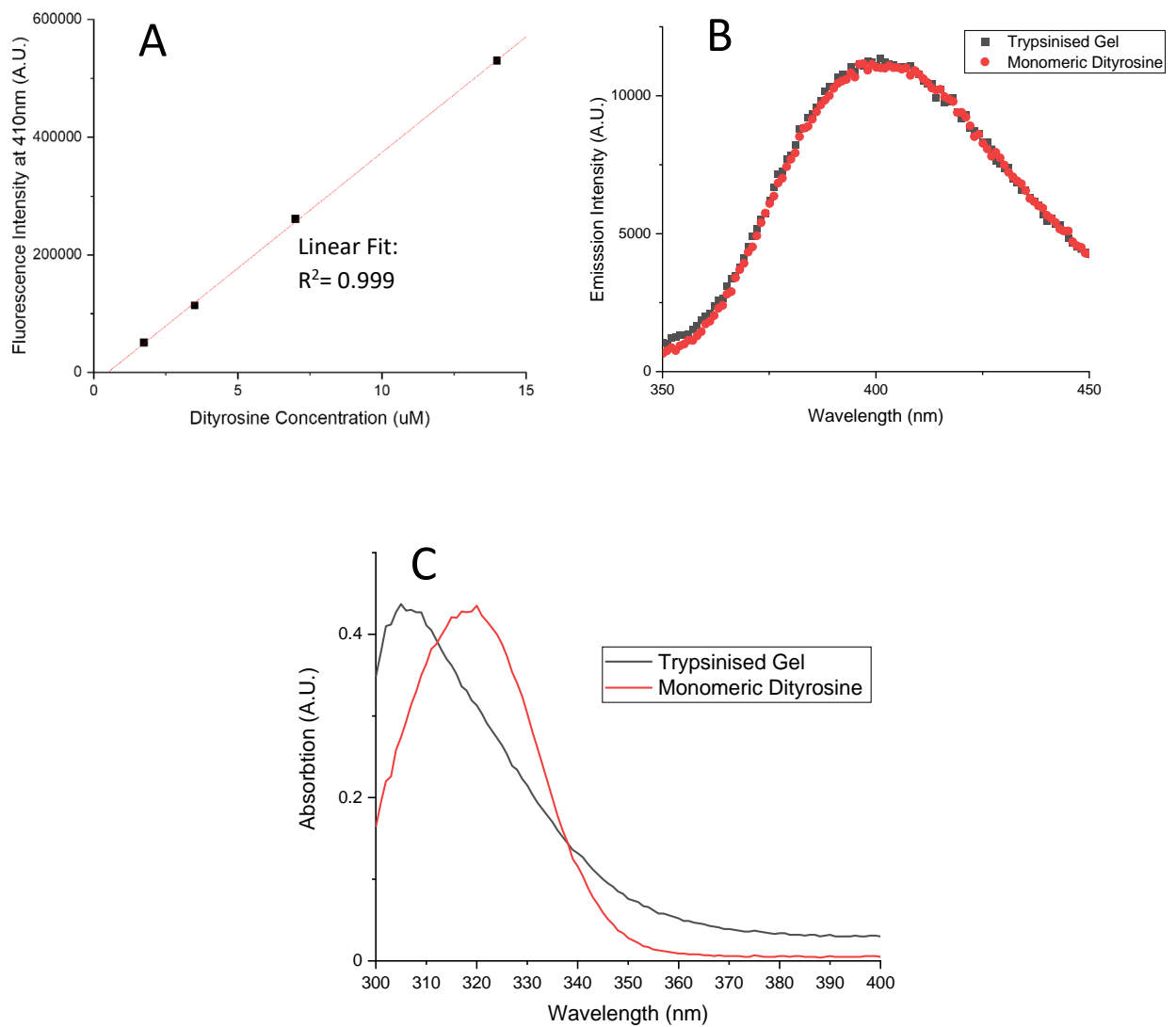
Crosslink efficiency was calculated as a percentage of the number of dityrosine molecules that would result upon all tyrosines present in the gel solution forming a crosslink. Each gel was formulated to a final concentration of 487.9 $\mu$ M polyprotein corresponding to 1.7<sup>-8</sup> mols of protein in 35 $\mu$ L volume. The maximum theoretical number of mols of dityrosine adducts that would be yielded by 100% crosslinking efficiency for each construct was therefore calculated as mols of protein multiplied by the number of tyrosines per monomer (127 pentamer), divided by 2. Fluorescence quantification of the dityrosine adducts present in each degraded gel sample was then performed using a standard curve of monomeric dityrosine fluorescence emission at 410nm; Figure 5.4A. To assess the validity of using monomeric dityrosine as a standard with which to quantify dityrosine adducts contained as part of a peptide fragment the photometric equivalency of these two species was assessed by measuring the similarity between the absorbance and fluorescence emission spectra of monomeric dityrosine and a trypsinised hydrogel sample. Using the published molar extinction coefficient for dityrosine of 5340 M<sup>-1</sup>cm<sup>-1</sup> [252] a dityrosine standard and a trypsinised gel sample were prepared to 80 $\mu$ M dityrosine and their absorption spectra obtained (Figure 5.4C). The absorbance spectrum of dityrosine in the trypsinised gel sample was obtained by subtracting the background absorbance spectrum of the digestion buffer (Section 2.2.5.3) from the convoluted trypsinised gel sample absorbance spectrum. The absorption peaks of the two dityrosine species were 310nm for the gel sample and 320nm for the monomeric dityrosine, most likely due to minor differences in pH. However the absorbance at 315nm was identical between the two samples as are their maximal absorbance intensities, demonstrating that when quantified using their 315nm absorbance coefficient of 5340M<sup>-1</sup>cm<sup>-1</sup> each species of dityrosine yields the same result. Figure 5.4B shows that when excited at 315nm both samples exhibited identical fluorescence spectra with identical maximum emission intensities. These



results prove that the two molecules are photometrically identical, and therefore monomeric dityrosine can be used as a standard with which to quantify the dityrosine adducts formed during hydrogel gelation. Fluorescence was used for quantification rather than absorbance due to the superior sensitivity of this technique.

---

Figure 5.4: A; Standard Curve of Monomeric Dityrosine Concentration versus Fluorescence Intensity with Linear Fitting, B; Fluorescence Spectra of 80 $\mu$ M Monomeric Dityrosine Standard and Trypsinised Hydrogel Sample Prepared to 80 $\mu$ M Dityrosine when Excited at 315nm, C; Absorbance Spectra of Monomeric Dityrosine and Dityrosine Adducts in Hydrogel Sample both Formulated to 80 $\mu$ M



### 5.2.3: Subtracting IAEDANS/Dityrosine Fluorescence from Full Hydrogel Fluorescence Spectra

The absorbance  $\lambda_{\text{max}}$ 's of dityrosine and IAEDANS are 315nm and 336nm respectively meaning that excitation of one does result in excitation and a degree of fluorescence from the other. Therefore in order to perform crosslink and unfolded fraction quantifications (Section 5.3) simultaneously the background spectrum of each must be subtracted from the full gel spectra to obtain a clean fluorescence spectrum of each fluorophore and allow accurate concentration calculations. When excited at either 315nm or 336nm the raw spectrum obtained from a degraded labelled hydrogel sample exhibited a convoluted spectrum with distinct peaks at 410nm and 490nm corresponding to dityrosine and IAEDANS fluorescence respectively. Two spectra of each gel sample were obtained with excitation at 315nm and 336nm, a dityrosine standard spectra was obtained using 336nm excitement, and an IAEDANS standard spectrum was obtained at 315nm excitement. The 410nm emission peak of the 336nm excited dityrosine standard spectrum was then normalized to the 410nm emission value of the 336nm excited gel spectrum, before being subtracted from the full gel spectrum. The same process was performed using the 315nm excited IAEDANS standard spectrum, 315nm excited gel spectrum, and normalized to the 490nm emission values. The subtraction of the standard spectra from the gel spectra resulted in separate emission spectra corresponding perfectly to the profile of dityrosine and IAEDANS and therefore suitable for fitting onto a standard curve to calculate crosslinking/unfolded efficiency. This is summarised in Figure 5.5.

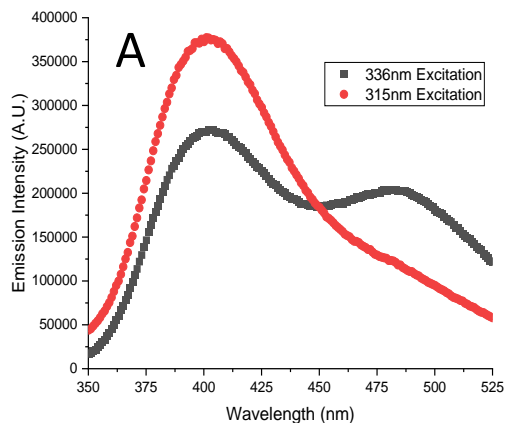
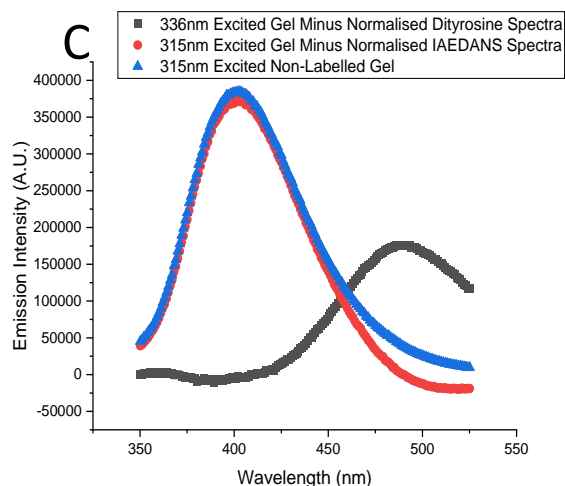
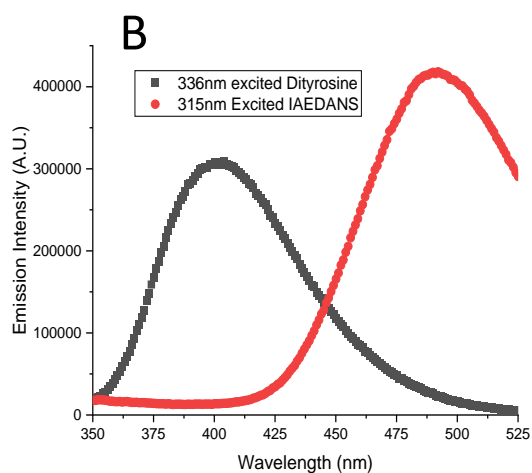


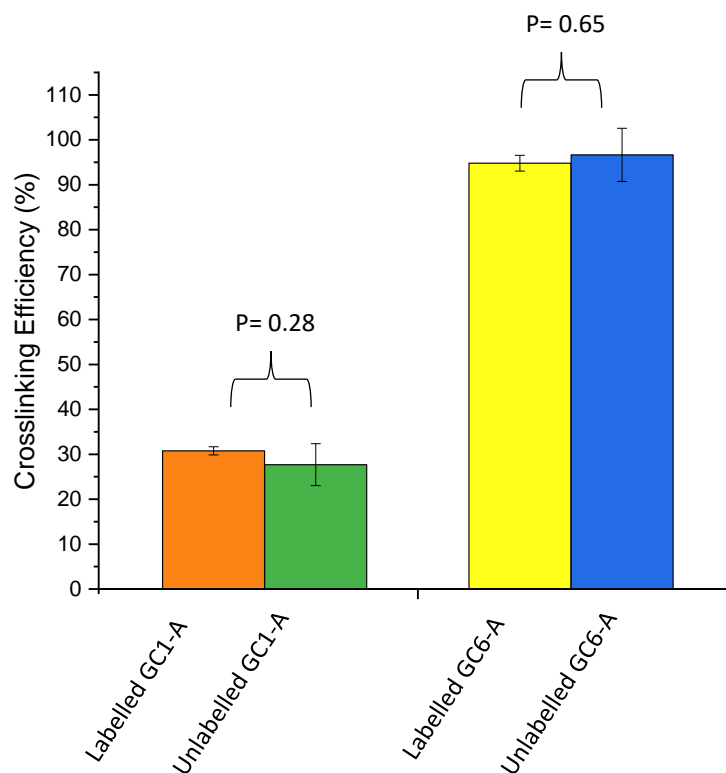
Figure 5.5: A; raw IAEDANS-labelled hydrogel fluorescence spectra at 315nm and 336nm excitation. B; Fluorescence spectra of dityrosine and IAEDANS standards measured at 336nm and 315nm respectively. C; Fluorescence spectra of labelled gel minus the normalised fluorescence spectra of the undesired second fluorophore, and the full raw spectra of an unlabelled gel excited at 315nm.



In order to validate this subtraction methodology crosslinking efficiency was calculated for technical repeats with and without labelling the gel with IAEDANS. The crosslinking efficiencies calculated for these labelled and unlabelled technical repeats showed no significant differences (Figure 5.6). This indicates that the subtraction of the IAEDANS spectrum from a labelled gel spectrum does not over or undercompensate for the background IAEDANS emission at 410nm, providing an accurate 410nm dityrosine emission value.

---

Figure 5.6: Average crosslinking efficiency with standard deviation error of same-biological repeat technical repeats of GC1-A and GC6-A hydrogels. T Test values show no statistically significant differences in crosslink quantification between labelled and unlabelled gels.



---

#### 5.2.4: Crosslinking Efficiency/Crosslinks per Monomer versus Geometry

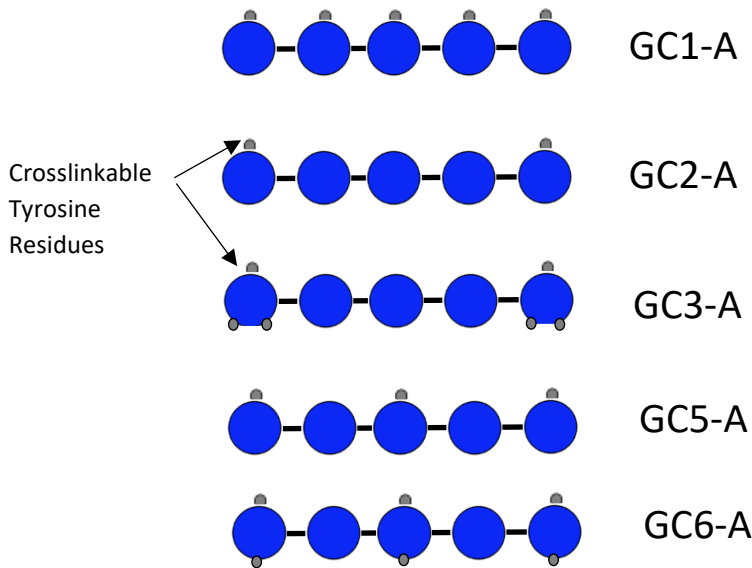
##### 5.2.4.1: Crosslinking Efficiency

Fluorescence quantification using the methodology previously described allowed firstly the concentration of dityrosine in a sample to be measured, followed by calculation of the total number of moles. The crosslinking efficiency was then calculated as a percentage; the number of moles present versus the theoretical maximum number that could have been formed with 100% efficiency had all tyrosines in every monomer formed a crosslink. This was then converted into an average number of crosslinks per monomer by multiplying the number of

tyrosines per monomer by the percentage crosslinking efficiency. A monomer is defined as a single I27 pentamer polyprotein molecule. The number and distribution of tyrosines in each construct is summarized in Figure 5.7.

---

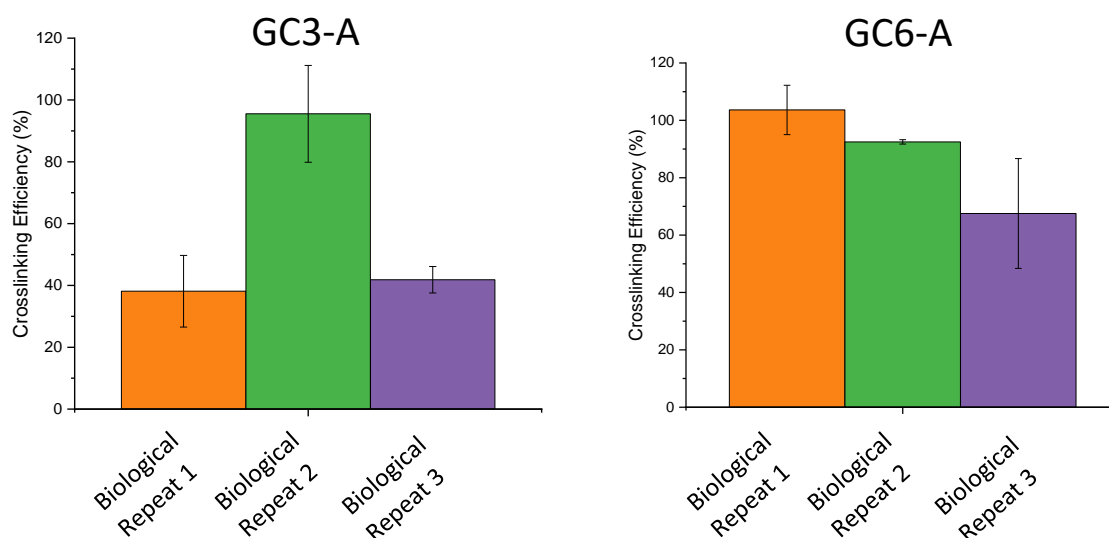
Figure 5.7: Schematic representation of the number and distribution of tyrosine residues throughout each polyprotein construct discussed in this Section.



Average crosslinking efficiency showed generally non-statistically significant variation between biological repeats, and low technical repeat variation. The constructs which showed some variation between biological repeats were GC3-A (between repeat 2 versus 1 and 3) and GC6-A (repeat 3 versus 1 and 2) as shown in Figure 5.8. The possible reasons for this are discussed in section 5.4 as are observations about protein re-suspension consistency.

---

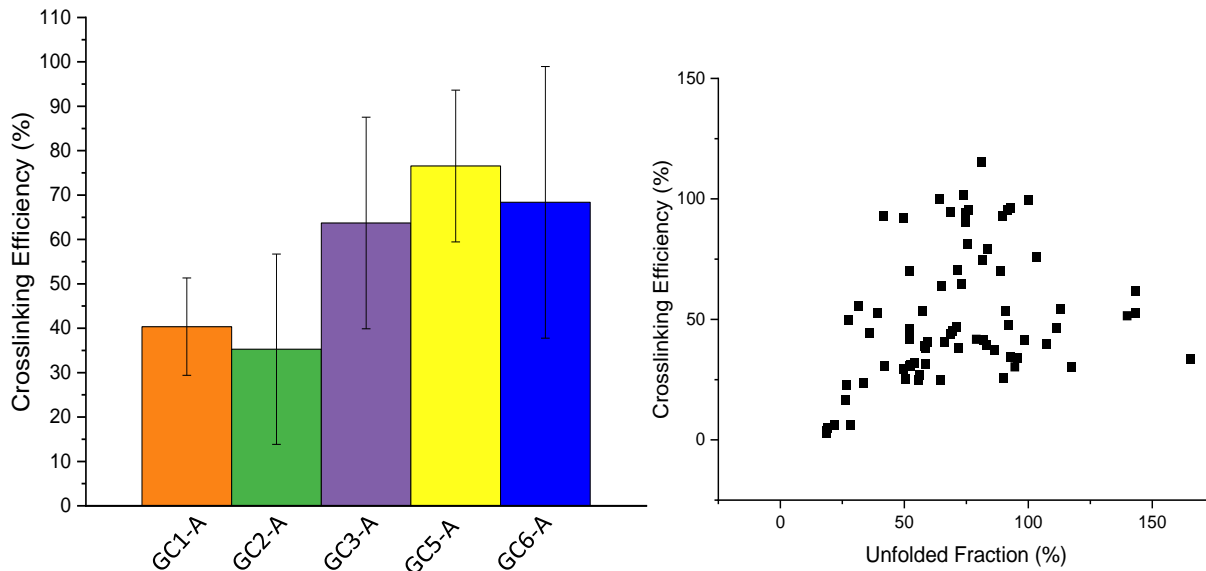
Figure 5.8: Summary of average crosslinking efficiency and standard deviation error of 3 biological repeats of GC3-A and GC6-A. Each biological repeat corresponds to n>3 technical repeats.



---

Analysis of the correlation between crosslinking efficiency and geometry was performed by averaging all individual measurements from all biological repeats. The increase in crosslinking efficiency of constructs containing more tyrosines than GC1-A (GC3-, 6-A), and the decrease in a construct containing less (GC2-A) indicates that variations in the APS : tyrosine excess of the crosslinking reaction is not the determining factor in efficiency (Table 5.1). In addition there is no correlation between the unfolded fraction and crosslinking efficiency (Figure 5.9B) indicating that it is the geometry of crosslinks that influences variations in efficiency, not the unfolded fraction prior to gelation.

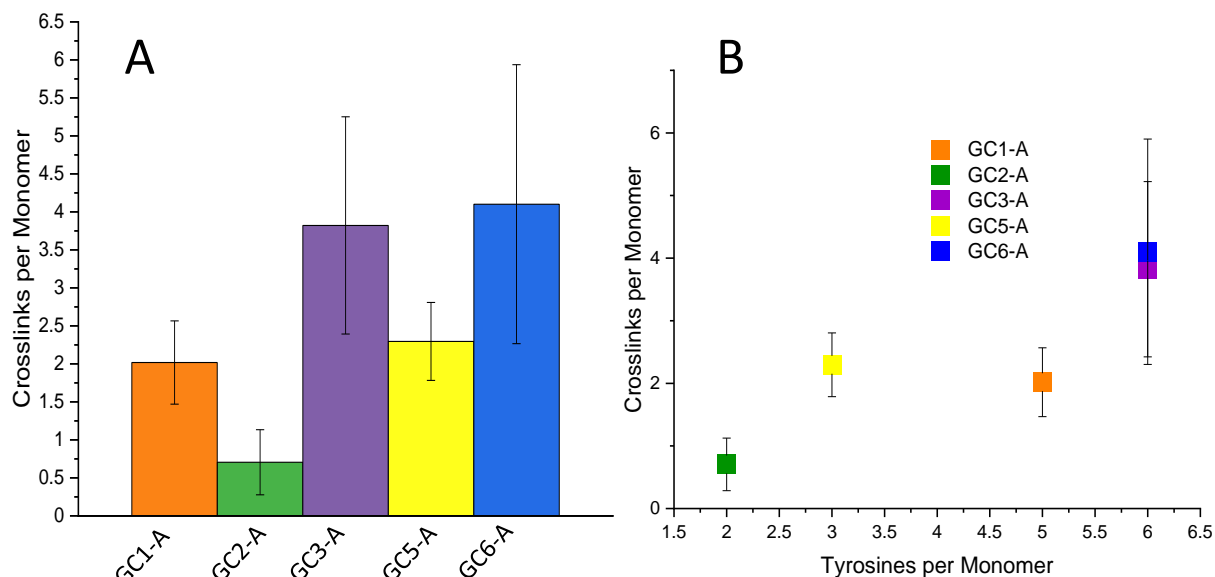
Figure 5.9: A; Summary of all constructs all samples average crosslinking efficiency and standard deviation, T-Test significance results Appendix 1 table 5.1, B; All Gel Samples Crosslink efficiency versus Unfolded Fraction Post-gelation.



#### 5.2.4.2: Crosslinks per Monomer

The aim of this study was to create a series of hydrogels with significant differences in crosslinking density and correlate these microscale alterations to changes in macroscale hydrogel network mechanics. As discussed previously, alterations in crosslinking efficiency are observed in response to changes in crosslink geometry/density, and on the molecular scale this translates to changes in the average number of crosslinks each monomer of material will form as an incorporated subunit of the network. By converting global crosslinking efficiency into crosslinks per monomer, a more nuanced picture of the relationship between crosslink geometry/density and network topology emerges.

Figure 5.10: A; Average Crosslinks per Monomer for each Construct. Calculated as number of tyrosines per monomer expressed as a function of crosslinking efficiency. Error is standard deviation, T-Test significance results Appendix 1 Table 5.2, B; Tyrosines present per monomer versus the average number of crosslinks formed per monomer.



The results summarized in Figure 5.10A show significant differences between the average number of crosslinks per monomer that each gel construct exhibits (Table 5.2). Figure 5.10B demonstrates a general correlation between the number of crosslinks per monomer and the number of tyrosines present in each monomer. An interesting exception to this rule is GC1-A which contains 5 tyrosines (1 per I27 domain) but averages only 2 crosslinks per monomer. For comparison GC5-A which contains only 3 tyrosines (I27 domains 1, 3, and 5) also averages 2 crosslinks per monomer. This suggests that the removal of the tyrosines in domains 2 and 4 has no effect on the number of crosslinks formed indicating that on average they do not become crosslinked in GC1-A, likely due to steric hindrance. We hypothesize that the two most commonly crosslinked tyrosines will be those on the two terminal domains (1 and 5) as these have the most solvent accessible surface area and are most likely to bind into the



network. Tyrosines in domains 2-4 are likely less sterically favorable to allow another molecule to diffuse to within the reaction distance for sufficient time to form a crosslink.

A second interesting comparison is between GC6-A and GC3-A (See Figure 5.7). These two constructs both contain 6 tyrosines but their distribution is radically different; GC6-A contains two tyrosines located on opposite sides of the I27 domain in repeats 1, 3, and 5, whilst GC3-A contains three tyrosines in both repeats 1 and 5. On average these constructs both form the same number of crosslinks per monomer (4) with an extremely similar standard deviation. This suggests that in GC3-A 2 out of 3 tyrosines on each of the terminal domains become crosslinked, and we can therefore assume that this will be the same in GC6-A with the two tyrosines in repeats 1 and 5 accounting for the average of 4 crosslinks per monomer. The two internal tyrosines of GC6-A are likely the source of variation in crosslinking efficiency. The significantly larger standard deviation of GC6-A relative to GC5-A further suggests that the addition of a tyrosine on both sides of the central domain increases the chance of at least one becoming crosslinked. This is further evidence of the role crosslinking geometry plays in determining the number of crosslinks per monomer.

GC2-A contains only 2 tyrosine residues; one on domain 1 and 5. This design geometry is predicted to allow the formation of elongated chains crosslinked via their N- and C-terminal domains, with little to no crossbridging via chemically crosslinked protein molecules. Instead it is hypothesized that physical interactions caused by the molecular entanglement of these long chains causes a sol-gel transition. This is discussed in detail in Section 6. A comparison between GC2-A and GC3-A provides convincing evidence of the previously mentioned hypothesis of preferential terminal domain crosslinking. The only difference between these two constructs is the addition of two crosslinking sites in each of the terminal domains of GC3-

A. These additions see a dramatic increase in crosslinking efficiency and crosslinks per monomer relative to GC2-A from  $<1$  to  $\sim 4$ . This contrasts with the addition of three internal tyrosines in GC1-A which is accompanied with an increase on average of only 1 crosslink per monomer. These results demonstrate that it is the addition of crosslink sites in the terminal domains which allows the most dramatic increase in crosslinks per monomer, indicating that terminal domains are the most likely to diffuse close enough to other molecules to crosslink into the network. The caveat to this conclusion is that an average number of crosslinks per monomer of  $<1$  as seen in GC2-A gels would rationally preclude the formation of N-C-terminal linked chains. We therefore hypothesize that due to the physical nature of the gel interactions a significant portion of lower molecular weight crosslinked chains are able to diffuse out of the gel during the labelling reaction. Therefore the average number of crosslinks per monomer may be higher in GC2-A than these results suggest.

### **5.3: Quantification of Hydrogel Folded Protein Fraction**

The fraction of I27 domains which existed in an unfolded state post-gelation was measured through the use of cysteine shotgun labelling with the thiol-reactive fluorescent dye 1,5-IAEDANS (IAEDANS) [110]. As described in Section 3 each I27 domain contains a single buried cysteine residue with 0 Angstrom solvent accessibility in the folded state. Conjugation of this residue to an IAEDANS molecule is therefore indicative of that individual domain becoming unfolded, with all 5 domains of a single polyprotein being individually assessed. As described in Section 2 post-gelation hydrogels were immersed in a 131:1 mols excess solution of IAEDANS : cysteine residues and labelling was allowed to proceed for 2 hours. The labelling reaction was then quenched by the removal of the labelling solution and addition of a  $\beta$ -

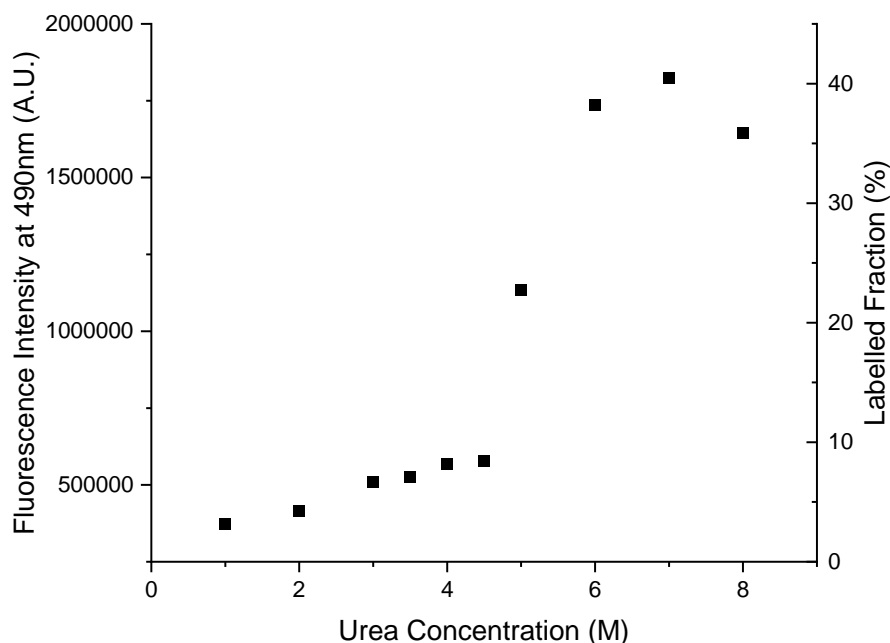
mercaptoethanol solution to scavenge the non-conjugated IAEDANS label. Trypsinisation of the gels then allowed fluorescence quantification of the IAEDANS-labelled cysteines in the sample, and back-calculations to yield an unfolded fraction.

### 5.3.1 Measuring IAEDANS Fluorescence Labelling versus Unfolded Fraction

The buried cysteine residue within each I27 domain of the polyprotein is inaccessible to the label in the folded state, and therefore quantification of the extent of cysteine labelling should reveal the total number of unfolded domains. To confirm this relationship protein hydrogels formulated to 10mg/mL were incubated with a cysteine reactive fluorophore (IAEDANS) in increasing concentrations of urea. All gels were formed in native conditions before immersion overnight in urea-containing buffer, before subsequently being immersed in an IAEDANS/urea labelling buffer for 2 hours. After quenching the labelling reactions the gels were degraded as previously described, before the fluorescence intensity at 490nm (336nm excited) was measured and corrected for volume. The resultant urea concentration versus fluorescence intensity profile (Figure 5.11) demonstrates a clear sigmoidal relationship in a fashion identical to that seen in a standard protein denaturation curve.

---

Figure 5.11: Urea concentration versus IAEDANS Fluorescence Intensity from Degraded 100mg/mL GC1-A Hydrogels



This clearly shows that the degree of IAEDANS labelling observed increases as the protein becomes destabilized. However calculation of the unfolded protein fraction suggested only ~35% unfolding in 8M urea (Figure 5.11). From these data we hypothesize that construct GC1-A exhibits 100% labelling efficiency of the material forming the crosslinked network, as indicated by the fluorescence intensity plateau above 6M urea, but that a significant proportion of the starting material is lost during the labelling experiment as it diffuses into the various immersion buffers. No construct exhibited 100% crosslinking efficiency (Figure 5.9) meaning that all gels formed from them are likely to experience the same loss of non-crosslinked material during the labelling process, making it impossible to repeatedly achieve 100% labelling in the presence of chemical denaturants. However the clear correlation between fluorescence intensity and denaturant concentration demonstrates that

quantification of the IAEDANS-labelled cysteines in the degraded hydrogel sample corresponds to the unfolded fraction of crosslinked protein in the gel network.

### **5.3.2 Quantification of Native Hydrogel Folded Fractions**

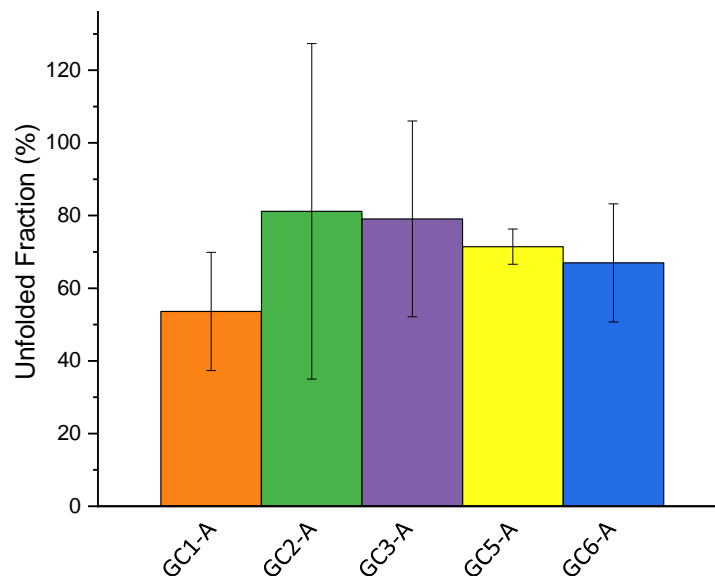
It is assumed that during the immersion of the hydrogels in labelling and quenching buffers a fraction of protein not bound to the network diffuses out of the gel into the surrounding liquid buffer and is therefore not measured in the unfolded fraction. As a result the unfolded fraction described here refers only to the protein molecules which have formed at least one chemical crosslink with the network. The thermodynamic characterization of the polyprotein constructs discussed in Section 4 show that the thermodynamic stability of all constructs is approximately equal in ambient conditions (25°C, 25mM NaPi pH7.4 buffer) with near identical secondary structures. This allows us to assume that prior to gelation all solutions exhibit an approximately equal unfolded fraction, with any differences post-gelation a consequence of differing hydrodynamic/mechanical forces incurred during gelation resulting from the different numbers of crosslinks formed by each construct.

The results summarized in Figure 5.12 show no significant difference in average unfolded fraction between any mutant constructs (GC2-, 3-, 5-, and 6-A). However a small but statistically significant difference in unfolded fraction does exist between these 4 constructs and the pseudo-wild type GC1-A. This indicates that the introduction of mutant domains has reduced the folded fraction post-gelation. This is likely due to a reduction in the mechanical stability of these mutant domains leading to more unfolding under the mechanical stresses of gelation. However this change in average unfolded fraction corresponds to an average of one additional I27 domain per monomer remaining folded in GC1-A hydrogels. An alternate explanation for this difference could be the result of lower crosslinking efficiency leading to a

greater fraction of protein diffusing away during labelling, but this would be expected to result in differences between the other constructs folded fractions for the same reason. In either case it is not anticipated to have a significant effect on hydrogel macromechanics relative to the changes in crosslinks per monomer. Mechanical differences observed between the mutant-containing constructs will not be attributable to differences in folded fraction as there are no statistically significant differences in unfolded fraction between any pair of constructs.

---

Figure 5.12: Summarised average unfolded fractions of all constructs. Error bars are standard deviation. Biological repeats n=3, technical repeats per biological repeat = >3. Significance T-Test results Appendix 1 Table 5.3.



---

## 5.4 Discussion

### 5.4.1 Unfolded Fractions

The thermodynamics of each construct discussed in Section 4 suggest that none have significantly reduced stability at room temperature in an optimal chemical environment relative to the “wild type”, meaning that significant differences in unfolded fraction in

solution between constructs was not anticipated. This pattern was observed as there were no significant differences between the unfolded fractions of hydrogels formed from different constructs. In addition there appears to be no correlation between unfolded fraction and crosslinking efficiency. The results discussed in section 4.3 regarding the continued folded state of the protein domains 45 minutes post gelation is somewhat at odds with the high unfolded fractions observed in section 5.3. The continued decline in  $\beta$ -sheet-like spectrum over time discussed in section 4.3 suggests that the high unfolded fraction  $\sim$ 2 hours post-gelation may be a result of the continued long-timescale unfolding of domains under global network stress. Therefore the results presented in section 5.3 represent the endpoint of the network evolution indicated over a 45 minute timescale in section 4.3.

#### 5.4.2 Crosslinking

Constructs exhibiting drastically different crosslink geometries (distribution and number of tyrosines in the monomer) yielded significantly differing crosslink efficiencies. These translated to even more significant differences in the average number of crosslinks formed per monomer of polyprotein. The results discussed in Section 5.2 prove that alterations in crosslinking geometry and density can be used to alter the average number of crosslinks per monomer and consequently the network topology of folded protein hydrogels. Furthermore we hypothesise that the addition/subtraction of crosslink sites from the terminal domain of polyprotein monomers has the greatest effect on crosslinking efficiency. This is evidence that the terminal domains are the most likely to bind into the network, and that steric hindrances make the rational design of efficient internal crosslinking sites more challenging. Alterations in the total number of crosslinks in the network and the addition of branching nodes such as the ones which must exist as the terminal domains of GC3-A will alter global network

topology. We hypothesize that internal domains lacking crosslinking sites can be used as spacers to tune the efficiency of other internal domains which do, but that more than one spacer domain between crosslinkable domains is necessary to achieve this in the case of 127 polyproteins. It is likely however that increasing the distance between domains and the persistence length of the molecule by increasing the length of the linker regions will allow a similar level of control over internal crosslinking site efficiency. The greater flexibility will likely allow more diffusion to within the reaction distance of internal crosslinking sites. This ability to rationally alter the network topology of a hydrogel network by changing the pattern of the crosslinks without otherwise altering the nature of the monomer building block provides a facile way to tune hydrogel mechanics. The relationship between changes to the microscale network and hydrogel macromechanics is discussed in Section 6.

During resuspension of lyophilized protein the surface area : volume ratio of the liquid resuspension volume and the resuspension container area can cause protein unfolding and aggregation due to the propensity of protein molecules to rearrange at the air-water interface, causing mechanical unfolding and aggregation [253]. Variations in concentration of protein during resuspension is likely to influence the magnitude of this effect. Visually this manifested itself as a discernable difference in the opacity of solution prior to final centrifugation before the concentration was measured. Due to the low-density/high volume nature of the lyophilized protein accurate and consistent weighing of protein into tubes was impossible making it difficult to prevent variations in resuspension quality. An in-depth study of lyophilization formulation to produce a more consistent and granular product would likely help to ameliorate this problem and improve consistency.



## Section 6: Rheological Characterisation of I27 Polyprotein Hydrogels

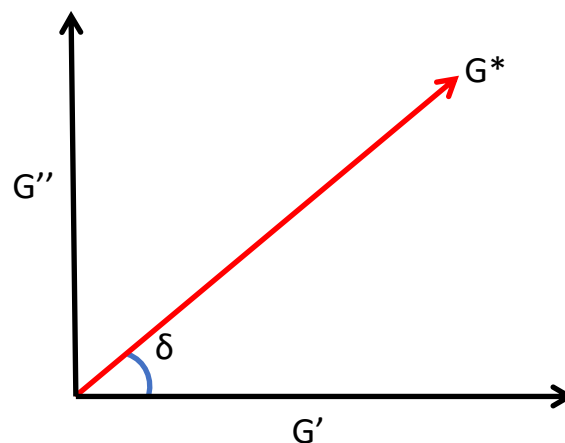
### 6.1 Introduction

The aim of this study is to characterise the relationship between the microscopic network topology and macroscopic mechanical properties of folded protein hydrogels. The microscale network topology of I27 polyprotein hydrogels was tuned as described in Section 5 by the alteration of crosslink geometry and density. Macroscale hydrogel mechanics were investigated by rheology, and are described in this Section.

All hydrogels are viscoelastic, defined as materials which exhibit both fluid like (viscous) and solid like (elastic) behaviour [254]. This can be described by the relationship between three parameters; the storage modulus ( $G'$ ), loss modulus ( $G''$ ) and the phase angle ( $\delta$ ). Together these constitute the complex modulus ( $G^*$ ), and the geometric relationship between them is shown in Figure 6.1. All these parameters are to some extent frequency dependent.

---

Figure 6.1: Geometric relationship between phase angle ( $\delta$ ),  $G''$ ,  $G'$ , and  $G^*$ .



---

Using trigonometry we can calculate the  $G'$  and  $G''$  moduli from the  $G^*$  and  $\delta$ :  $G' = G^* \cos \delta$ ,  $G'' = G^* \sin \delta$ ,  $G^* = \sqrt{G'^2 + G''^2}$ , and  $\tan \delta = G''/G'$ .

These moduli together are sufficient to describe the nature of the material, though not its full range of viscoelastic responses [186], [190], [255]. Small amplitude oscillatory tests (SAOT) are used to extract these parameters as the sample is strained around its equilibrium position in an oscillating pattern. One full oscillation is defined as a 360° rotation of the geometry probehead, or a 2π radian revolution. A full oscillation does not necessarily correspond to a particular strain in the sample, and so a target strain or stress is requested from the motor at a certain frequency, and the degree of oscillation is determined by the distance required to achieve the set stress/strain. The amplitude of the experiment is therefore expressed as maximum stress/frequency, or number of oscillations/second. The ratio of applied stress to the strain achieved is the complex modulus ( $G^*$ ).  $G^*$  is a quantitative measure of a materials stiffness, i.e. its resistance to deformation, and is expressed in Equation 11 [190] where  $\sigma$  is stress and  $\gamma$  is strain :

$$G^* = \sigma_{max}/\gamma_{max}$$

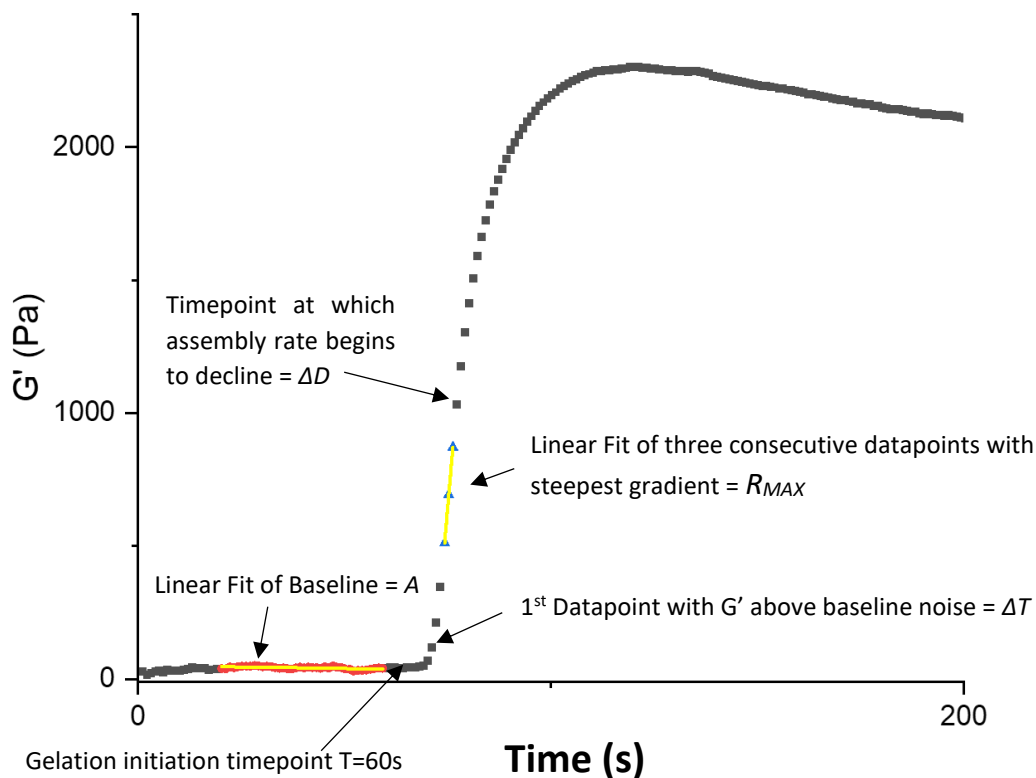
The simplest SAOT experiment is a timesweep were  $G^*$  is tracked over time in order to measure changes in elastic or viscous contribution to behaviour as a function of time or in response to environmental stimuli.

### 6.1.1 Small Amplitude Oscillatory Tests

An initial timesweep measurement was performed to measure the evolution of the hydrogels storage modulus during network formation and subsequent relaxation. This measurement allowed the extraction of several kinetic network assembly and relaxation parameters. Firstly the lag time ( $\Delta T$ ) was extracted by fitting a linear baseline ( $A$ ) through datapoints 20 seconds to 60 seconds in order to prevent early timepoint instrument settling confounding the baseline fit. After the 60 second gelation initiation timepoint each datapoint was fitted to the

extrapolated baseline. Once a datapoint reached a  $G'$  value  $>2$  standard deviations above the baseline this timepoint was defined as the start of measurable network assembly. This timepoint minus 60 seconds was defined as the  $\Delta T$  (s) between crosslinking initiation and the start of macroscale network formation. After the  $\Delta T$  datapoint each subsequent moving 3 datapoints were fitted to a linear model and the gradient calculated. The steepest single 3-point gradient calculated was defined as the maximum assembly rate ( $R_{MAX}$  (Pa/s)). Once 3 consecutive 3-point fittings after the  $R_{MAX}$  showed a decline in gradient the earliest datapoint in the first decline gradient fitting was defined as the decline time ( $\Delta D$  (s)) at which the rate of network formation begins to slow. This is summarised in Figure 6.2.

Figure 6.2: Example Gelation Curve with Assembly Kinetic Parameters Highlighted

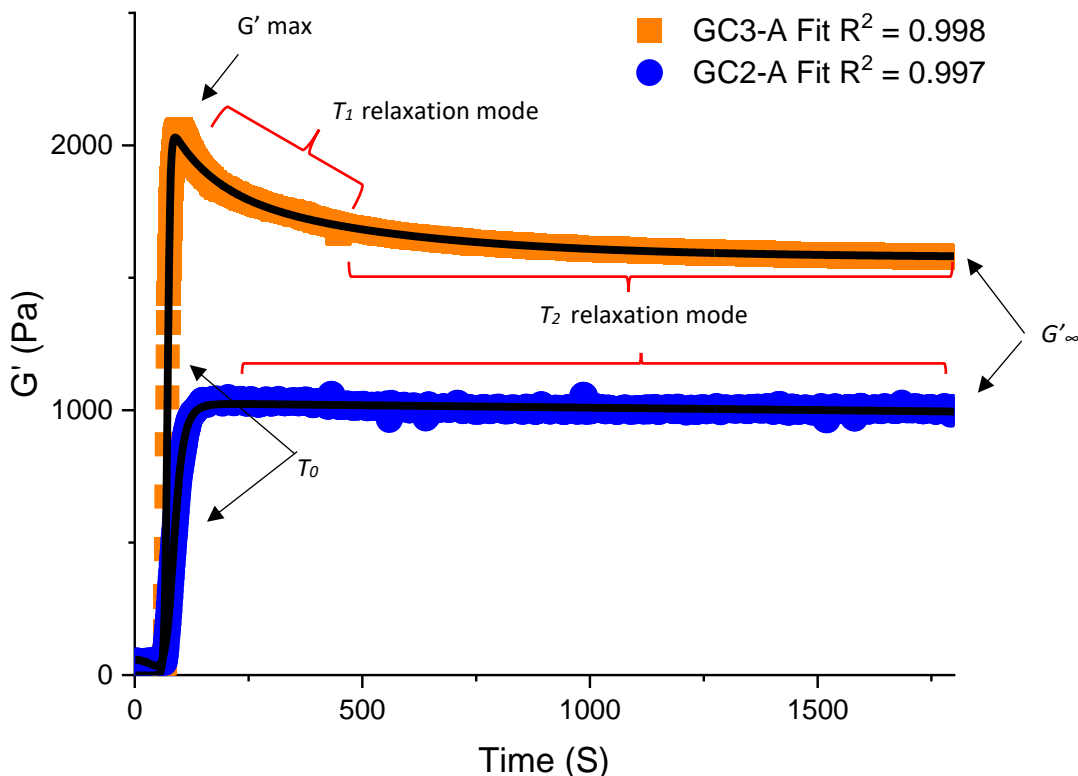


$G'$  evolution followed a sigmoidal regimen as after  $\Delta D$   $G'$  increase underwent a continuous decline in rate until a steady-state linear plateau was reached and  $G'$  either remained constant or declined at a steady rate. Some hydrogel constructs exhibited almost no  $G'$  decline post-crosslinking, whilst others showed significant declines in  $G'$  over two significantly different timescales;  $t_1$  which has been hypothesised to correspond to network rearrangement, and  $t_2$  which is attributed to protein unfolding [249]. Relaxation time constant kinetics were extracted using a previously reported model [249] according to Equation 6.1:

$$\text{Equation 6.1: } G_t' = \frac{1}{(1+e^{-c(t-t_0)})} * \left( G_\infty' + B_1 e^{-\frac{t}{t_1}} + B_2 e^{-\frac{t}{t_2}} \right) + G_0'$$

Described in this equation is an initial sigmoidal fit where  $C$  is the rate of  $G'$  increase and  $t_0$  the midpoint of the maximum  $G'$  value at infinite time ( $G_\infty'$ ), which models the initial assembly up to the maximum  $G'$  value (not  $G_\infty'$ ). This is followed by two exponential terms which describe two relaxation modes of differing timescales;  $t_1$  and  $t_2$ . Most previously described model systems have exhibited these two relaxation modes but in this study certain constructs gelation curves did not contain the shorter  $t_1$  relaxation timescale. Instead certain constructs exhibited only one relaxation mode, and in these cases the fixing of one exponential variable ( $B_1$  or  $B_2$ ) to zero allowed the accurate calculation of a single time constant. This fitting process allowed the extraction of the one or more relaxation time constants to give a measure of the degree of network rearrangement exhibited by the various gel constructs, and a final storage modulus at  $T=\infty$  ( $G_\infty'$ ). These parameters are illustrated in Figure 6.3 and an example fit of two different shaped gelation curves is provided.

Figure 6.3: Example full gelation curves for GC2-A and GC3-A hydrogels with parameters illustrated. Fit  $R^2$  refers to the quality of the fit achieved during final free-fitting of the curves and is a measure of the reliability of the results obtained from them.

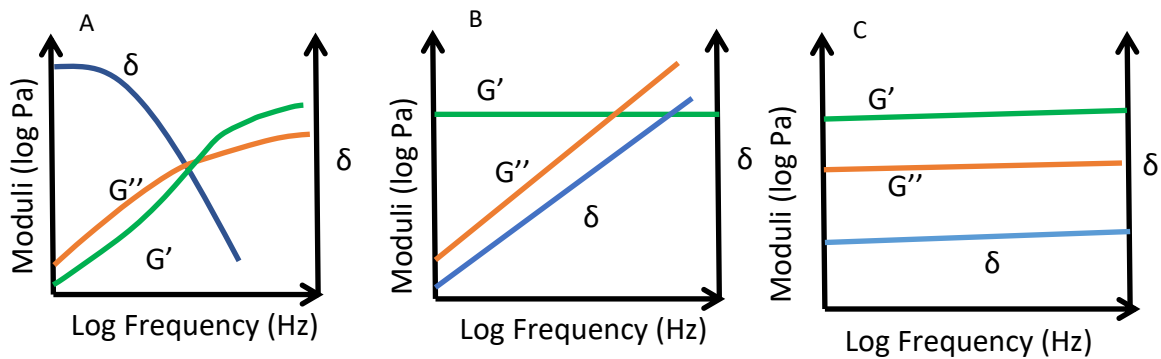


The example fit of GC3-A contains both  $t_1$  and  $t_2$  relaxation exponential functions to model the two distinct relaxation regimens present in the curve. The example fit of GC2-A was limited by fixing the  $B_1$  value to zero in order to model the single relaxation exponential present in the data. In gelation curves which exhibit only one relaxation mode this mode is always designated as  $t_2$  as they are consistently of timescales previously associated with protein unfolding and the  $t_2$  timescales of curves containing both relaxation modes.

Frequency sweep tests were performed between 0.01 and 10Hz. A 10Hz limit was selected because above this limit the viscous turbulence caused by the extreme speed of stress oscillation causes a wave effect which can manifest itself in an apparent dramatic increase in  $G'$ . This makes data  $>10\text{Hz}$  generally unreliable [219]. All hydrogels exhibited shallow gradient

increases in  $G'$  as frequency of oscillation was increased. The shape of the curve produced on a Log-Log plot of frequency versus  $G'$ ,  $G''$ , and  $\delta$  allows easy determination between viscoelastic liquids, solids, and gels (Figure 6.4 [190]). In the case of gel-like materials, at low frequencies the timescale of network rearrangement is slow and increases in line with frequency. At higher frequencies the rate of counter-force application and opposite direction deformation exceeds the timescale of network rearrangement leading to shear thickening and a gradual increase in  $G'$  [256]–[259]. The linear gradient of  $G'$  increase across frequencies is therefore a measurement of the materials timescale of network rearrangement with steeper gradients indicating slower rearrangement with greater viscous retardation and shallower indicating faster. This is independent of total resistance to flow and therefore information can be drawn independently of a materials final storage modulus. [186][260].

Figure 6.4: Example Frequency response profiles for; A- a viscoelastic liquid, B- a viscoelastic solid, C- a gel-like material. Adapted from Malvern Panalytical White Paper “A Basic Introduction to Rheology”.



Viscoelastic fluids are generally non-Newtonian meaning that at high frequencies they undergo a phase transition as  $G'$  exceeds  $G''$  (Figure 6.4A). Viscoelastic solids undergo an increase in phase transition as frequency increases as the total energy input exceeds the elastic properties of the material leading to crossover point at which the total energy in the system causes melting of the material into a fluid (Figure 6.4B).

### 6.1.2 Recovery and Relaxation

By definition viscoelastic materials display elastic recovery. The rate and degree to which recovery can take place in both the continued application and after the removal of stress can be indicative of internal network structure as a function of the contribution of the viscous or elastic moduli. Simply put a material with a dominant complex viscosity ( $\eta^*$ ) will be unable to recover its initial state after stress, whilst one with a dominant  $G^*$  modulus will be able to recover, but only if the yield stress has not been exceeded. Real viscoelastic materials sit somewhere in between, and are usually able to fully recover up to a certain sub-yield stress, and may exhibit a time lag between removal of stress and recovery as the elastic recovery is retarded by the viscous component [190], [203].

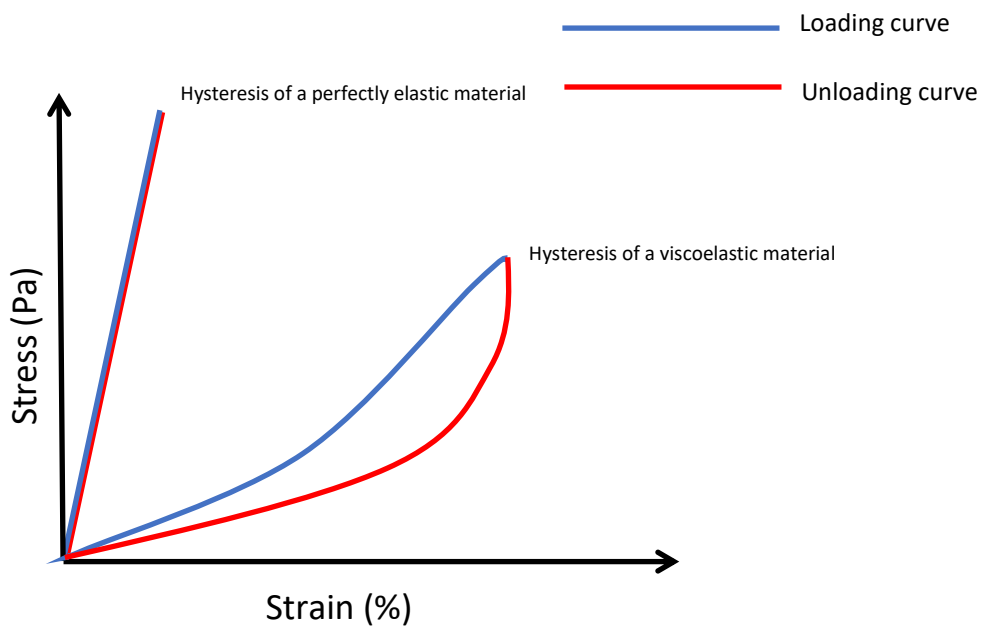
During this study all hydrogels were subjected to a stress-strain ramp test using a dynamic strain controlled rheometer (Section 2.1.16) during which the material was strained up to a target strain at a set rate, and the resultant stress at each increment of strain was measured. This constitutes the material loading curve. Once the pre-set maximum strain is reached the strain is reduced by identical increments back down to zero. In a perfectly elastic material all the energy put into the system is re-emitted at an identical rate, meaning that on the relaxation curve the same stress induces exactly the same strain as the loading curve. As a result both the loading and relaxation curves should overlay perfectly, as no energy has been dissipated during the stress-strain cycle and no permanent network damage has been done. A perfect fluid has no capacity to recover its initial state after stress, and as a result all energy put into the system is dissipated mostly as heat [261]. Viscoelastic materials exhibit behaviour between these two extremes called hysteresis [190], [262]. Essentially the loading and relaxation curves of a viscoelastic material do not overlap perfectly, but will yield close to

identical maximum and minimum strains. This creates a loop, the area of which can be used to calculate the total energy dissipation during the stress-relaxation cycle (Figure 6.5) [187]. The area of the hysteresis loop/100 is equal to total energy dissipated on unloading in J/m<sup>3</sup> [261]. The efficiency of energy recovery can also be calculated using Equation 6.2 where A is loop area, and  $Ef$  is efficiency (%):

$$Ef = \left( 1 - \left[ \frac{A}{\sigma_{max} * \gamma_{max}} \right] \right) * 100$$

The more a materials complex modulus is dominated by viscosity the lower efficiency will become due to a greater dissipation of energy by the viscous component [263].

Figure 6.5: example hysteresis loop of a viscoelastic material generated by a stress-strain cycle.



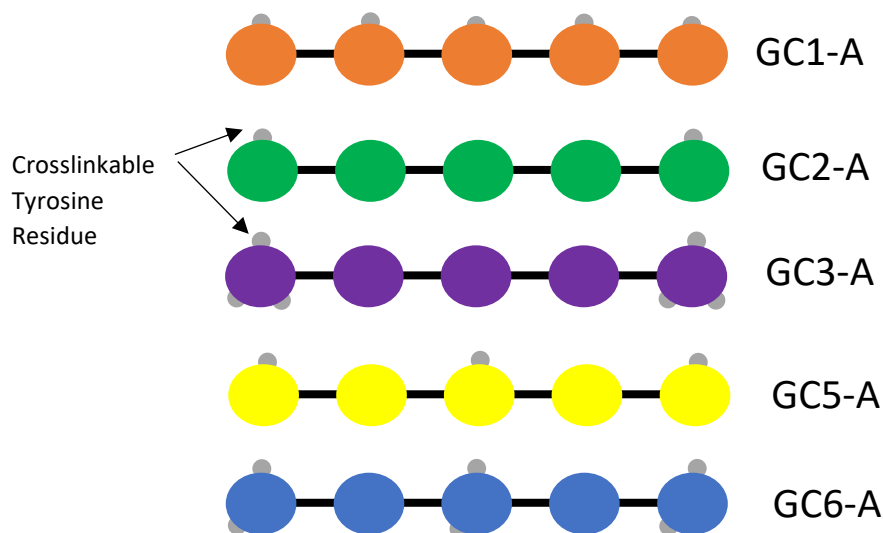


## 6.2 Gelation Analysis

As described in Section 6.1.1 gelation kinetics were measured by timesweep measurement. All data presented is an average of 3 biological repeats each consisting of >3 technical repeats. Errors are standard deviation calculated across all samples of each hydrogel species. The hydrogel solutions were measured for 60 seconds prior to crosslinking initiation (lamp on) to allow the instrument to settle and a 40 second baseline to be fitted to the data. At 60 seconds the lamp (Section 2.1.24) was switched on for 300 seconds and during this time all samples underwent a sol-gel transition. Once the lamp was turned off samples were continuously measured for 25 minutes to observe post-crosslinking relaxation behaviours. Kinetics were analysed as described in Section 6.1.1 to extract the  $\Delta T$ ,  $\Delta D$ , and  $R_{MAX}$  values.

---

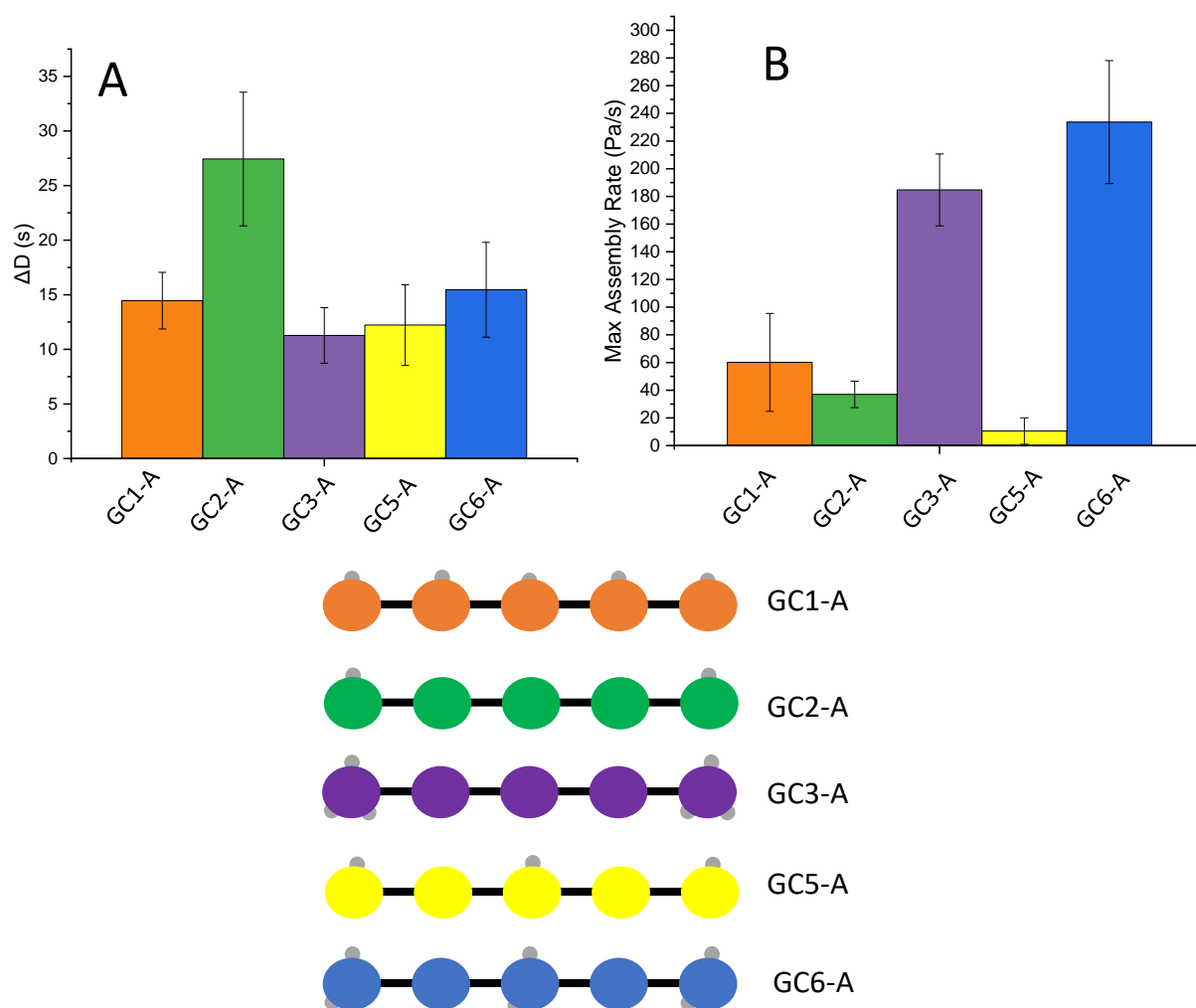
Figure 6.6: Schematic representation of the number and distribution of tyrosine residues throughout each polyprotein construct discussed in this Section.



## 6.2.1 Network Assembly

The variation in  $\Delta D$ , and  $R_{MAX}$  values between hydrogel construct species are summarised in Figure 6.7. There was no significant variation in  $\Delta T$  between any constructs.

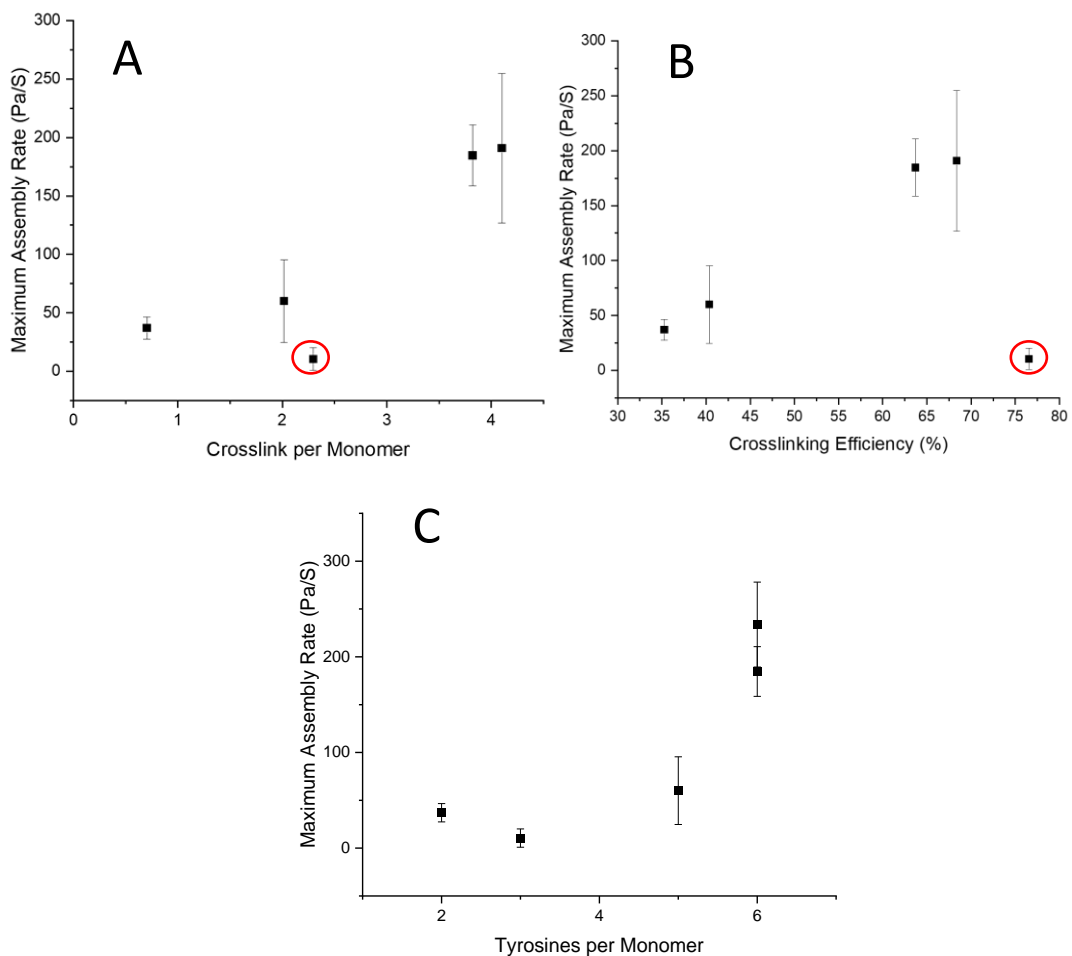
Figure 6.7: A; Summarised Decline Start times ( $\Delta D$ ) with standard deviation. B; Summarised maximum assembly rates ( $R_{MAX}$ ) with standard deviation.



All constructs exhibit near-identical  $\Delta D$ 's (Figure 6.6A). This may suggest that the rate of crosslink formation-dominated network assembly slows in response to factors independent of crosslink geometry or number. GC2-A has a significantly later  $\Delta D$  and is discussed in detail in Section 6.4.1.

$R_{MAX}$  (Figure 6.7B) correlates with both crosslinking efficiency and crosslinks per monomer (Figures 6.8A and B) suggesting that the rate of network assembly is driven by the rate of crosslink formation. However there is no correlation with the number of tyrosines per monomer (Figure 6.8C). This indicates that crosslink site geometry can alter the efficiency of crosslink formation, and that this in turn will alter the  $R_{MAX}$ .

Figure 6.8: A; Crosslinks per Monomer versus  $R_{MAX}$ , B; Crosslinking Efficiency versus  $R_{MAX}$ , C; Tyrosines per Monomer versus  $R_{MAX}$ . Datapoints in red circles indicate GC5-A.



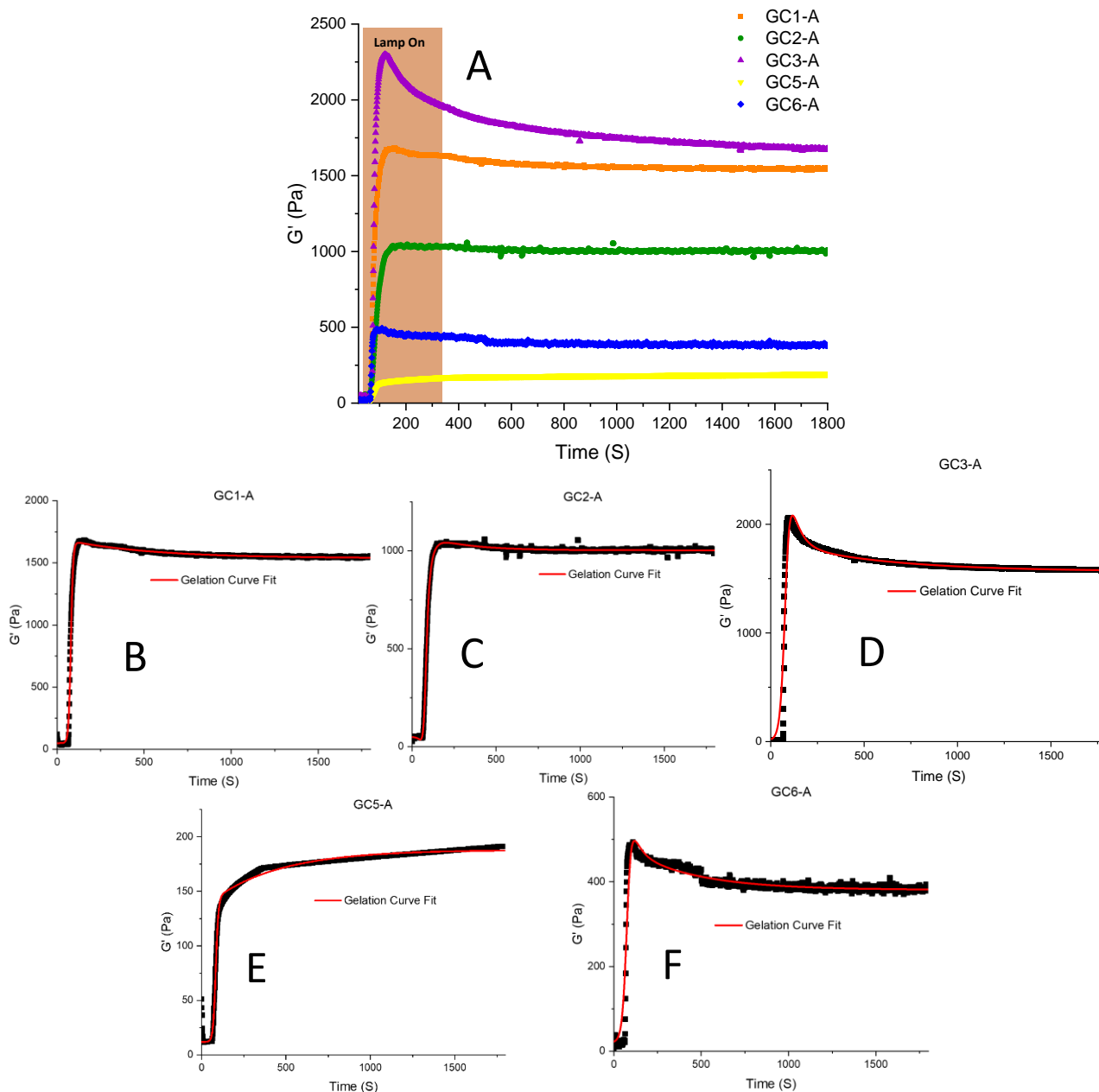
Interestingly there is an exception to this rule; GC5-A which has a high crosslinking efficiency and number of crosslinks per monomer but sits as an outlier on both graphs in Figures 6.8A and B highlighted in red circles. This may be indicative of an interesting result of a change in crosslink geometry. GC2-A contains one crosslink site per terminal domain (total of 2) and is

able to form a hydrogel with a  $G'_{\infty}$  of  $1087 \pm 158$  Pa and has an  $R_{MAX}$   $37.0 \pm 9.6$  Pa/S. The addition of a single crosslink site in the central domain as seen in GC5-A drastically reduces the  $G'_{\infty}$  to  $432 \pm 250$  Pa and the  $R_{MAX}$  to  $10.5 \pm 9.6$  Pa/S. This suggests that a single additional crosslink site has reduced the ability of the construct to form a stiff network not by reducing the number of crosslinks (which increases) as seen in other construct examples but purely by geometry most likely altering the network topology. Potential mechanisms for this are discussed in Section 6.4.

### 6.2.2 Network Relaxation

During gelation a transition from a steep gradient increase to a steady-state plateau is observed. This is accompanied by a degree of network relaxation seen as a decline in  $G'$  from a peak maximum. The gelation curves of different constructs exhibit drastically different relaxation behaviours, with representative examples shown in Figure 6.9 along with example fittings according to Equation 6.1 [249].

Figure 6.9: A; Representative example gelation curves of each hydrogel construct. B-F; Example gelation curves with relaxation behaviour fit lines shown in red.



Broadly speaking there are two distinct relaxation regimes seen; GC2-, 3-, and 6-A demonstrate two timescales of relaxation, whilst GC1- and 5-A have only one. This was determined by iterative fitting of individual gelation curves until an R-Squared of  $>0.99$  was reached. These fittings yielded one of two results; either  $t_1$  and  $t_2$  were calculated as two numbers an order of magnitude apart indicating a true separation of timescales and therefore

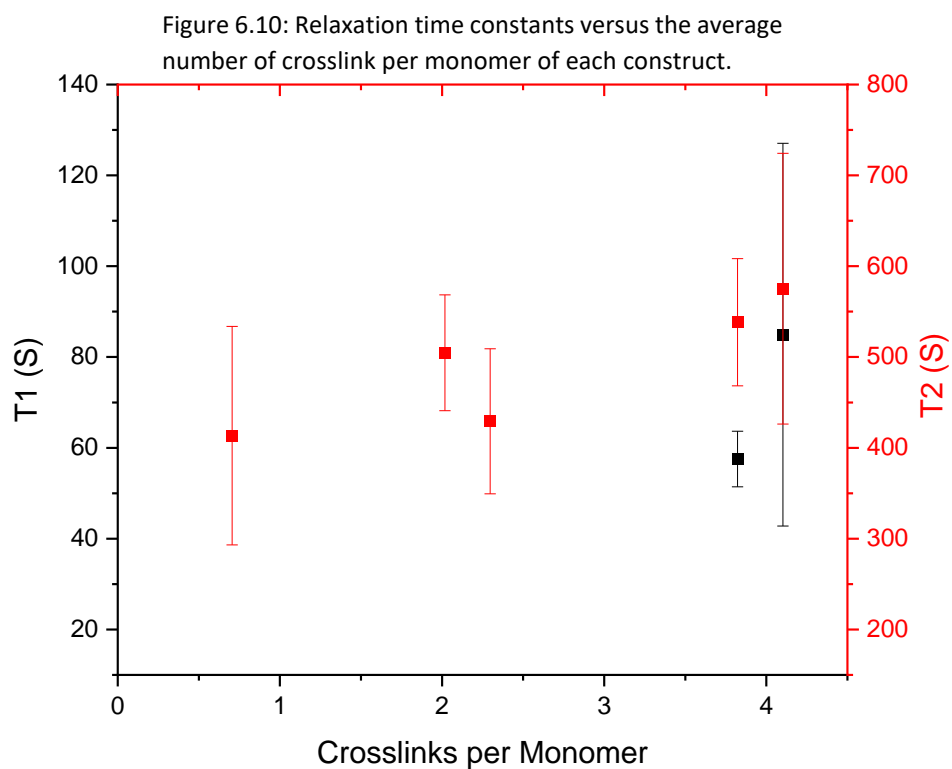
two relaxation modes. Alternatively both  $t_1$  and  $t_2$  were calculated to be either the same number or one was calculated to be extremely small and equivalent to zero. This indicated that the fitting function was only able to achieve a high-accuracy fit by merging both time constants into one by splitting it in half between both exponential functions or by making one zero. In these cases the  $B_2$  value was fixed to zero and iterative fitting re-performed to yield an R-squared fitting  $>0.99$  with a single time constant calculated. The time constant values calculated in these cases were invariably of the order of magnitude associated with protein unfolding and the  $t_2$  timescales previously reported, and so were designated as such. All reported fits achieved an R-squared  $>0.99$ . All fitting results are summarised in Table 6.1 and correspond to the combined average and standard deviations of 3 biological repeat datasets each comprising  $\geq 3$  technical repeats.

	GC1-A	GC2-A	GC3-A	GC5-A	GC6-A
$t_1$ (s)	N/A	N/A	57.5 $\pm$ 6.1	N/A	84.9 $\pm$ 42.2
$t_2$ (s)	504.7 $\pm$ 63.8	413.4 $\pm$ 120.3	538.4 $\pm$ 70.0	429.3 $\pm$ 79.9	575.2 $\pm$ 149.0
$B_1$	N/A	N/A	386.4 $\pm$ 263.5	N/A	81.2 $\pm$ 57.4
$B_2$	107.2 $\pm$ 70.2	55.1 $\pm$ 7.5	261.4 $\pm$ 64.0	-27.1 $\pm$ 34.1	111.2 $\pm$ 64
$B_1 + B_2$	107.2 $\pm$ 70.2	55.1 $\pm$ 7.5	539.8 $\pm$ 322.7	-27.1 $\pm$ 34.1	192.3 $\pm$ 8.2
$G'_{max} - G'_{\infty}$	128.9 $\pm$ 64.8	65.7 $\pm$ 24.4	560.9 $\pm$ 170.2	-13.8 $\pm$ 38.4	170.2 $\pm$ 26.1
$G'_{\infty}$ (Pa)	1368.0 $\pm$ 200.9	1087.2 $\pm$ 157.8	1538.4 $\pm$ 161.0	432.5 $\pm$ 260.5	735.3 $\pm$ 344.5

In cases where an increase in  $G'$  occurs after 360s (lamp off) the  $B_1$  and  $B_2$  values become negative as this reflects a gain in  $G'$  rather than a relaxation as normally occurs, as

demonstrated by GC5--A. Time constants are still reflective of this timescale of network evolution.  $B_1$  and  $B_2$  are functions of the total loss/gain of  $G'$  (network evolution) therefore their sum should be approximately equal to the difference between  $G'$  max and  $G'_{\infty}$ .

The time constants extracted show significant variation suggesting that the rate of network relaxation is altered by network crosslink density. This is illustrated in Figure 6.10 which shows a correlation between an increase in  $t_2$  and the number of crosslinks per monomer.



From this we can surmise that as the global network crosslink density increases there is an accompanying retardation of network rearrangement and rate of protein unfolding. This may be a result of the reduction in the ability of the network subunits/branches/clusters to move past each other and rearrange as they are more tightly tethered in place with more points of attachment, and an increase in global enthalpy. This is discussed in detail in Section 6.4. A

final interesting point is that the order of magnitude for  $t_1$  and  $t_2$  in a previously reported globular Maltose Binding Protein hydrogel model system was reported as 100's and 1000's of seconds respectively, whereas in this polyprotein system both are one order of magnitude faster. This may suggest that the change in monomer molecule shape from spherical (MBP) to rod-like (I27 pentamer) has an effect on the rate of network relaxation with rod-like molecules able to more easily to move past each other and therefore relax faster.

### 6.3 Macromechanics

All samples were subjected to both timesweep and frequency sweep SAOT measurements before being subjected to a stress-strain ramp.

#### 6.3.1 Final Storage Moduli ( $G'_{\infty}$ ) and Frequency Response

All hydrogel species exhibit a decline from a  $G'$  maximum achieved during primary gelation into a steady state  $G'$  decline (Section 6.2.2). Therefore when comparing the storage moduli of gels it is the final storage modulus at infinite time which should be used to define and compare the stiffness of FGPs as this will give the theoretical final steady state gel stiffness independently of differences in the  $t_2$  relaxation time constant. This parameter is  $G'_{\infty}$  and was extracted during the gelation curve fitting process described in Sections 6.1.2 and 6.2.2. This was not measured experimentally due to the potentially extraordinarily long periods of time that would be required to reach a true zero-gradient steady state. Some gelation curves to the naked eye appeared to reach a steady state indicating relaxation was complete and therefore the final experimentally recorded  $G'$  value was near-identical to the calculated  $G'_{\infty}$



but in these cases the  $G'_{\infty}$  is still reported. In addition despite the use of a low-viscosity silicone oil to surround the edge of the gel and prevent evaporation, over long timescales the gradual rate of gel dehydration would likely prevent a zero-gradient steady state being achieved. The  $G'_{\infty}$  values of each hydrogel species are summarised in Figure 6.11B with standard deviation error, and in Table 6.1 (Section 6.2.2).

Figure 6.11: A; Final Storage Moduli versus average Crosslinks per Monomer, B; Summarised Storage Moduli for each construct, C; Final Storage Moduli versus Tyrosines per Monomer.

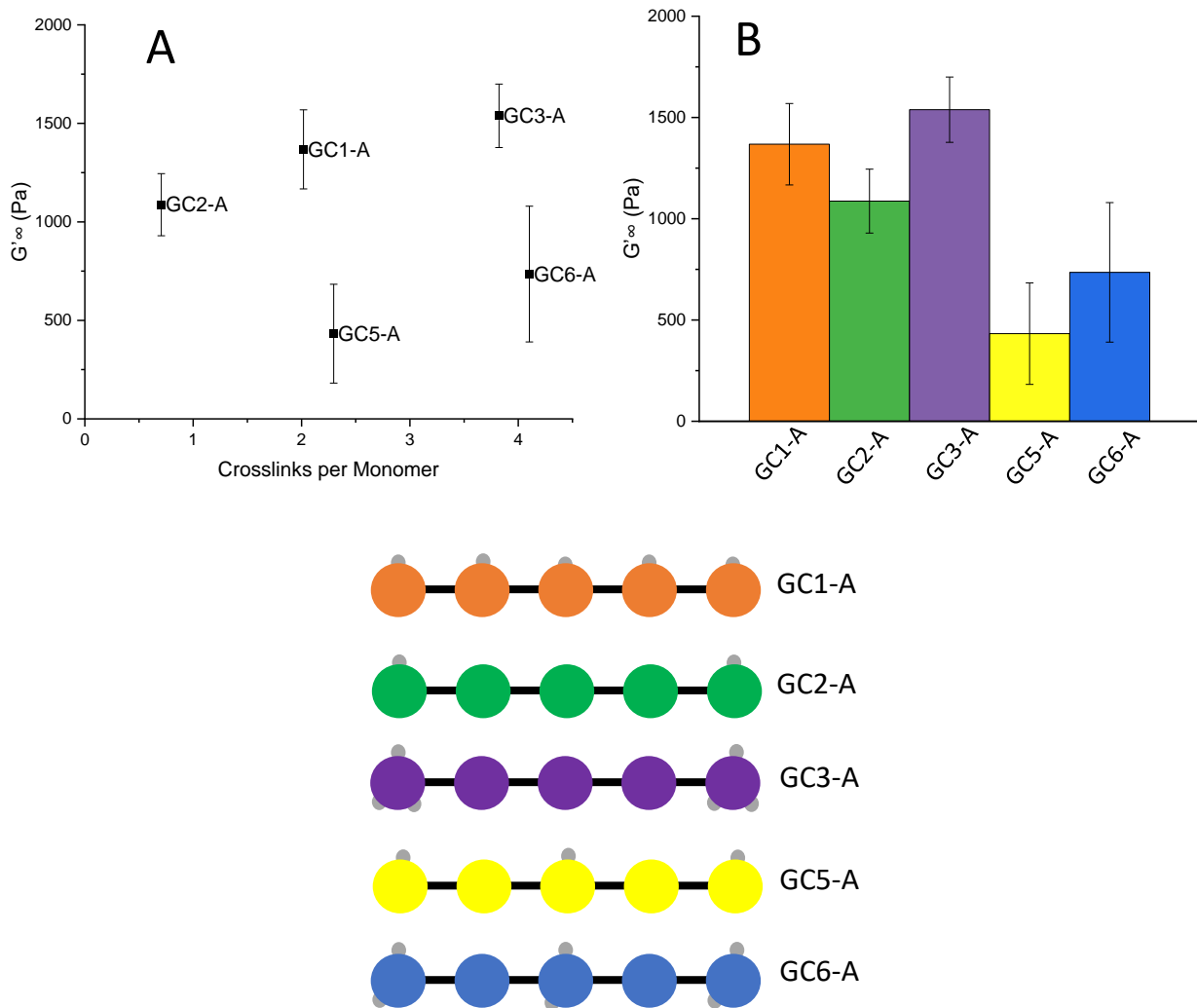
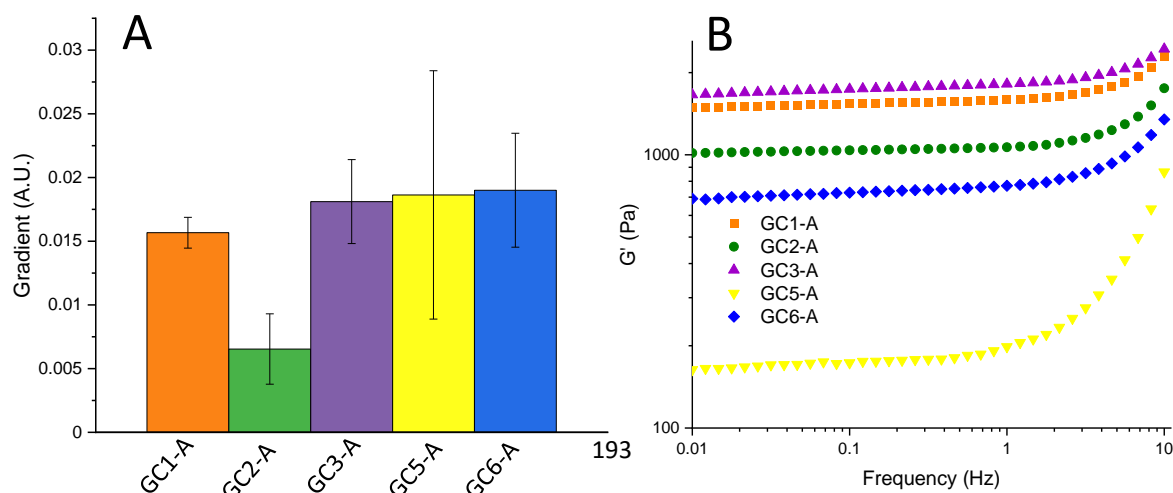


Figure 6.11B shows significant variation in the  $G'_{\infty}$  values of different hydrogel species. Most markedly between GC5-A and GC6-A and the other three. There is also a statistically significant difference between GC2-A and GC1- and 3-A, whilst GC1-A and GC3-A appear to have approximately the same  $G'_{\infty}$ . Figure 6.11A shows no correlation between the number of crosslinks per monomer, with two examples of gels exhibiting an ~equal number of crosslinks per monomer but significantly different  $G'_{\infty}$ 's. This may suggest that it is the underlying network topology determined by crosslink geometry which is the biggest determinant of  $G'_{\infty}$  with interactions other than chemical crosslinks also contributing to gel stiffness. Potential mechanisms for this variation are discussed in Section 6.4.2.

The frequency response of each hydrogel construct was measured as described in Section 6.11. Frequency- $G'$  gradients were fitted through data between 0.01 and 1 Hz because of the appearance of viscous wave-related  $G'$  increases (Section 6.1.1) above 1 Hz demonstrated in Figure 6.12B. Figure 6.12A shows that no significant differences in average gradient exist between hydrogel species with the exception of GC2-A versus all other constructs. This lack of variation suggests that in the frequency range of 0.01-1 Hz all gels except GC2-A have an approximately equal rate of network rearrangement.

Figure 6.12: A; Summarised average gradient of  $G'$  increase between 0.01-1Hz frequency of each construct. B; Example frequency sweep plots of each construct.

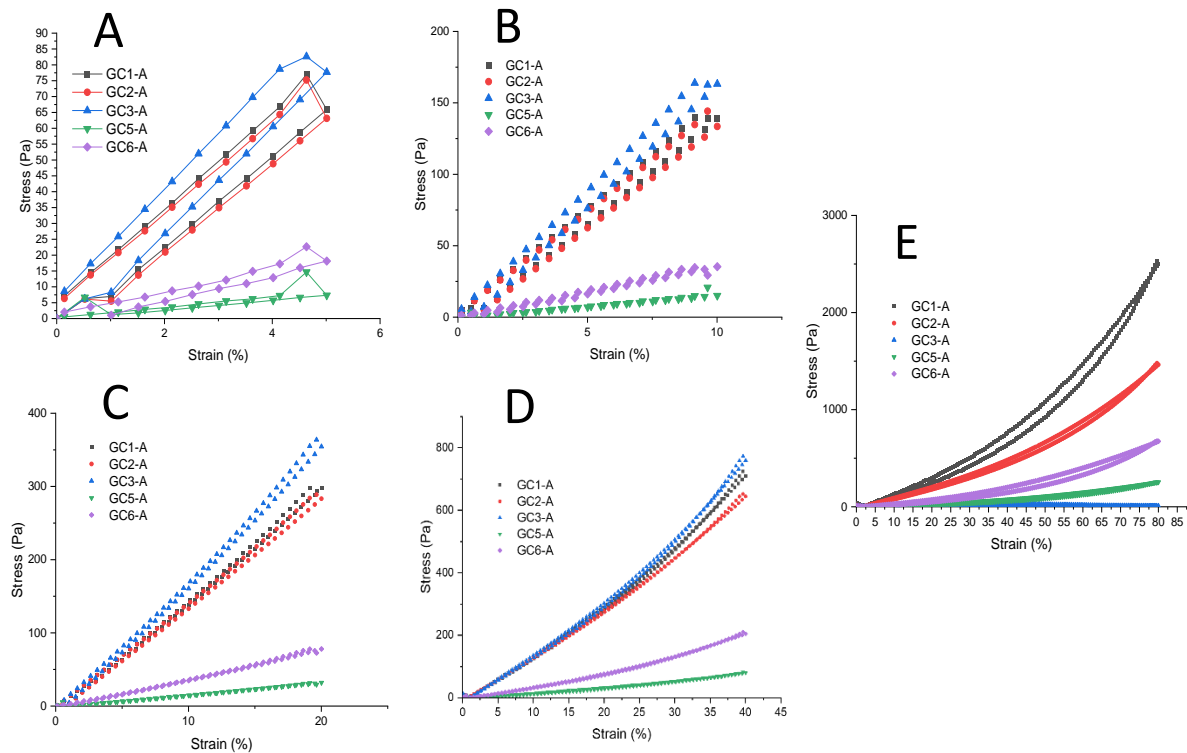


The dramatically shallower gradient seen in GC2-A hydrogels (Figure 6.12A) suggests a significantly faster rate of network rearrangement reducing the frequency dependence of the material [264]. This supports the view of a reliance on physical entanglements of GC2-A gels and is discussed in Section 6.4.2.

### 6.3.2 Hysteresis

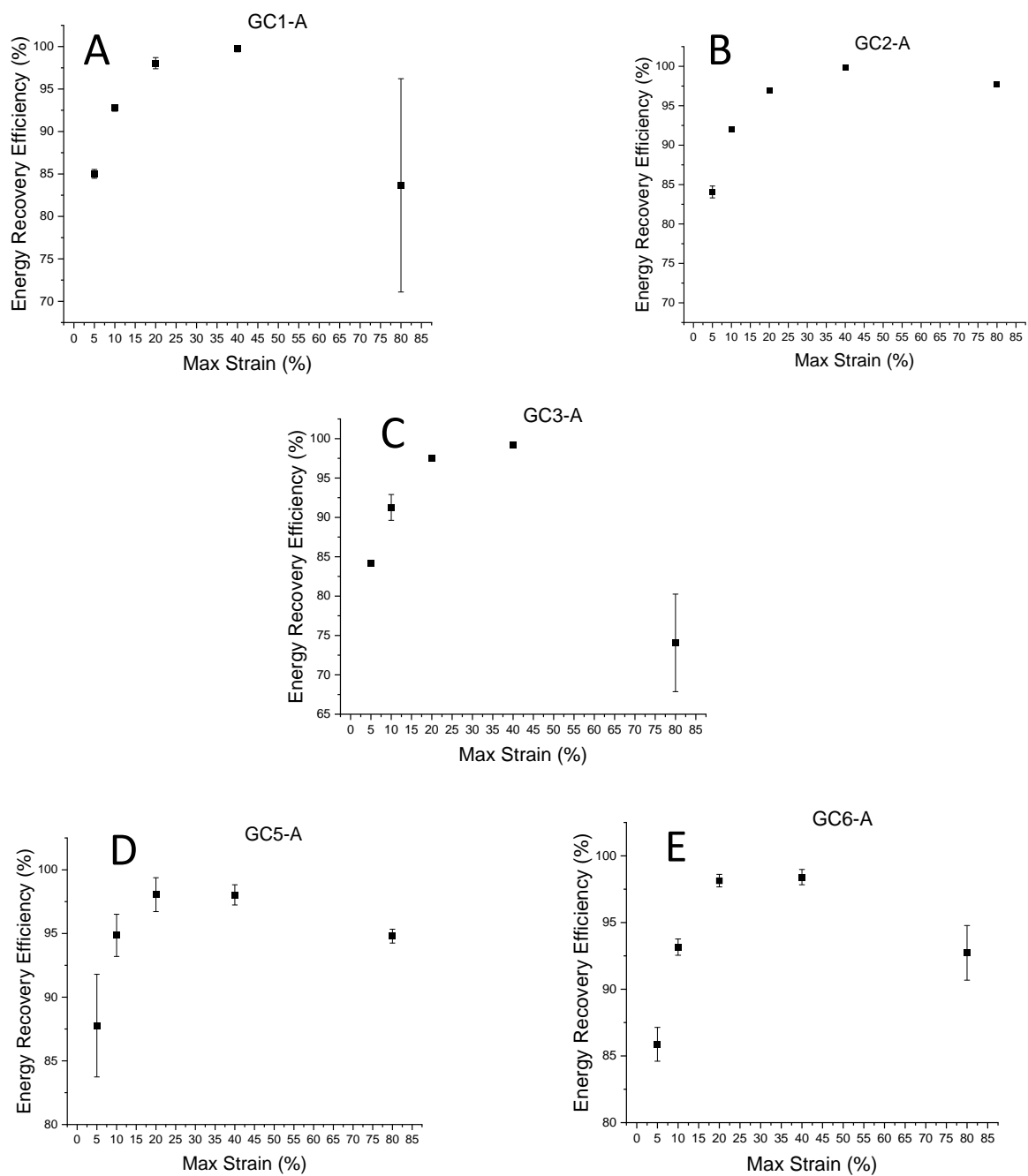
All samples were subjected to a stress-strain ramp from 5% to 80% strain. The areas of the hysteresis loops were then used to calculate the efficiency of energy recovery of each gel according to Equation 6.2. Representative hysteresis loops for each construct across each maximum strain cycle are shown in Figure 6.13. All hysteresis data is averaged across 3 biological repeats each comprising  $\geq 2$  technical repeats.

Figure 6.13: Representative hysteresis loops for all constructs at A; 5% maximum strain, B; 10% maximum strain, C; 20% maximum strain, D; 40% maximum strain, E; 80% maximum strain.



When average energy recovery efficiency is related to the maximum strain reached during each stress-strain cycle, a trend towards higher efficiency at higher strain is observed for all gels. This is followed by a dramatic decline indicating a breakdown in the network at 40/80% strain depending on hydrogel species. This is illustrated in Figure 6.14.

Figure 6.14: A-E; Energy recovery efficiency of each hydrogel construct as a function of the maximum induced strain. F-J; Summarised energy recovery efficiencies for all constructs at each maximum strain with standard deviation.

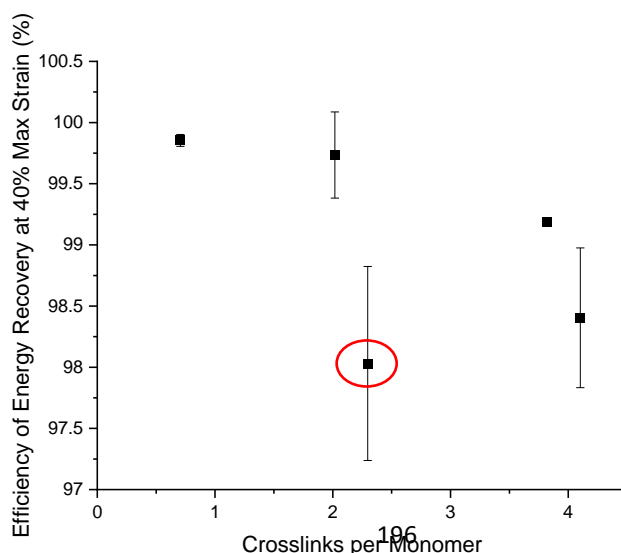


An increase in efficiency in line with strain is an interesting observation from this data. Efficiency is generally believed to be a measure of elasticity, with a material stressed below its elastic limit theoretically yielding 100% elastic recovery efficiency. Therefore we would predict that as maximum strain is increased the material gets closer to (or further beyond) its elastic limit and efficiency decreases. Figures 6.14 A-E show that the opposite relationship is true of these protein hydrogel systems. This is true up until 80% strain at which point efficiency decreases significantly indicating that the elastic limit has been reached and permanent network rearrangement/damage has occurred. The differences in recovery efficiency at these maximum strains are therefore indicative of the gels different elastic limits and potentially their network topologies [265].

Figures 6.14 F-J show that at 5, 10, and 20% maximum strain there is no significant difference in the energy recovery efficiency of the hydrogel constructs. However at 40% strain we see significant differences emerging. The efficiency of energy recovery at 40% strain correlates with the number of crosslinks per monomer with the outlier of GC5-A once again (Figure 6.16). This further illustrates the potentially unique network topology of GC5-A hydrogels discussed in detail in Section 6.4.2.

---

Figure 6.16: Efficiency of Energy recovery versus Crosslinks per Monomer. GC5-A is highlighted in a red circle.



This trend may indicate that a greater density of crosslinks in the bulk network leads to less perfect elastic behaviour at high strains. This may be further indication that increasing the number of crosslinks per monomer reduces the ability of the network to rearrange even temporarily in response to stress.

## 6.4 Discussion

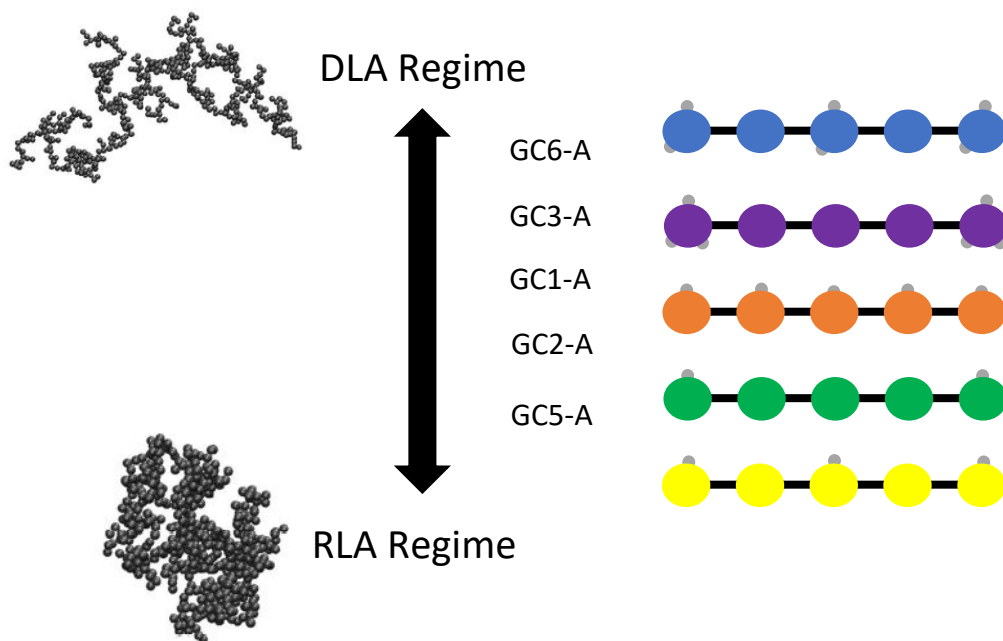
Firstly these results strongly suggest that alterations in crosslink density and geometry of the monomer building block can have a significant effect on all mechanical properties of folded protein hydrogels. Whilst there are certain parameters that can be correlated to the number of crosslinks per monomer within the gel network ( $R_{MAX}$ ,  $t_1$  and  $t_2$ , and hysteresis at high strain) it does not appear that alterations in these parameters are a direct result of changes to crosslink density but instead are symptomatic of different network topologies [266].

### 6.4.1 Assembly Kinetics

To understand how these differences come about we must start from the earliest indications of divergence between constructs; the gelation kinetics. As discussed in Section 6.2.1 the variation in  $\Delta T$  between constructs is minimal indicating that the increase in solution viscosity during early gelation occurs at a constant rate independent of crosslink geometry, and therefore the growth in size of multiple network nuclei appears to be similar across all species. However during the rapid assembly phase constructs show large differences in  $R_{MAX}$ . Figures 6.8A and B suggest that this rapid assembly is correlated to the availability of tyrosines to form crosslinks, and so from this we can hypothesise that a high/low  $R_{MAX}$  is indicative of efficient/inefficient crosslink geometry in the monomer building blocks. Network formation

theories discussed in Section 1.2.4.2 inform us that if the chance of forming a crosslink when within the reaction distance is close to 1 then the system will tend towards a DLA regime, and that as this chance gets further from 1 the system will shift towards an RLA regime. Therefore we can hypothesise that constructs with more sterically efficient crosslink sites tend towards DLA regime network topologies whilst those with sterically inaccessible crosslink sites will tend more towards RLA regime network topologies; Figures 1.12 and 1.13 [157][158]. Working from this hypothesis we can rank the constructs from most DLA-prone to most RLA-prone based on their  $R_{MAX}'S$ ;

Figure 6.17: Schematic summary of the order of constructs between DLA and RLA assembly regimes.



This would suggest that constructs which tend towards DLA will exhibit smaller clusters and a more highly branched network, and constructs tending towards RLA will have larger clusters and less branching. Where each construct sits on this spectrum will cause variation in its final network topology. By changing the crosslink geometry the dominance of RLA or DLA is altered during early stage assembly which in turn alters the dominance of DLCA or RLCA in later stage

assembly [267], [268]. This may lead to significant variations in network topology and in turn hydrogel mechanics. The  $\Delta D$ 's of each construct show no significant differences. This is unexpected but may be an indication that depending on each constructs DLA/RLA dominance during early gelation the DLCA/RLCA balance in later assembly is altered. An RLA regime will likely run out of available tyrosines to crosslink faster than a DLA regime slowing down assembly rate. Meanwhile an RLA regime will allow  $R_{MAX}$  to be maintained longer, manifesting in a common  $\Delta D$  [267]. The exception to this is GC2-A. This construct was designed to be unique; containing single crosslink sites on each of its terminal domains it was anticipated that GC2-A would be unable to form a cross-bridged crosslinked network. Instead it was anticipated to form only long chains crosslinked via their N- and C-termini with gelation caused by physical entanglements of these elongated chains. This is seen to manifest as a low  $R_{MAX}$  probably because crosslinking alone does not allow network formation and instead the retarded evolution of physical interactions afterwards causes gelation. The longer these chains become the higher the rate of physical network formation is likely to be. The lag between the crosslinking-driven formation of elongated chains and the physical network formation is therefore seen as a late  $\Delta D$  due to the utterly distinct mechanism of network growth.

The rank order above of constructs on the DLA-RLA regime spectrum next informs our rationalisation of the relaxation kinetics. As discussed in Section 6.2.2 GC3-A, GC6-A, and GC2-A all exhibit two distinct relaxation time constants, whilst GC1-A and GC5-A contain only one. If we rank the constructs on the magnitude of their  $t_1$  time constants an order emerges similar to that for DLA/RLA regime network assembly:



Construct	$T_1$ Relaxation time Constant (s)
GC6-A	84.9±42.2
GC3-A	57.5±6.1
GC2-A	24.0±3.0
GC1-A/GC5-A	N/A

For the reasons discussed above GC2-A can likely be discounted from this ranking because of its distinct gelation regime making its mechanism of network relaxation not directly comparable. With this in mind we can therefore see a clear divide between the constructs; those which we hypothesise will tend towards an DLA regime exhibit a faster  $t_1$  timescale of network rearrangement, whilst those tending further towards an RLA regime lack this short-timescale relaxation mode. If we assume that this is a reflection of their different network topologies we can hypothesise that more highly branched networks undergo a settling of the network once crosslinking has ended, whilst those with less highly branched and more cluster-dense networks do not. This may be because more densely clustered networks form within an entropic minima meaning there is no energetic need for significant rearrangement, whilst branched networks experience a greater increase in enthalpy leading to a settling downwards into an entropic minima once enthalpy-increasing crosslinking has ceased.

Whilst this interpretation could provide a useful framework for understanding the assembly of protein hydrogels, it does not fully capture all their behaviour. The key example of behaviour not captured in this study is the different crosslinking topologies resulting from the different distributions of tyrosines throughout the various polyprotein constructs. It has been possible to measure the global efficiency of crosslinking and from this we have hypothesised the possibility of terminal-domain located tyrosines being more efficient. But this is very

coarse grained and a more detailed understanding of each individual tyrosine's crosslinking efficiency is still required to give a more nuanced and detailed understanding of how to use crosslink distribution to regulate network topology. This study has captured the microscale network topology, not the nanoscale crosslinking behaviour of individual protein domains.

#### 6.4.2 Hydrogel Mechanics

Beyond this point our analysis of the data presented here becomes more speculative because there are no clear patterns between the ranking discussed so far and hydrogel mechanics. Instead we must consider individual constructs and speculate on how we believe their network topologies may affect their mechanics. Firstly the frequency dependence of all hydrogel species appears to be the same with the exception of GC2-A (Section 6.3.1). This indicates that despite alterations to network topologies and densities of crosslinking, chemically crosslinked protein hydrogels do not exhibit significant variation in frequency dependence as a result of crosslink geometry. The key conclusion we can draw from the frequency response data is that GC2-A is different and has a lower frequency dependence than the other constructs. As discussed in Section 6.1.1, a shallower gradient in a frequency sweep is indicative of a faster rate of network rearrangement and recovery, which is itself a function of elastic behaviour. This would suggest that a shallower gradient will correspond to a greater number of interactions [264], [269]. The lack of variation seen in the frequency responses of the other constructs may therefore be due to the relatively small variation in the number of interactions each monomer forms with a range of crosslinks per monomer of only 2. GC2-A averages <1 crosslink per monomer but once crosslinked into elongated chains these chains can be described as the new monomer species for actual gelation. We hypothesise that the average number of physical/ionic interactions each of these chains forms is significantly

greater than the number of crosslinks in any other gel species, possibly by more than an order of magnitude. This may be the cause of both GC2-A's high  $G'_{\infty}$  (1087.2±157.8Pa) and lower frequency-dependence as the greater number of weaker physical interactions is more elastic than a smaller number of chemical crosslinks [270]. In real terms therefore GC2-A can probably be classified as a photo-polymerised physical hydrogel, not a chemical one.

By referring back to the DLA-RLA rank order and accompanying relaxation behaviours discussed above we can attempt to rationalise the mechanics of the other constructs. We believe that both GC1-A and GC5-A tend towards an RLA regime and form on average ~2 crosslinks per monomer. However, they have drastically different  $G'_{\infty}$ 's; 1368.0±200.9Pa and 432.5±260.5Pa respectively. This suggests that despite similar crosslink densities their geometries engender different network topologies, further supported by GC1-A having a significantly greater  $R_{MAX}$  than GC5-A (60.1±35.4 Pa/s versus (10.5±9.6 Pa/s). We hypothesise that GC5-A is more dominated by DLA assembly than GC1-A leading to less crosslinking between larger clusters, and that this relative difference in cluster interconnectedness manifests in differences in stiffness. We believe that GC3-A and GC6-A both tend towards a DLA regime and form an average of ~4 crosslinks per monomer, and as a result have more highly branched networks. However GC6-A is ~half as stiff as GC3-A;  $G'_{\infty}$  GC3-A= 1538.4±161.0Pa, GC6-A= 735.3±344.5Pa. This may reflect that crosslink geometry in branched networks can cause changes in network topology just as significant as changing between RLA and DLA dominated assembly systems. We do not know if GC3-A and GC6-A hydrogels exhibit differences in average cluster size or homogeneity of network. The biggest crosslink geometry difference between the two is the presence of two crosslink sites in the third domain of GC6-A. This suggests that it is this potential internal crosslinking point which leads to the alteration in stiffness. We hypothesise that the marginally greater average number of crosslinks per

monomer formed by GC6-A ( $4.1 \pm 1.8$  versus  $3.8 \pm 1.4$ ) reflects the low-efficiency crosslinking of one central crosslink site in GC6-A. This occasional additional crosslink may cause a degree of reduction in branching on the microscale which translates to a large topological difference on the macroscale. This conjecture is relatively weakly supported but may be evidence that significant tuneability of hydrogel mechanics is possible with extremely minor alterations on the microscale.

As mentioned in Section 6.3.2 an unusual hysteresis relationship exists in all constructs between maximum strain and energy recovery efficiency, where efficiency recovery increases in line with strain despite the material getting closer to its elastic limit. There are two possible explanations for this behaviour; firstly it may be due to the low distance of displacement at low strains causing little network engagement as only the topmost layers of the material actually move. This could reduce the ratio between stress stored in the network and the stress dissipated into the top layers of viscous fluid. The second possible explanation may be related to the nature of the hydrogels complex viscosity. All hydrogels exhibit a degree of fluid (viscous) behaviour (Section 1) with viscous fluids categorised as either Newtonian or non-Newtonian [271]. Non-Newtonian fluids have a non-linear relationship between stress and strain and can either become more or less viscous as strain is increased. These results may therefore suggest that the viscous component of these hydrogels is non-Newtonian, causing an increase in the elastic component of the complex viscosity of the hydrogel, manifesting as an increase in energy recovery efficiency as strain increases.

### 6.4.3 Conclusions

We believe that these results demonstrate alterations in the crosslinking geometry of monomer building blocks effects hydrogel micromechanics. We hypothesise that variations in crosslinking efficiency as a result of sterically-limiting crosslink geometries define the balance between RLA and DLA assembly regimes. It is therefore the balance between these two regimes which alters the network topology on the macroscale, with distinct bulk topologies imbuing the material with unique mechanical characteristics. Our understanding of this relationship is imperfect and requires verification by SANS experiments to measure the network topologies of these distinct hydrogel species. However these results likely suggest that altering the crosslinking geometry on the microscale translates to predictable alterations in network assembly regime and therefore network topology and bulk material mechanics.

## Section 7: Final Conclusions and Future Directions

The aim set out at the beginning of this project was twofold; firstly to develop a robust methodology by which the microscopic network topology of protein hydrogels (crosslink density and folded fraction) could be measured, and secondly to demonstrate that alterations in these parameters correlates to alterations in macroscale hydrogel mechanics. Both of these aims have been achieved but our understanding of how microscale network changes translate to the macroscale mechanics is still far from perfect. The mechanism by which alterations in crosslink density engender alterations in micromechanics requires further intermediate length-scale investigation.

### 7.1: Microscale Network Topology

In this thesis (Section 5) two assays were developed and described; firstly to measure the crosslinking efficiency of the different pentameric I27 monomer hydrogel species, and secondly to measure their unfolded fraction. The first of these assays showed that the different geometries of crosslink sites in the polyprotein monomers resulted in significant variations in crosslinking efficiency, which in turn corresponded to even more marked differences in the average number of crosslinks each monomer unit formed (Figures 5.9 and 5.10). This demonstrates that by altering the crosslinking geometry/density of the monomer building block of an FGP hydrogel the crosslink density of the macroscale network topology can be tuned. We hypothesize from these results that in polyprotein building blocks it is crosslink sites located on the two terminal domains which are most sterically accessible and therefore likely to bind into the network. Additionally we believe that internal domain crosslink sites will generally exhibit low crosslinking efficiency due to steric hindrance, but

that even rare crosslinking of these sites can result in significant changes to network topology and so their rational design can still be used as a tool with which to tune hydrogel mechanics. The quantification of the folded fraction of protein post-gelation was performed using two separate methods to observe two separate timescales; firstly solid-state CD spectra of each gel species was obtained prior to, immediately post, and every 15 minutes after gelation up to +45 minutes (Section 4.3). This allowed the relative folded fractions to be observed and the continued presence of secondary structure to be demonstrated, and the observation of the progression of unfolding on a short to medium timescale. This allows us to confidently state that I27 polyprotein hydrogels are folded protein hydrogels. The CD measurements showed that all construct solutions exhibited approximately equal folded fractions, and that all underwent a degree of unfolding during gelation with some variation. Secondly a cysteine-shotgun labelling assay was developed to measure the absolute fraction of unfolded I27 domains ~2 hours post-gelation (Section 5.3). These results indicated that all species of I27 polyprotein hydrogel reached an approximately equal steady state unfolded fraction (Figure 5.12). From this we can be confident that macroscale mechanical differences between different hydrogel constructs are unlikely to be the result of differential folded fractions. Both these assays and CD measurements demonstrated robustness and reproducibility. Therefore we believe that they offer an adaptable and useful methodology with which to characterize the microscale network topologies of FGP hydrogels.

## 7.2 Translation of Microscale Network Topology to Macroscale Mechanics

The results described in Section 6 strongly suggest that the different crosslink geometries/densities which exist in the different I27 polyproteins cause changes in the

macroscale mechanical properties of the hydrogels they form. Further discussion in Section 6.4 presents potential mechanisms by which these geometries may correlate to certain mechanical characteristics. These hypotheses stem from our interpretation of how network assembly kinetics (Section 6.2.1, 6.4.1) correlate to alterations in the DLA/RLA regime of hydrogel network formation. We believe these results show that FGP hydrogel mechanics can be tuned by regulating the dominance of DLA/RLA and DLCA/RLCA network formation, and that this balance is determined by the crosslink geometry of the monomer building block. The intricacies of this relationship are not yet fully understood, although we can infer that the average steric accessibility of crosslink sites in the monomer is the factor which most affects the probability of forming a crosslink ( $p$ ). The introduction of less sterically accessible crosslink sites will move the average  $p$  further from 1 thereby shifting assembly balance towards RLA. If crosslink sites are uniformly highly accessible then  $p$  will increase and shift assembly towards DLA. It is therefore of crucial importance when designing the monomer building block to use carefully positioned crosslink sites to either maximize or minimize the steric hindrance caused by the potential formation of a crosslink on the accessibility of all others.

### 7.3 Final Thoughts and Future Directions

The study described in this thesis illustrates that in the absence of any other variable such as volume fraction or folded fraction, changes in crosslink geometry can be used to tune the mechanics of folded protein hydrogels. This study is ultimately incomplete as we cannot yet describe in detail how different network topologies correspond to different macromechanics because whilst we have investigated and characterized the network on the microscale and the macroscale we have not gathered data regarding the intermediate length scale of cluster



sizes. Information on this length scale will likely allow direct rationalization between crosslink geometry and micromechanics in conjunction with the data presented here. In addition we do anticipate that folded fraction will have an effect on hydrogel mechanics, and so in the future a study focusing on variations in this parameter independent of crosslink geometry either by mutagenesis or by use of chemical denaturants will be necessary to further understand FGP hydrogel systems.

We have shown that in this system a significant fraction of protein remains folded post-gelation, and that the monomer building block can be tuned in a facile way to regulate gel mechanics. The next stage in this development must therefore focus on incorporating folded domains which exhibit some form of enzymatic/catalytic activity, and demonstration of their continued activity in an FGP hydrogel. This will represent a significant advancement in hydrogel adaptability as it will allow the marriage of both catalytic protein function and inherent micromechanics into the property portfolio of hydrogels.

# Appendix

## Appendix 1: Section 5 T-Test Results Summary Tables

Construct 1	Construct 2	T-test Result p-value
GC1-A	GC2-A	0.301
GC1-A	GC3-A	0.0002
GC1-A	GC5-A	1.21E-09
GC1-A	GC6-A	0.001
GC2-A	GC3-A	0.0001
GC2-A	GC5-A	1.43E-09
GC2-A	GC6-A	0.0004
GC3-A	GC5-A	0.044
GC3-A	GC6-A	0.598
GC5-A	GC6-A	0.321

Construct 1	Construct 2	T-test Result p-value
GC1-A	GC2-A	1.33E-09
GC1-A	GC3-A	5.06E-06
GC1-A	GC5-A	0.1088
GC1-A	GC6-A	0.0002
GC2-A	GC3-A	2.35E-10
GC2-A	GC5-A	1.51E-14
GC2-A	GC6-A	3.79E-07
GC3-A	GC5-A	4.81E-05
GC3-A	GC6-A	0.5984
GC5-A	GC6-A	0.0007

Construct 1	Construct 2	T-test Result p-value
GC1-A	GC2-A	0.0195
GC1-A	GC3-A	0.0024
GC1-A	GC5-A	0.0006
GC1-A	GC6-A	0.0255
GC2-A	GC3-A	0.8655
GC2-A	GC5-A	0.3672
GC2-A	GC6-A	0.2145
GC3-A	GC5-A	0.2812
GC3-A	GC6-A	0.1308
GC5-A	GC6-A	0.3570

## Appendix 2: DNA and Protein Sequences

### Appendix 2.1: I27 Monomers in pGem-T-Easy Vectors

“Pseudo Wild Type”:

```
ACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTAGCCTAATAGAAGTG
GAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTAC
GGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCT
GATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAA
AGTGAAAGAATTGTAAACGCG
```

Y9S Mutant:

```
ACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTAGCCTAATAGAAGTG
GAAAAGCCTCTGTACGGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTAC
GGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCT
GATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAA
AGTGAAAGAATTGTAAACGCG
```

H31Y Mutant:

```
ACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTAGCCTACTAGAAGTG
GAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTAT
GGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTATCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCT
GATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAA
AGTGAAAGAATTGTAAACGCG
```

H31Y/S44Y Mutant:

```
ACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTAGCCTACTAGAAGTG
GAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTAT
GGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTATCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCT
GATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAA
AGTGAAAGAATTGTAAACGCG
```

## Appendix 2.2: I27 Pentamers in pET14b Vectors

GC1:

DNA;

```
AACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGAGCCTAATAGAAGTGGAAAAGCCTCTGTACGG
AGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATTATCACGGCCAGTGGAAGCT
GAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAAC
TCTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAA
TTGCTGAGCGTGGGCGCGACCATTCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAC
AGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCC
CCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAG
GTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAATTGACCGTTATTGGTCTGGCGAGCC
TAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTG
AACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATG
GAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGCTAATGCCAA
ATCTGCAGCCAATCTGAAAGTAAAAGAATTGGCGCTGAGCGGCACCATTGTGCTAATAGAAGTGGAAAAGCCTCTGTA
CGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGA
GCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCAT
AACTCTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAA
GAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGA
AACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATTATCACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGC
TCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGA
GAGGTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAATTGTGTTGTTGAATATGTCGA
GGATCC
```

Protein Sequence:

```
MHHHHHSSLIEVEKPLYGVEVFVGETAHFEIELSEPDYHGQWKLKGQPLTASPDSEIIEDGKKHILIHNSQLGMTGEVSFQAA
NAKSAANLKVKELLSVGATILIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILIHNSQLGM
TGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHIL
ILIHNSQLGMTGEVSFQAANAKSAANLKVKELALSGLTIVLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPD
EIEDGKKHILIHNSQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVFVGETAHFEIELSEPDYHGQWKLK
GQPLTASPDSEIIEDGKKHILIHNSQLGMTGEVSFQAANAKSAANLKVKELCC
```

pI: 5.26

MW: 52579.97 Da

Extinction coefficient: 35075 M<sup>-1</sup>cm<sup>-1</sup>

GCO:

DNA;

AACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGAGCCTAATAGAAGTGGAAAAGCCTCTGTCGGG  
AGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGAAGCT  
GAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAAC  
TCTCAGCTGGGTATGACAGGAGAGGTTTCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAA  
TTGCTGAGCGTGGGCGGACCACTTAATAGAAGTGGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGAAC  
AGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGAAGCTGAAAGGACAGCCTTTGACAGCTTCC  
CCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAG  
GTTTCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAAATTGACCGTTATTGGTCTGGCGAGCC  
TAATAGAAGTGGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTG  
AACCTGATGTTACGGCCAGTGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATG  
GAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTTCTCCAGGCTGCTAATGCCAA  
ATCTGCAGCCAATCTGAAAGTGAAGAAATTGGCGCTGAGCGGCACCACTTGTGCTAATAGAAGTGGAAAAGCCTCTGT  
GGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAA  
GCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCAT  
AACTCTCAGCTGGGTATGACAGGAGAGGTTTCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAG  
GAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGA  
AACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGAAGCTGAAAGGACAGCCTTTGACAGC  
TCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGA  
GAGGTTTCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAAATTGTGTTGTTGAATATGTCGA  
GGATCC

Protein Sequence :

MHHHHHSSLIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILILHNSQLGMTGEVSFQAA  
NAKSAANLKVKELLSVGATILIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILILHNSQLGM  
TGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHIL  
ILHNSQLGMTGEVSFQAANAKSAANLKVKELALSIVLIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDS  
EIIEDGKKHILILHNSQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLK  
GQPLTASPDEIIEDGKKHILILHNSQLGMTGEVSFQAANAKSAANLKVKELCC

Molecular weight: 52199.48 Da

Theoretical pI: 5.26

Ext. coefficient 27500 M<sup>-1</sup>cm<sup>-1</sup>

GC0-A:

DNA;

TATATTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTAGCCTAATAGAA  
GTGGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGAT  
GTTACAGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAAGAA  
GCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCA  
GCCAATCTGAAAGTGAAAGAATTGCTGAGCGTGGGCGCGACCATTCTAATAGAAGTGAAAAGCCTCTGTCGGGAGT  
AGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACAGGCCAGTGGAAGCTGAA  
AGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAAGAAGCATATTCTGATCCTTCATAACTGT  
CAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTG  
ACCGTTGCGCTGAGCGGCACCATTGTGCTAATAGAAGTGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGA  
AACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACAGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGC  
TCCCCTGACTCTGAAATCATTGAGGATGGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGA  
GAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGTTTATTACCGGTAGCCTG  
GCGCTAATAGAAGTGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTT  
TCTGAACCTGATGTTACAGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAG  
GATGGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAAT  
GCCAATCTGCAGCCAATCTGAAAGTGAAAGAATTGTGAGCTCGAGGATCC

Protein Sequence:

MHHHHHSENLYFQGSLLVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELLSVGATILIEVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHIL  
HNCQLGMTGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEI  
IEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALSGTIVLIEVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKG  
QPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLSGVEVFGVGETAHFEIELSEPDV  
HGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKEL

Molecular weight: 52925.42 Da

Theoretical pI: 5.23

Ext. coefficient 28990 M<sup>-1</sup>cm<sup>-1</sup>

GC1-A:

DNA;

TAATTTTGTAACTTTAAGAAGGAGATATAACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTA  
GCCTAATAGAAGTGAAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCACTTTGAAATTGAACTTT  
CTGAACCTGATGTTACGGCCAGTGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGG  
ATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATG  
CCAAATCTGCAGCCAATCTGAAAGTGAAGAATTGCTGAGCGTGGGCGCGACCATTCTAATAGAAGTGAAAAAGCCT  
CTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGT  
GGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAGAAGCATATTCTGATCC  
TTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAGAATTGACCGTTATTGGTCTGGCGAGCCTAATAGAAGTGAAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGG  
GTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGAAGCTGAAAGGACAGCCTTTGA  
CAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGAC  
AGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAATTGGCGCTGAGCGGCA  
CCATTGTGCTAATAGAAGTGAAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTG  
AACTTTCTGAACCTGATGTTACGGCCAGTGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCAT  
TGAGGATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGC  
TAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGAAAA  
GCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGC  
CAGTGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAGAAGCATATTCTG  
ATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGA  
AAGTGAAGAATTGTGAGCTCGAGGATCC

Protein Sequence;

MHHHHHSENLYFQGSLLVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELLSVGATILIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILI  
LHNCQLGMTGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDS  
EIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALSGLTIVLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLK  
GQPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVFVGETAHFEIELSEP  
DVHGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKEL

Molecular weight: 53305.91 Da

Theoretical pI: 5.23

Ext. coefficient  $36440 \text{ M}^{-1}\text{cm}^{-1}$

GC2-A:

DNA;

TAATTTTGTAACTTTAAGAAGGAGATATAACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTA  
GCCTAATAGAAGTGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTAAAACAGCCCACTTTGAAATTGAACTTT  
CTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGG  
ATGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATG  
CCAAATCTGCAGCCAATCTGAAAGTAAAAGATTGCTGAGCGTGGGCGCGACCATTCTAATAGAAGTGAAAAGCCT  
CTGTCGGGAGTAGAGGTGTTTGTGGTAAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGT  
GGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGAAAAGAAGCATATTCTGATCC  
TTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAGAATTGACCGTTATTGGTCTGGCGAGCCTAATAGAAGTGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTG  
GTGAAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGA  
CAGCTTCCCCTGACTCTGAAATCATTGAGGATGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGAC  
AGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAATTGGCGCTGAGCGGCA  
CCATTGTGCTAATAGAAGTGAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTAAAACAGCCCACTTTGAAATTG  
AACTTTCTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCAT  
TGAGGATGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGC  
TAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGAAA  
GCCTCTGTACGGAGTAGAGGTGTTTGTGGTAAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGC  
CAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGAAAAGAAGCATATTCTG  
ATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGA  
AAGTAAAAGAATTGTGAGCTCGAGGATCC

Protein Sequence;

MHHHHHSENLYFQGSLLVEKPLYGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTAS  
PDSEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELLSVGATILIEVEKPL  
SGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLSGVEVFGVGETAHFEIELSEPDVHG  
QWKLKGQPLTASPDSEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALS  
GTIVLIEVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKH  
ILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVFGVGETAH  
FEIELSEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILHNCQLGMTGEVSFQAANAKS  
AANLKVKEL

Molecular weight: 53077.62 Da

Theoretical pI: 5.23

Ext. coefficient 31970 M<sup>-1</sup>cm<sup>-1</sup>



GC3-A:

DNA;

TAATTTTGTTTAACTTTAAGAAGGAGATATAACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTA  
GCCTAATAGAAGTGAAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTT  
CTGAACCTGATGTTTATGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTATCCTGACTCTGAAATCATTGAGG  
ATGGAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATG  
CCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGCTGAGCGTGGGCGCGACCATTCTAATAGAAGTGAAAAAGCCT  
CTGTCGGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGT  
GGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCC  
TTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAGAATTGACCGTTATTGGTCTGGCGAGCCTAATAGAAGTGAAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTG  
GTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGA  
CAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGAC  
AGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGGCGCTGAGCGGCA  
CCATTGTGCTAATAGAAGTGAAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTG  
AACTTTCTGAACCTGATGTTACGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCAT  
TGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGC  
TAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGAAAA  
GCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTATGGC  
CAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTATCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTG  
ATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGA  
AAGTGAAAGAATTGTGAGCTCGAGGATCC

Protein Sequence;

MHHHHHSENLYFQGSLLVEKPLYGVEVFVGETAHFEIELSEPDVYQWKLKGQPLTAYPDSEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELLSVGATILIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPSEIIEDGKKHIL  
HNCQLGMTGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPSEI  
IEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALSGLTIVLIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKG  
QPLTASPSEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVFVGETAHFEIELSEPDV  
YQWKLKGQPLTAYPDSEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKEL

Molecular weight: 53281.88 Da

Theoretical pI: 5.16

Ext. coefficient 37930 M<sup>-1</sup>cm<sup>-1</sup>

GC4-A:

DNA;

TAATTTTGTTTAACTTTAAGAAGGAGATATACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTAG  
CCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGGTGGTAAAACAGCCCACTTTGAAATTGAACTTC  
TGAACCTGATGTTTATGGCCAGTGGAAAGCTGAAAAGGACAGCCTTTGACAGCTTATCCTGACTCTGAAATCATTGAGGA  
TGGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGCTAATGCC  
AAATCTGCAGCCAATCTGAAAGTAAAAGAATTGCTGAGCGTGGGCGCGACCATTCTAATAGAAGTGGAAAAGCCTCT  
GTACGGAGTAGAGGTGTTTGGTGGTAAAACAGCCCACTTTGAAATTGAACTTTCTGAACTGATGTTTATGGCCAGTG  
GAAGCTGAAAGGACAGCCTTTGACAGCTTCCCTGACTCTGAAATCATTGAGGATGGAAAAGAAGCATATTCTGATCCT  
TCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAAGAATTGACCGTTATTGGTCTGGCGAGCCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGGT  
GTAAAACAGCCCACTTTGAAATTGAACTTTCTGAACTGATGTTTATGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGA  
CAGCTTCCCTGACTCTGAAATCATTGAGGATGGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGAC  
AGGAGAGGTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAATTGGCGCTGAGCGGCA  
CCATTGTGCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGGTGGTAAAACAGCCCACTTTGAAATTG  
AACTTTCTGAACTGATGTTTATGGCCAGTGGAAAGCTGAAAAGGACAGCCTTTGACAGCTTCCCTGACTCTGAAATCAT  
TGAGGATGGAAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGC  
TAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGGAAA  
GCCTCTGTACGGAGTAGAGGTGTTTGGTGGTAAAACAGCCCACTTTGAAATTGAACTTTCTGAACTGATGTTTATGGC  
CAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTATCCTGACTCTGAAATCATTGAGGATGGAAAAGAAGCATATTCTG  
ATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCCTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGA  
AAGTAAAAGAATTGTGAGCTCGAGGATCC

Protein Sequence;

MHHHHHSENLYFQGSLLVEKPLYGVEVVFVGETAHFEIELSEPDVYGQWKLKGQPLTAYPDSEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELLSVGATILIEVEKPLYGVEVVFVGETAHFEIELSEPDVYGQWKLKGQPLTAYPDSEIIEDGKKHIL  
HNCQLGMTGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLYGVEVVFVGETAHFEIELSEPDVYGQWKLKGQPLTAYPDSEI  
IEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALSGLTIVLIEVEKPLYGVEVVFVGETAHFEIELSEPDVYGQWKLKGQ  
PLTAYPDSEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVVFVGETAHFEIELSEPDVY  
GQWKLKGQPLTAYPDSEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKEL

Molecular weight: 53588.28 Da

Theoretical pI: 5.06

Ext. coefficient 46870 M<sup>-1</sup>cm<sup>-1</sup>

GC5-A:

DNA;

TAATTTTGTAACTTTAAGAAGGAGATATAACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTA  
GCCTAATAGAAGTGAAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTT  
CTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGG  
ATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATG  
CCAAATCTGCAGCCAATCTGAAAGTGAAGAATTGCTGAGCGTGGGCGGACCATTCTAATAGAAGTGAAAAAGCCT  
CTGTCGGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGT  
GGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAGAAGCATATTCTGATCC  
TTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAGAATTGACCGTTATTGGTCTGGCGAGCCTAATAGAAGTGAAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTG  
GTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGA  
CAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGAC  
AGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAATTGGCGCTGAGCGGCA  
CCATTGTGCTAATAGAAGTGAAAAAGCCTCTGTCGGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTG  
AACTTTCTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCAT  
TGAGGATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGC  
TAATGCCAAATCTGCAGCCAATCTGAAAGTGAAGAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGAAAA  
GCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGC  
CAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAGAAGCATATTCTG  
ATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGA  
AAGTGAAGAATTGTGAGCTCGAGGATCC

Protein Sequence;

MHHHHHSENLYFQGSLLVEKPLYGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELLSVGATILIEVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHIL  
HNCQLGMTGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLYGVEVFGVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEI  
IEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALSGTIVLIEVEKPLSGVEVFGVGETAHFEIELSEPDVHGQWKLKG  
QPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVFGVGETAHFEIELSEPDV  
HGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKEL

Molecular weight: 53153.71 Da

Theoretical pI: 5.23

Ext. coefficient 33460 M<sup>-1</sup>cm<sup>-1</sup>

GC6-A:

DNA;

TAATTTTGTTTAACTTTAAGAAGGAGATATAACCATGCATCACCATCACCATCACTCGGAAAATCTTTATTTTCAAGGTA  
GCCTAATAGAAGTGGAAGAAAGCCTCTGTACGGAGTAGAGGTGTTTGGTGGTAAACAGCCCACTTTGAAATTGAACTTT  
CTGAACCTGATGTTTATGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGG  
ATGGAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATG  
CCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGCTGAGCGTGGGCGGACCATTCTAATAGAAGTGGAAGAAAGCCT  
CTGTCGGGAGTAGAGGTGTTTGGTGGTAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGT  
GGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCC  
TTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAGAATTGACCGTTATTGGTCTGGCGAGCCTAATAGAAGTGGAAGAAAGCCTCTGTACGGAGTAGAGGTGTTTGGT  
GTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTATGGCCAGTGGAAGCTGAAAGGACAGCCTTTGA  
CAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGAC  
AGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGGCGCTGAGCGGCA  
CCATTGTGCTAATAGAAGTGGAAGAAAGCCTCTGTCGGGAGTAGAGGTGTTTGGTGGTAAACAGCCCACTTTGAAATTG  
AACTTTCTGAACCTGATGTTACGGCCAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCAT  
TGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGC  
TAATGCCAAATCTGCAGCCAATCTGAAAGTGAAAGAATTGGTTATTACCGGTAGCCTGGCGCTAATAGAAGTGGAAGAA  
GCCTCTGTACGGAGTAGAGGTGTTTGGTGGTAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTTATGGC  
CAGTGGAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTG  
ATCCTTCATAACTGTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGA  
AAGTGAAAGAATTGTGAGCTCGAGGATCC

Protein Sequence;

MHHHHHSENLYFQGSLLVEKPLYGVEVFVGETAHFEIELSEPDVYGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMT  
GEVSFQAANAKSAANLKVKELLSVGATILIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHIL  
HNCQLGMTGEVSFQAANAKSAANLKVKELTVIGLASLIEVEKPLYGVEVFVGETAHFEIELSEPDVYGQWKLKGQPLTASPDEI  
IEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELALSGTIVLIEVEKPLSGVEVFVGETAHFEIELSEPDVHGQWKLKG  
QPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKELVITGSLALIEVEKPLYGVEVFVGETAHFEIELSEPDV  
YGQWKLKGQPLTASPDEIIEDGKKHILHNCQLGMTGEVSFQAANAKSAANLKVKEL

Molecular weight: 53231.82 Da

Theoretical pI: 5.13

Ext. coefficient 37930 M<sup>-1</sup>cm<sup>-1</sup>

(127)<sub>5</sub> [179]:

DNA;

ATGCATCACCATCACCATCACTCGAGCCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAAC  
AGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTCCCC  
TGA CTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTT  
CCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAACTAGTAGAGGCTCGACTAATAGAAGTGGAA  
AAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGC  
CAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGAT  
CCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGT  
GAAAGAATTGCTTATCGAAGCGCGCCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACA  
GCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCT  
GACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTTC  
CTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAAATTGCTGAGCTCGGCTCGACTAATAGAAGTGG  
AAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGG  
GCCAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTTCCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTG  
ATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAA  
GTGAAAGAATTGCTTATCGAAGCACGGGCCCTAATAGAAGTGGAAAAGCCTCTGTACGGAGTAGAGGTGTTTGTGGTG  
AAACAGCCCACTTTGAAATTGAACTTTCTGAACCTGATGTTACGGCCAGTGGAAAGCTGAAAGGACAGCCTTTGACAGCTT  
CCCCTGACTCTGAAATCATTGAGGATGGAAAGAAGCATATTCTGATCCTTCATAACTCTCAGCTGGGTATGACAGGAGAG  
GTTTCTTCCAGGCTGCTAATGCCAAATCTGCAGCCAATCTGAAAGTAAAAGAAATTGTGTTGTTAAACGCGT

Protein Sequence;

MHHHHHSSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILIHNSQLGMTGEVSFQAA  
NAKSAANLKVKELVEARLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILIHNSQLGMTG  
EVSFQAANAKSAANLKVKELLIEARLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGKKHILIHNS  
QLGMTGEVSFQAANAKSAANLKVKELLSSARLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDEIIEDGK  
KHILIHNSQLGMTGEVSFQAANAKSAANLKVKELLIEARLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASP  
DEIIEDGKKHILIHNSQLGMTGEVSFQAANAKSAANLKVKELCC

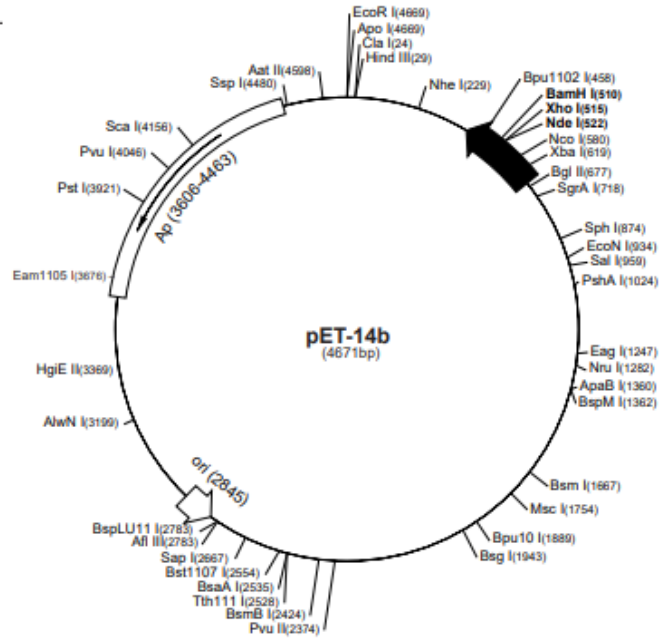
Molecular weight: 52219.15 Da

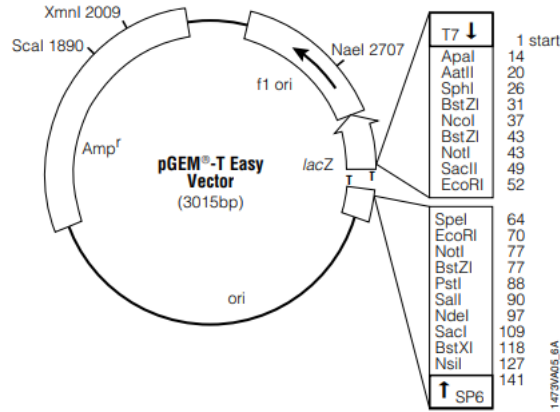
Theoretical pI 5.36

Extinction coefficient = 34970 M<sup>-1</sup>cm<sup>-1</sup>

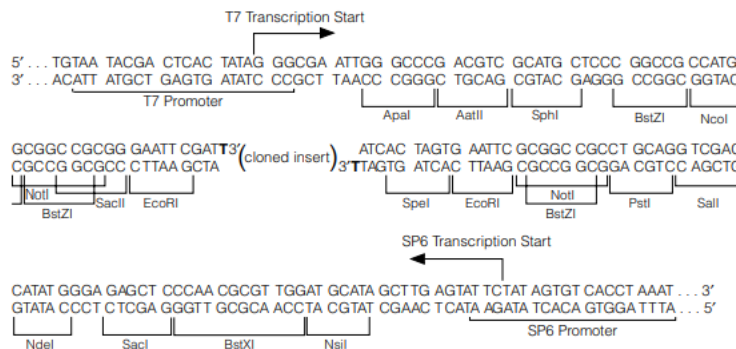
## Appendix 2.3: Vectors

pET-14b sequence landmarks	
T7 promoter	646-662
T7 transcription start	645
His-Tag coding sequence	554-571
Multiple cloning sites ( <i>Nde</i> I - <i>Bam</i> HI)	510-526
T7 terminator	404-450
pBR322 origin	2845
<i>bla</i> coding sequence	3606-4463





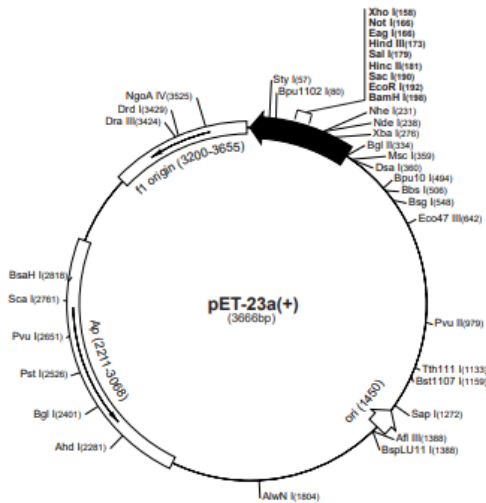
## Sequence and Multiple Cloning Site of pGEM-T-Easy Vector (Promega)



**pET-23a(+) sequence landmarks**

T7 promoter	303-319
T7 transcription start	302
T7*Tag coding sequence	207-239
Multiple cloning sites ( <i>Bam</i> H I - <i>Xho</i> I)	158-203
His*Tag coding sequence	140-157
T7 terminator	26-72
pBR322 origin	1450
<i>bla</i> coding sequence	2211-3068
<i>f1</i> origin	3200-3655

The maps for pET-23b(+), pET-23c(+), and pET-23d(+) are the same as pET-23a(+) (shown) with the following exceptions:  
 pET-23b(+) is a 3663bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198.  
 pET-23c(+) is a 3664bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.  
 pET-23d(+) is a 3663bp plasmid; the *Bam*H I site is in the same reading frame as in pET-23c(+). An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 238 of pET-23c(+). As a result, *Nco* I cuts pET-23d(+) at 234, and *Nhe* I cuts at 229. For the rest of the sites, subtract 3bp from each site beyond position 239 in pET-23a(+). *Nde* I does not cut pET-23d(+). Note also that *Sty* I is not unique in pET-23d(+).

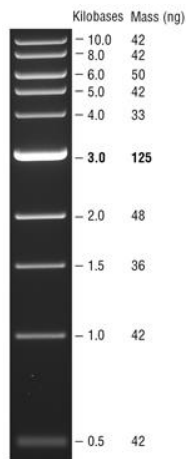


### Appendix 3: SDS and Agarose Gel Ladders

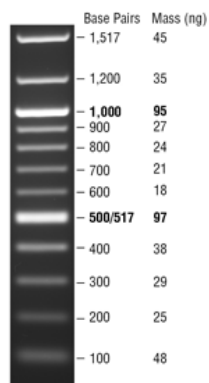
Precision plus protein dual colour standards protein ladder (BioRad, CA, USA):



1kb DNA Ladder (NEB, NE, USA):



100kb DNA Ladder (NEB, NE, USA):





## References

- [1] S. Macromolecules, "Structural Macromolecules and Supramolecular Organisation of the Vitreous Gel," *Prog. Retin. Eye Res.*, vol. 19, no. 3, 2000.
- [2] A. S. Hoffman, "Hydrogels for biomedical applications," *Advanced Drug Delivery Reviews*, vol. 64, no. SUPPL. Elsevier, pp. 18–23, Dec. 01, 2012, doi: 10.1016/j.addr.2012.09.010.
- [3] S. Van Vlierberghe, P. Dubruel, and E. Schacht, "Biopolymer-based hydrogels as scaffolds for tissue engineering applications: A review," *Biomacromolecules*, vol. 12, no. 5. American Chemical Society, pp. 1387–1408, May 09, 2011, doi: 10.1021/bm200083n.
- [4] Y. S. Zhang and A. Khademhosseini, "Advances in engineering hydrogels," *Science (80-. )*, vol. 356, no. 6337, 2017, doi: 10.1126/science.aaf3627.
- [5] "Theory of Viscoelasticity: An Introduction - R Christensen - Google Books." [https://books.google.co.uk/books?hl=en&lr=&id=-k2-WE7QIkYc&oi=fnd&pg=PP1&dq=intro+to+viscoelasticity&ots=FjExAc-rbN&sig=zzRvKB1FiD-9eRnVNsFOTkFdU6E&redir\\_esc=y#v=onepage&q=intro to viscoelasticity&f=false](https://books.google.co.uk/books?hl=en&lr=&id=-k2-WE7QIkYc&oi=fnd&pg=PP1&dq=intro+to+viscoelasticity&ots=FjExAc-rbN&sig=zzRvKB1FiD-9eRnVNsFOTkFdU6E&redir_esc=y#v=onepage&q=intro to viscoelasticity&f=false) (accessed May 15, 2020).
- [6] Y. Yeo, E. Bellas, W. Firestone, R. Langer, and D. S. Kohane, "Complex coacervates for thermally sensitive controlled release of flavor compounds," *J. Agric. Food Chem.*, vol. 53, no. 19, pp. 7518–7525, Sep. 2005, doi: 10.1021/jf0507947.
- [7] A. Lazcano and S. L. Miller, "The origin and early evolution of life: Prebiotic chemistry, the pre-RNA world, and time," *Cell*, vol. 85, no. 6. Cell Press, pp. 793–798, Jun. 14,

- 1996, doi: 10.1016/S0092-8674(00)81263-5.
- [8] G. Sengle, A. Eisenführ, P. S. Arora, J. S. Nowick, and M. Famulok, "Novel RNA catalysts for the Michael reaction," *Chem. Biol.*, vol. 8, no. 5, pp. 459–473, May 2001, doi: 10.1016/S1074-5521(01)00026-6.
- [9] A. F. Silva, M. A. Alves, and M. S. N. Oliveira, "Rheological behaviour of vitreous humour," *Rheol. Acta*, vol. 56, no. 4, pp. 377–386, Apr. 2017, doi: 10.1007/s00397-017-0997-0.
- [10] C. Gambini, B. Abou, A. Ponton, and A. J. M. Cornelissen, "Micro- and macrorheology of jellyfish extracellular matrix," *Biophys. J.*, vol. 102, no. 1, pp. 1–9, Jan. 2012, doi: 10.1016/j.bpj.2011.11.4004.
- [11] T. A. Duncombe *et al.*, "Hydrogel Pore-Size Modulation for Enhanced Single-Cell Western Blotting," *Adv. Mater.*, vol. 28, no. 2, pp. 327–334, Jan. 2016, doi: 10.1002/adma.201503939.
- [12] S. J. Kim, S. J. Park, and S. I. Kim, "Swelling behavior of interpenetrating polymer network hydrogels composed of poly(vinyl alcohol) and chitosan," *React. Funct. Polym.*, vol. 55, no. 1, pp. 53–59, Feb. 2003, doi: 10.1016/S1381-5148(02)00214-6.
- [13] S. Kou, Z. Yang, and F. Sun, "Protein hydrogel microbeads for selective uranium mining from seawater," *ACS Appl. Mater. Interfaces*, vol. 9, no. 3, pp. 2035–2039, 2017, doi: 10.1021/acsami.6b15968.
- [14] G. W. Ashley, J. Henise, R. Reid, and D. V. Santi, "Hydrogel drug delivery system with predictable and tunable drug release and degradation rates," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 6, pp. 2318–2323, Feb. 2013, doi: 10.1073/pnas.1215498110.

- [15] F. T. Bosman and I. Stamenkovic, "Functional structure and composition of the extracellular matrix," *J. Pathol.*, vol. 200, no. 4, pp. 423–428, Jul. 2003, doi: 10.1002/path.1437.
- [16] C. Frantz, K. M. Stewart, and V. M. Weaver, "The extracellular matrix at a glance," *J. Cell Sci.*, vol. 123, no. 123, pp. 4195–4200, 2010, doi: 10.1242/jcs.023820.
- [17] H. Järveläinen, A. Sainio, M. Koulu, T. N. Wight, and R. Penttinen, "Extracellular matrix molecules: Potential targets in pharmacotherapy," *Pharmacological Reviews*, vol. 61, no. 2. American Society for Pharmacology and Experimental Therapeutics, pp. 198–223, Jun. 01, 2009, doi: 10.1124/pr.109.001289.
- [18] D. E. Kipp, M. McElvain, D. B. Kimmel, M. P. Akhter, R. G. Robinson, and B. P. Lukert, "Scurvy results in decreased collagen synthesis and bone density in the guinea pig animal model," *Bone*, vol. 18, no. 3, pp. 281–288, Mar. 1996, doi: 10.1016/8756-3282(95)00481-5.
- [19] "13.1: Introduction to Extracellular Matrix and Cell Adhesion - Biology LibreTexts." [https://bio.libretexts.org/Bookshelves/Cell\\_and\\_Molecular\\_Biology/Book%3A\\_Cells\\_and\\_Molecules\\_and\\_Mechanisms\\_\(Wong\)/13%3A\\_Extracellular\\_Matrix\\_and\\_Cell\\_Adhesion/13.01%3A\\_Introduction\\_to\\_Extracellular\\_Matrix\\_and\\_Cell\\_Adhesion](https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Cells_and_Molecules_and_Mechanisms_(Wong)/13%3A_Extracellular_Matrix_and_Cell_Adhesion/13.01%3A_Introduction_to_Extracellular_Matrix_and_Cell_Adhesion) (accessed May 14, 2020).
- [20] P. Fratzl, "Collagen: Structure and mechanics, an introduction," in *Collagen: Structure and Mechanics*, Springer US, 2008, pp. 1–13.
- [21] O. De Wever, P. Demetter, M. Mareel, and M. Bracke, "Stromal myofibroblasts are drivers of invasive cancer growth," *Int. J. Cancer*, vol. 123, no. 10, pp. 2229–2238,

- Nov. 2008, doi: 10.1002/ijc.23925.
- [22] S. Sibilla, M. Godfrey, S. Brewer, A. Budh-Raja, and L. Genovese, "An overview of the beneficial effects of hydrolysed collagen as a nutraceutical on skin properties: Scientific background and clinical studies," *Open Nutraceuticals J.*, vol. 8, no. 1, pp. 29–42, Jan. 2015, doi: 10.2174/1876396001508010029.
- [23] M. D. Shoulders and R. T. Raines, "Collagen Structure and Stability," 2009, doi: 10.1146/annurev.biochem.77.032207.120833.
- [24] T. Rozario and D. W. DeSimone, "The extracellular matrix in development and morphogenesis: A dynamic view," *Developmental Biology*, vol. 341, no. 1. Academic Press Inc., pp. 126–140, May 01, 2010, doi: 10.1016/j.ydbio.2009.10.026.
- [25] D. W. Urry *et al.*, "Elastin: a representative ideal protein elastomer," doi: 10.1098/rstb.2001.1023.
- [26] S. M. Mithieux and A. S. Weiss, "Elastin," *Adv. Protein Chem.*, vol. 70, pp. 437–461, Jan. 2005, doi: 10.1016/S0065-3233(05)70013-9.
- [27] S. G. Wise and A. S. Weiss, "Tropoelastin," *International Journal of Biochemistry and Cell Biology*, vol. 41, no. 3. Pergamon, pp. 494–497, Mar. 01, 2009, doi: 10.1016/j.biocel.2008.03.017.
- [28] S. G. Wise, S. M. Mithieux, and A. S. Weiss, "Engineered tropoelastin and elastin-based biomaterials.," *Advances in protein chemistry and structural biology*, vol. 78. Academic Press, pp. 1–24, Jan. 01, 2009, doi: 10.1016/s1876-1623(08)78001-5.
- [29] H. A. Lucero and H. M. Kagan, "Lysyl oxidase: An oxidative enzyme and effector of cell function," *Cellular and Molecular Life Sciences*, vol. 63, no. 19–20. Springer, pp. 2304–

- 2316, Oct. 11, 2006, doi: 10.1007/s00018-006-6149-9.
- [30] “Fibronectins - Richard O. Hynes - Google Books.”  
[https://books.google.co.uk/books?hl=en&lr=&id=N6oJCAAAQBAJ&oi=fnd&pg=PR5&ots=u2u6CR7DWT&sig=N7gVhQjIBu\\_apPfyLy32OznS4aM&redir\\_esc=y#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=N6oJCAAAQBAJ&oi=fnd&pg=PR5&ots=u2u6CR7DWT&sig=N7gVhQjIBu_apPfyLy32OznS4aM&redir_esc=y#v=onepage&q&f=false) (accessed May 18, 2020).
- [31] J. C. Friedland, M. H. Lee, and D. Boettiger, “Mechanically Activated Integrin Switch Controls a 5 b 1 Function,” doi: 10.1126/science.1168441.
- [32] P. Singh, C. Carraher, and J. E. Schwarzbauer, “Assembly of Fibronectin Extracellular Matrix,” *Annu. Rev. Cell Dev. Biol.*, vol. 26, no. 1, pp. 397–419, Nov. 2010, doi: 10.1146/annurev-cellbio-100109-104020.
- [33] S.-I. Aota, M. Nomizu, and K. M. Yamada, “The Short Amino Acid Sequence Pro-His-Ser-Arg-Asn in Human Fibronectin Enhances Cell-adhesive Function\*,” 1994.
- [34] M. L. Smith *et al.*, “Force-induced unfolding of fibronectin in the extracellular matrix of living cells,” *PLoS Biol.*, vol. 5, no. 10, pp. 2243–2254, Oct. 2007, doi: 10.1371/journal.pbio.0050268.
- [35] K. C. Ingham, S. A. Brew, and H. P. Erickson, “Localization of a Cryptic Binding Site for Tenascin on Fibronectin\* Downloaded from,” *J. Biol. Chem.*, vol. 279, no. 27, pp. 28132–28135, 2004, doi: 10.1074/jbc.M312785200.
- [36] J. D. Mott and Z. Werb, “Regulation of matrix biology by matrix metalloproteinases,” *Current Opinion in Cell Biology*, vol. 16, no. 5. Elsevier Current Trends, pp. 558–564, Oct. 01, 2004, doi: 10.1016/j.ceb.2004.07.010.
- [37] A. Jabłońska-Trypuć, M. Matejczyk, and S. Rosochacki, “Matrix metalloproteinases

- (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs," *J. Enzyme Inhib. Med. Chem.*, vol. 31, no. sup1, pp. 177–183, Nov. 2016, doi: 10.3109/14756366.2016.1161620.
- [38] H. Birkedal-Hansen *et al.*, "Matrix metalloproteinases: A review," *Critical Reviews in Oral Biology and Medicine*, vol. 4, no. 2. International and American Associations for Dental Research, pp. 197–250, Jan. 01, 1993, doi: 10.1177/10454411930040020401.
- [39] H. Nagase, R. Visse, and G. Murphy, "Structure and function of matrix metalloproteinases and TIMPs," *Cardiovascular Research*, vol. 69, no. 3. Oxford Academic, pp. 562–573, Feb. 15, 2006, doi: 10.1016/j.cardiores.2005.12.002.
- [40] D. R. Eyre, M. A. Paz, and P. M. Gallop, "CROSS-LINKING IN COLLAGEN AND ELASTIN," 1984. Accessed: May 19, 2020. [Online]. Available: [www.annualreviews.org](http://www.annualreviews.org).
- [41] L. Cassereau, Y. A. Miroshnikova, G. Ou, J. Lakins, and V. M. Weaver, "A 3D tension bioreactor platform to study the interplay between ECM stiffness and tumor phenotype," *J. Biotechnol.*, vol. 193, pp. 66–69, Jan. 2015, doi: 10.1016/j.jbiotec.2014.11.008.
- [42] S. A. Langhans, "Three-dimensional in vitro cell culture models in drug discovery and drug repositioning," *Frontiers in Pharmacology*, vol. 9, no. JAN. Frontiers Media S.A., p. 6, Jan. 23, 2018, doi: 10.3389/fphar.2018.00006.
- [43] R. Edmondson, J. J. Broglie, A. F. Adcock, and L. Yang, "Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors," *Assay and Drug Development Technologies*, vol. 12, no. 4. Mary Ann Liebert Inc., pp. 207–218, May 01, 2014, doi: 10.1089/adt.2014.573.

- [44] A. B. Bonhome-Espinosa *et al.*, “Effect of particle concentration on the microstructural and macromechanical properties of biocompatible magnetic hydrogels,” *Soft Matter*, vol. 13, no. 16, pp. 2928–2941, 2017, doi: 10.1039/c7sm00388a.
- [45] J. Zhu, “Bioactive modification of poly(ethylene glycol) hydrogels for tissue engineering,” *Biomaterials*, vol. 31, no. 17. Elsevier, pp. 4639–4656, Jun. 01, 2010, doi: 10.1016/j.biomaterials.2010.02.044.
- [46] C. C. Lin, A. Raza, and H. Shih, “PEG hydrogels formed by thiol-ene photo-click chemistry and their effect on the formation and recovery of insulin-secreting cell spheroids,” *Biomaterials*, vol. 32, no. 36, pp. 9685–9695, Dec. 2011, doi: 10.1016/j.biomaterials.2011.08.083.
- [47] Y. C. Chiu *et al.*, “The role of pore size on vascularization and tissue remodeling in PEG hydrogels,” *Biomaterials*, vol. 32, no. 26, pp. 6045–6051, Sep. 2011, doi: 10.1016/j.biomaterials.2011.04.066.
- [48] J. J. Roberts, G. D. Nicodemus, E. C. Greenwald, and S. J. Bryant, “Degradation improves tissue formation in (Un)loaded chondrocyte-laden hydrogels,” in *Clinical Orthopaedics and Related Research*, 2011, vol. 469, no. 10, pp. 2725–2734, doi: 10.1007/s11999-011-1823-0.
- [49] W. Shen, K. Zhang, J. A. Kornfield, and D. A. Tirrell, “Tuning the erosion rate of artificial protein hydrogels through control of network topology,” *Nat. Mater.*, vol. 5, no. 2, pp. 153–158, Feb. 2006, doi: 10.1038/nmat1573.
- [50] T. Y. Cheng, M. H. Chen, W. H. Chang, M. Y. Huang, and T. W. Wang, “Neural stem cells encapsulated in a functionalized self-assembling peptide hydrogel for brain

- tissue engineering," *Biomaterials*, vol. 34, no. 8, pp. 2005–2016, Mar. 2013, doi: 10.1016/j.biomaterials.2012.11.043.
- [51] S. B. Anderson, C. C. Lin, D. V. Kuntzler, and K. S. Anseth, "The performance of human mesenchymal stem cells encapsulated in cell-degradable polymer-peptide hydrogels," *Biomaterials*, vol. 32, no. 14, pp. 3564–3574, May 2011, doi: 10.1016/j.biomaterials.2011.01.064.
- [52] R. H. Schmedlen, K. S. Masters, and J. L. West, "Photocrosslinkable polyvinyl alcohol hydrogels that can be modified with cell adhesion peptides for use in tissue engineering," *Biomaterials*, vol. 23, no. 22, pp. 4325–4332, Nov. 2002, doi: 10.1016/S0142-9612(02)00177-1.
- [53] I. L. Kim, R. L. Mauck, and J. A. Burdick, "Hydrogel design for cartilage tissue engineering: A case study with hyaluronic acid," *Biomaterials*, vol. 32, no. 34, pp. 8771–8782, Dec. 2011, doi: 10.1016/j.biomaterials.2011.08.073.
- [54] K. M. Galler, L. Aulisa, K. R. Regan, R. N. D'Souza, and J. D. Hartgerink, "Self-assembling multidomain peptide hydrogels: Designed susceptibility to enzymatic cleavage allows enhanced cell migration and spreading," *J. Am. Chem. Soc.*, vol. 132, no. 9, pp. 3217–3223, Mar. 2010, doi: 10.1021/ja910481t.
- [55] K. M. Galler, R. N. D'Souza, J. D. Hartgerink, and G. Schmalz, "Scaffolds for Dental Pulp Tissue Engineering," *Adv. Dent. Res.*, vol. 23, no. 3, pp. 333–339, Jul. 2011, doi: 10.1177/0022034511405326.
- [56] G. Sun *et al.*, "Dextran hydrogel scaffolds enhance angiogenic responses and promote complete skin regeneration during burn wound healing," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 52, pp. 20976–20981, Dec. 2011, doi: 10.1073/pnas.1115973108.



- [57] D. Shav, S. Einav, and B. Professor of Vascular Bioengineering, "The effect of mechanical loads in the differentiation of precursor cells into mature cells," *Ann. N.Y. Acad. Sci.*, doi: 10.1111/j.1749-6632.2009.05079.x.
- [58] M. Doi, *Introduction to polymer physics*. 1996.
- [59] P. L. Chandran, D. C. Paik, and J. W. Holmes, "Structural mechanism for alteration of collagen gel mechanics by glutaraldehyde crosslinking," *Connect. Tissue Res.*, vol. 53, no. 4, pp. 285–297, Aug. 2012, doi: 10.3109/03008207.2011.640760.
- [60] A. Simionescu, D. Simionescu, and R. Deac, "Lysine-enhanced glutaraldehyde crosslinking of collagenous biomaterials," *J. Biomed. Mater. Res.*, vol. 25, no. 12, pp. 1495–1505, Dec. 1991, doi: 10.1002/jbm.820251207.
- [61] K. Nakajima *et al.*, "Gelatin Hydrogel Enhances the Engraftment of Transplanted Cardiomyocytes and Angiogenesis to Ameliorate Cardiac Function after Myocardial Infarction," 2015, doi: 10.1371/journal.pone.0133308.
- [62] O. Wichterle and D. Lím, "Hydrophilic Gels for Biological Use," *Nature*, vol. 185, no. 4706, pp. 117–118, 1960, doi: 10.1038/185117a0.
- [63] "Apparatus and method for satisfying disposable contact lens prescriptions," Jul. 1987.
- [64] J. R. Cope, S. A. Collier, H. Nethercut, J. M. Jones, K. Yates, and J. S. Yoder, "Risk Behaviors for Contact Lens–Related Eye Infections Among Adults and Adolescents — United States, 2016," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 66, no. 32, pp. 841–845, Aug. 2017, doi: 10.15585/mmwr.mm6632a2.
- [65] J. R. Cope *et al.*, "Contact lens wearer demographics and risk behaviors for contact

- lens-related eye infections — United States, 2014,” *Morb. Mortal. Wkly. Rep.*, vol. 64, no. 32, pp. 865–870, Aug. 2015, doi: 10.15585/mmwr.mm6432a2.
- [66] C. H. Lin, Y. H. Yeh, W. C. Lin, and M. C. Yang, “Novel silicone hydrogel based on PDMS and PEGMA for contact lens application,” *Colloids Surfaces B Biointerfaces*, vol. 123, pp. 986–994, Nov. 2014, doi: 10.1016/j.colsurfb.2014.10.053.
- [67] R. T. C. Cleophas, M. Riool, H. C. Quarles Van Ufford, S. A. J. Zaat, J. A. W. Kruijtzter, and R. M. J. Liskamp, “Convenient preparation of bactericidal hydrogels by covalent attachment of stabilized antimicrobial peptides using thiol-ene click chemistry,” *ACS Macro Lett.*, vol. 3, no. 5, pp. 477–480, May 2014, doi: 10.1021/mz5001465.
- [68] T. O. Leiknes, “The effect of coupling coagulation and flocculation with membrane filtration in water treatment: A review,” *J. Environ. Sci.*, vol. 21, no. 1, pp. 8–12, Jan. 2009, doi: 10.1016/S1001-0742(09)60003-6.
- [69] O. Legrini, E. Oliveros, and A. M. Braun, “Photochemical Processes for Water Treatment,” *Chem. Rev.*, vol. 93, no. 2, pp. 671–698, 1993, doi: 10.1021/cr00018a003.
- [70] G. S. Simate, S. E. Iyuke, S. Ndlovu, M. Heydenrych, and L. F. Walubita, “Human health effects of residual carbon nanotubes and traditional water treatment chemicals in drinking water,” *Environment International*, vol. 39, no. 1. Elsevier Ltd, pp. 38–49, Feb. 01, 2012, doi: 10.1016/j.envint.2011.09.006.
- [71] J. Kim *et al.*, “Chlorination by-products in surface water treatment process,” *Desalination*, vol. 151, no. 1, pp. 1–9, Jan. 2003, doi: 10.1016/S0011-9164(02)00967-0.
- [72] X. Zeng, D. T. McCarthy, A. Deletic, and X. Zhang, “Silver/Reduced Graphene Oxide

- Hydrogel as Novel Bactericidal Filter for Point-of-Use Water Disinfection,” *Adv. Funct. Mater.*, vol. 25, no. 27, pp. 4344–4351, Jul. 2015, doi: 10.1002/adfm.201501454.
- [73] J.-B. Fan *et al.*, “Directly Coating Hydrogel on Filter Paper for Effective Oil-Water Separation in Highly Acidic, Alkaline, and Salty Environment,” *Adv. Funct. Mater.*, vol. 25, no. 33, pp. 5368–5375, Sep. 2015, doi: 10.1002/adfm.201501066.
- [74] N. Gogoi, M. Barooah, G. Majumdar, and D. Chowdhury, “Carbon dots rooted agarose hydrogel hybrid platform for optical detection and separation of heavy metal ions,” *ACS Appl. Mater. Interfaces*, vol. 7, no. 5, pp. 3058–3067, Feb. 2015, doi: 10.1021/am506558d.
- [75] Y. Yang *et al.*, “Sixty-five years of the long march in protein secondary structure prediction: the final stretch?,” *Brief. Bioinform.*, vol. 19, no. 3, pp. 482–494, 2018, doi: 10.1093/bib/bbw129.
- [76] H. Li, N. Kong, B. Laver, and J. Liu, “Hydrogels Constructed from Engineered Proteins,” *Small*, vol. 12, no. 8, pp. 973–987, Feb. 2016, doi: 10.1002/smll.201502429.
- [77] A. Kwong, B. Gründig, J. Hu, S. Bioscan, and R. Renneberg, “Comparative study of hydrogel-immobilized L-glutamate oxidases for a novel thick-film biosensor and its application in food samples Phytate biosensor View project Microfluidic Biosensor Disposables View project,” *Artic. Biotechnol. Lett.*, 2000, doi: 10.1023/A:1005694704872.
- [78] M. A. Elsliger, R. M. Wachter, G. T. Hanson, K. Kallio, and S. J. Remington, “Structural and spectral response of green fluorescent protein variants to changes in pH,” *Biochemistry*, vol. 38, no. 17, pp. 5296–5301, Apr. 1999, doi: 10.1021/bi9902182.

- [79] Y. Feng and S. Walter Englander, "Salt-Dependent Structure Change and Ion Binding in Cytochrome c Studied by Two-Dimensional Proton NMR," *Biochemistry*, vol. 29, no. 14, pp. 3505–3509, Apr. 1990, doi: 10.1021/bi00466a012.
- [80] J. Ramprakash *et al.*, "Comparison of the chemical and thermal denaturation of proteins by a two-state transition model," *Anal. Biochem.*, vol. 374, no. 1, pp. 221–230, Mar. 2008, doi: 10.1016/j.ab.2007.10.005.
- [81] I. Hayakawa, Y. Y. Linko, and P. Linko, "Mechanism of high pressure denaturation of proteins," *LWT - Food Sci. Technol.*, vol. 29, no. 8, pp. 756–762, Dec. 1996, doi: 10.1006/fstl.1996.0118.
- [82] B. Kelley, "Industrialization of mAb production technology: The bioprocessing industry at a crossroads," *mAbs*, vol. 1, no. 5. Landes Bioscience, pp. 443–452, 2009, doi: 10.4161/mabs.1.5.9448.
- [83] J. F. Buyel, R. M. Twyman, and R. Fischer, "Very-large-scale production of antibodies in plants: The biologization of manufacturing," *Biotechnology Advances*, vol. 35, no. 4. Elsevier Inc., pp. 458–465, Jul. 01, 2017, doi: 10.1016/j.biotechadv.2017.03.011.
- [84] A. P. Chapman, P. Antoniw, M. Spitali, S. West, S. Stephens, and D. J. King, "Therapeutic antibody fragments with prolonged in vivo half-lives," *Nat. Biotechnol.*, vol. 17, no. 8, pp. 780–783, Aug. 1999, doi: 10.1038/11717.
- [85] K. Bojkowska *et al.*, "Measuring in vivo protein half-life," *Chem. Biol.*, vol. 18, no. 6, pp. 805–815, Jun. 2011, doi: 10.1016/j.chembiol.2011.03.014.
- [86] R. F.-A. T. in Biophysics and undefined 2007, "Protein Nanomechanics-asStudied by AFM Single-Molecule Force Spectroscopy," *books.google.com*, Accessed: May 26,

2020. [Online]. Available:  
[https://books.google.com/books?hl=en&lr=&id=ZvluccGskBUC&oi=fnd&pg=PA163&q=dynamic+force+spectroscopy+of+proteins+intro&ots=3n94SRmM2r&sig=zvN42cQc97GgCdFkV9K2FrLeg\\_w](https://books.google.com/books?hl=en&lr=&id=ZvluccGskBUC&oi=fnd&pg=PA163&q=dynamic+force+spectroscopy+of+proteins+intro&ots=3n94SRmM2r&sig=zvN42cQc97GgCdFkV9K2FrLeg_w).
- [87] C. P. Broedersz and F. C. Mackintosh, "Modeling semiflexible polymer networks," *Rev. Mod. Phys.*, vol. 86, no. 3, pp. 995–1036, Jul. 2014, doi: 10.1103/RevModPhys.86.995.
- [88] "Statistical thermodynamics of semi-flexible chain molecules," *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, vol. 234, no. 1196, pp. 60–73, Jan. 1956, doi: 10.1098/rspa.1956.0015.
- [89] D. Baker, "What has de novo protein design taught us about protein folding and biophysics?," *Protein Sci.*, vol. 28, no. 4, pp. 678–683, Apr. 2019, doi: 10.1002/pro.3588.
- [90] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, "Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations," *J. Chem. Phys.*, vol. 101, no. 7, pp. 6052–6062, Oct. 1994, doi: 10.1063/1.467320.
- [91] J. Skolnick and A. Kolinski, "Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics," *J. Mol. Biol.*, vol. 221, no. 2, pp. 499–531, Sep. 1991, doi: 10.1016/0022-2836(91)80070-B.
- [92] H. Li and Y. Cao, "Protein mechanics: From single molecules to functional biomaterials," *Acc. Chem. Res.*, vol. 43, no. 10, pp. 1331–1341, Oct. 2010, doi: 10.1021/ar100057a.

- [93] N. Kong, Q. Peng, and H. Li, "Rationally Designed Dynamic Protein Hydrogels with Reversibly Tunable Mechanical Properties," *Adv. Funct. Mater.*, vol. 24, no. 46, pp. 7310–7317, Dec. 2014, doi: 10.1002/adfm.201402205.
- [94] J. Fang and H. Li, "A facile way to tune mechanical properties of artificial elastomeric proteins-based hydrogels," *Langmuir*, vol. 28, no. 21, pp. 8260–8265, May 2012, doi: 10.1021/la301225w.
- [95] M. D. G. Hughes, S. E. Cussons, N. Mahmoudi, D. J. Brockwell, and L. Dougan, "Single Molecule Protein Stabilisation Translates to Macromolecular Mechanics of Protein Network," *Soft Matter*, 2020, doi: 10.1039/c9sm02484k.
- [96] N. Kong, L. Fu, Q. Peng, and H. Li, "Metal Chelation Dynamically Regulates the Mechanical Properties of Engineered Protein Hydrogels," *ACS Biomater. Sci. Eng.*, vol. 3, no. 5, pp. 742–749, May 2017, doi: 10.1021/acsbio.6b00374.
- [97] S. Lv, T. Bu, J. Kayser, A. Bausch, and H. Li, "Towards constructing extracellular matrix-mimetic hydrogels: An elastic hydrogel constructed from tandem modular proteins containing tenascin FnIII domains," *Acta Biomater.*, vol. 9, no. 5, pp. 6481–6491, May 2013, doi: 10.1016/j.actbio.2013.01.002.
- [98] J. Wu *et al.*, "Rationally designed synthetic protein hydrogels with predictable mechanical properties," *Nat. Commun.*, vol. 9, no. 1, pp. 1–11, 2018, doi: 10.1038/s41467-018-02917-6.
- [99] M. A. Da Silva, S. Lenton, M. Hughes, D. J. Brockwell, and L. Dougan, "Assessing the Potential of Folded Globular Polyproteins As Hydrogel Building Blocks," *Biomacromolecules*, vol. 18, no. 2, pp. 636–646, Feb. 2017, doi: 10.1021/acs.biomac.6b01877.

- [100] J. Fang *et al.*, “ARTICLE Forced protein unfolding leads to highly elastic and tough protein hydrogels,” *Nat. Commun.*, 2013, doi: 10.1038/ncomms3974.
- [101] N. Kong, L. Fu, Q. Peng, and H. Li, “Metal Chelation Dynamically Regulates the Mechanical Properties of Engineered Protein Hydrogels,” *ACS Biomater. Sci. Eng.*, vol. 3, no. 5, pp. 742–749, 2017, doi: 10.1021/acsbiomaterials.6b00374.
- [102] M. A. Da Silva, S. Lenton, M. Hughes, D. J. Brockwell, and L. Dougan, “Assessing the Potential of Folded Globular Polyproteins As Hydrogel Building Blocks,” *Biomacromolecules*, vol. 18, no. 2, pp. 636–646, 2017, doi: 10.1021/acs.biomac.6b01877.
- [103] L. F. Willis *et al.*, “Using extensional flow to reveal diverse aggregation landscapes for three IgG1 molecules,” *Biotechnol. Bioeng.*, vol. 115, no. 5, pp. 1216–1225, May 2018, doi: 10.1002/bit.26543.
- [104] W. Wang and C. J. Roberts, “Protein aggregation – Mechanisms, detection, and control,” *International Journal of Pharmaceutics*, vol. 550, no. 1–2. Elsevier B.V., pp. 251–268, Oct. 25, 2018, doi: 10.1016/j.ijpharm.2018.08.043.
- [105] M. Ahmadi *et al.*, “Small amounts of sub-visible aggregates enhance the immunogenic potential of monoclonal antibody therapeutics,” *Pharm. Res.*, vol. 32, no. 4, pp. 1383–1394, Oct. 2015, doi: 10.1007/s11095-014-1541-x.
- [106] C. J. Roberts, “Therapeutic protein aggregation: Mechanisms, design, and control,” *Trends in Biotechnology*, vol. 32, no. 7. Elsevier Ltd, pp. 372–380, Jul. 01, 2014, doi: 10.1016/j.tibtech.2014.05.005.
- [107] R. J. Ellis and A. P. Minton, “Protein aggregation in crowded environments,” *Biological*

- Chemistry*, vol. 387, no. 5. De Gruyter, pp. 485–497, May 01, 2006, doi: 10.1515/BC.2006.064.
- [108] J. Wang *et al.*, “Feather keratin hydrogel for wound repair: Preparation, healing effect and biocompatibility evaluation,” *Colloids Surfaces B Biointerfaces*, vol. 149, pp. 341–350, Jan. 2017, doi: 10.1016/j.colsurfb.2016.10.038.
- [109] C. S. Fox, H. A. Berry, and S. Pedigo, “Development and Characterization of Calmodulin-Based Copolymeric Hydrogels,” *Biomacromolecules*, 2020, doi: 10.1021/acs.biomac.0c00043.
- [110] J. Fang *et al.*, “Forced protein unfolding leads to highly elastic and tough protein hydrogels,” *Nat. Commun.*, vol. 4, no. 1, pp. 1–10, Dec. 2013, doi: 10.1038/ncomms3974.
- [111] B. Özerdem and A. Tözeren, “Physical response of collagen gels to tensile strain,” *J. Biomech. Eng.*, vol. 117, no. 4, pp. 397–401, Nov. 1995, doi: 10.1115/1.2794198.
- [112] T. Miyazaki, C. Yomota, and S. Okada, “Development and release characterization of hyaluronan-doxycycline gels based on metal coordination,” *J. Control. Release*, vol. 76, no. 3, pp. 337–347, Oct. 2001, doi: 10.1016/S0168-3659(01)00453-9.
- [113] G. Gentile, F. Greco, and D. Larobina, “Stress-relaxation behavior of a physical gel: Evidence of co-occurrence of structural relaxation and water diffusion in ionic alginate gels,” *Eur. Polym. J.*, vol. 49, no. 12, pp. 3929–3936, Dec. 2013, doi: 10.1016/j.eurpolymj.2013.08.023.
- [114] J. Y. Xiong, J. Narayanan, X. Y. Liu, T. K. Chong, S. B. Chen, and T. S. Chung, “Topology evolution and gelation mechanism of agarose gel,” *J. Phys. Chem. B*, vol. 109, no. 12,



- pp. 5638–5643, Mar. 2005, doi: 10.1021/jp044473u.
- [115] S. Dumitriu and E. Chornet, “Inclusion and release of proteins from polysaccharide-based polyion complexes,” *Advanced Drug Delivery Reviews*, vol. 31, no. 3. Elsevier Sci B.V., pp. 223–246, May 04, 1998, doi: 10.1016/S0169-409X(97)00120-8.
- [116] J. J. Schmidt, J. Rowley, and H. J. Kong, “Hydrogels used for cell-based drug delivery,” *J. Biomed. Mater. Res. Part A*, vol. 87A, no. 4, pp. 1113–1122, Dec. 2008, doi: 10.1002/jbm.a.32287.
- [117] “An Introduction to Rheology - Howard A. Barnes, John Fletcher Hutton, Kenneth Walters - Google Books.”  
[https://books.google.co.uk/books?hl=en&lr=&id=VKj8BAAAQBAJ&oi=fnd&pg=PP1&dq=intro+to+rheology&ots=E-KHhJgKZv&sig=TLiEqDAUUqQFV\\_Q264p-up-kD1A&redir\\_esc=y#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=VKj8BAAAQBAJ&oi=fnd&pg=PP1&dq=intro+to+rheology&ots=E-KHhJgKZv&sig=TLiEqDAUUqQFV_Q264p-up-kD1A&redir_esc=y#v=onepage&q&f=false) (accessed May 19, 2020).
- [118] S. E. Quiñones-Cisneros, C. K. Zéberg-Mikkelsen, and E. H. Stenby, “The friction theory (f-theory) for viscosity modeling,” *Fluid Phase Equilib.*, vol. 169, no. 2, pp. 249–276, Mar. 2000, doi: 10.1016/S0378-3812(00)00310-1.
- [119] E. P. Honig, G. J. Roeberson, and P. H. Wiersema, “Effect of hydrodynamic interaction on the coagulation rate of hydrophobic colloids,” *J. Colloid Interface Sci.*, vol. 36, no. 1, pp. 97–109, May 1971, doi: 10.1016/0021-9797(71)90245-1.
- [120] M. C. I. Mohd Amin, N. Ahmad, N. Halib, and I. Ahmad, “Synthesis and characterization of thermo- and pH-responsive bacterial cellulose/acrylic acid hydrogels for drug delivery,” *Carbohydr. Polym.*, vol. 88, no. 2, pp. 465–473, Apr. 2012, doi: 10.1016/j.carbpol.2011.12.022.

- [121] P. Gupta, K. Vermani, and S. Garg, "Hydrogels: From controlled release to pH-responsive drug delivery," *Drug Discovery Today*, vol. 7, no. 10. Elsevier Current Trends, pp. 569–579, May 15, 2002, doi: 10.1016/S1359-6446(02)02255-9.
- [122] Y. Qiu and K. Park, "Environment-sensitive hydrogels for drug delivery," *Advanced Drug Delivery Reviews*, vol. 53, no. 3. Elsevier, pp. 321–339, Dec. 31, 2001, doi: 10.1016/S0169-409X(01)00203-4.
- [123] B. Hames, *Gel electrophoresis of proteins: a practical approach*. 1998.
- [124] L. C. Lopérgolo, A. B. Lugão, and L. H. Catalani, "Direct UV photocrosslinking of poly(N-vinyl-2-pyrrolidone) (PVP) to produce hydrogels," *Polymer (Guildf)*, vol. 44, no. 20, pp. 6217–6222, Sep. 2003, doi: 10.1016/S0032-3861(03)00686-4.
- [125] "Protein Crosslinking | Thermo Fisher Scientific - UK."  
<https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking.html#> (accessed May 19, 2020).
- [126] S. Wong and D. Jameson, *Chemistry of protein and nucleic acid cross-linking and conjugation*. 2011.
- [127] J. O. Edwards and R. G. Pearson, "The Factors Determining Nucleophilic Reactivities," *J. Am. Chem. Soc.*, vol. 84, no. 1, pp. 16–24, Jan. 1962, doi: 10.1021/ja00860a005.
- [128] C. E. Hoyle and C. N. Bowman, "Thiol-Ene Click Chemistry," *Angew. Chemie Int. Ed.*, vol. 49, no. 9, pp. 1540–1573, Feb. 2010, doi: 10.1002/anie.200903924.
- [129] C. E. Hoyle and C. N. Bowman, "Thiol-Ene Click Chemistry," *Angew. Chemie Int. Ed.*, vol. 49, no. 9, pp. 1540–1573, Feb. 2010, doi: 10.1002/anie.200903924.
- [130] C. E. Hoyle, A. B. Lowe, and C. N. Bowman, "Thiol-click chemistry: A multifaceted

- toolbox for small molecule and polymer synthesis,” *Chemical Society Reviews*, vol. 39, no. 4. The Royal Society of Chemistry, pp. 1355–1387, Mar. 22, 2010, doi: 10.1039/b901979k.
- [131] A. L. MacKinnon and J. Taunton, “Target Identification by Diazirine Photo-Cross-Linking and Click Chemistry,” *Curr. Protoc. Chem. Biol.*, vol. 1, no. 1, pp. 55–73, Dec. 2009, doi: 10.1002/9780470559277.ch090167.
- [132] A. F. Gomes and F. C. Gozzo, “Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS,” *J. Mass Spectrom.*, vol. 45, no. 8, pp. 892–899, Aug. 2010, doi: 10.1002/jms.1776.
- [133] “Photoreactive Crosslinker Chemistry | Thermo Fisher Scientific - UK.” <https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/photoreactive-crosslinker-chemistry.html#4> (accessed May 20, 2020).
- [134] M. Hashimoto and Y. Hatanaka, “Recent progress in diazirine-based photoaffinity labeling,” *European Journal of Organic Chemistry*, vol. 2008, no. 15, pp. 2513–2523, May 2008, doi: 10.1002/ejoc.200701069.
- [135] D. A. Fancy and T. Kodadek, “Chemistry for the analysis of protein–protein interactions: Rapid and efficient cross-linking triggered by long wavelength light,” *Proc. Natl. Acad. Sci.*, vol. 96, no. 11, pp. 6020–6024, May 1999, doi: 10.1073/PNAS.96.11.6020.
- [136] K. Kalyanasundaram, “Photophysics, photochemistry and solar energy conversion with tris(bipyridyl)ruthenium(II) and its analogues,” *Coord. Chem. Rev.*, vol. 46, no. C, pp. 159–244, Oct. 1982, doi: 10.1016/0010-8545(82)85003-0.

- [137] J. Van Houten and R. J. Watts, "Photochemistry of Tris(2,2'-bipyridyl)ruthenium(II) in Aqueous Solutions," *Inorg. Chem.*, vol. 17, no. 12, pp. 3381–3385, 1978, doi: 10.1021/ic50190a016.
- [138] A. Lewandowska-Andralojc and D. E. Polyansky, "Mechanism of the quenching of the tris(bipyridine)ruthenium(II) emission by persulfate: Implications for photoinduced oxidation reactions," *J. Phys. Chem. A*, vol. 117, no. 40, pp. 10311–10319, 2013, doi: 10.1021/jp407573d.
- [139] D. A. Fancy *et al.*, "Scope, limitations and mechanistic aspects of the photo-induced cross-linking of proteins by water-soluble metal complexes," *Chem. Biol.*, vol. 7, no. 9, pp. 697–708, Sep. 2000, doi: 10.1016/S1074-5521(00)00020-X.
- [140] I. Migneault, C. Dartiguenave, M. J. Bertrand, and K. C. Waldron, "Glutaraldehyde: Behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking," *BioTechniques*, vol. 37, no. 5. Future Science Ltd London, UK, pp. 790–802, Nov. 06, 2004, doi: 10.2144/04375rv01.
- [141] D. Barati, S. Kader, S. R. Pajoum Shariati, S. Moeinzadeh, R. H. Sawyer, and E. Jabbari, "Synthesis and Characterization of Photo-Cross-Linkable Keratin Hydrogels for Stem Cell Encapsulation," *Biomacromolecules*, vol. 18, no. 2, pp. 398–412, Feb. 2017, doi: 10.1021/acs.biomac.6b01493.
- [142] A. Bigi, G. Cojazzi, S. Panzavolta, K. Rubini, and N. Roveri, "Mechanical and thermal properties of gelatin films at different degrees of glutaraldehyde crosslinking," *Biomaterials*, vol. 22, no. 8, pp. 763–768, Apr. 2001, doi: 10.1016/S0142-9612(00)00236-2.
- [143] G. Bulaj, "Formation of disulfide bonds in proteins and peptides," *Biotechnology*

- Advances*, vol. 23, no. 1. Elsevier Inc., pp. 87–92, Jan. 01, 2005, doi:  
10.1016/j.biotechadv.2004.09.002.
- [144] J. Gitschier, B. Moffat, D. Reilly, W. I. Wood, and W. J. Fairbrother, “Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase,” *Nat. Struct. Biol.*, vol. 5, no. 1, pp. 47–54, 1998, doi: 10.1038/nsb0198-47.
- [145] A. Miseta and P. Csutora, “Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms,” *Mol. Biol. Evol.*, vol. 17, no. 8, pp. 1232–1239, Aug. 2000, doi: 10.1093/oxfordjournals.molbev.a026406.
- [146] W. Zhou and C. R. Freed, “Tyrosine-to-Cysteine Modification of Human  $\alpha$ -Synuclein Enhances Protein Aggregation and Cellular Toxicity,” *J. Biol. Chem.*, vol. 279, no. 11, pp. 10128–10135, Mar. 2004, doi: 10.1074/jbc.M307563200.
- [147] T. J. Sanborn, P. B. Messersmith, and A. E. Barron, “In situ crosslinking of a biomimetic peptide-PEG hydrogel via thermally triggered activation of factor XIII,” *Biomaterials*, vol. 23, no. 13, pp. 2703–2710, Jul. 2002, doi: 10.1016/S0142-9612(02)00002-9.
- [148] I. E. Dzyaloshinskii, E. M. Lifshitz, and L. P. Pitaevskii, “The general theory of van der Waals forces,” *Adv. Phys.*, vol. 10, no. 38, pp. 165–209, 1961, doi:  
10.1080/00018736100101281.
- [149] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H. E. Gaub, “How strong is a covalent bond,” *Science (80-. )*, vol. 283, no. 5408, pp. 1727–1730, Mar. 1999, doi: 10.1126/science.283.5408.1727.
- [150] J. Ziólkowski, “New relation between ionic radii, bond length, and bond strength,” *J. Solid State Chem.*, vol. 57, no. 3, pp. 269–290, May 1985, doi: 10.1016/0022-

4596(85)90152-5.

- [151] H. J. Kong, E. Alsberg, D. Kaigler, K. Y. Lee, and D. J. Mooney, "Controlling Degradation of Hydrogels via the Size of Crosslinked Junctions," *Adv. Mater.*, vol. 16, no. 21, pp. 1917–1921, Nov. 2004, doi: 10.1002/adma.200400014.
- [152] V. S. Volkov and G. V. Vinogradov, "Molecular theories of nonlinear viscoelasticity of polymers," *Rheol. Acta*, vol. 23, no. 3, pp. 231–237, May 1984, doi: 10.1007/BF01332187.
- [153] T. A. Witten and L. M. Sander, "Diffusion-limited aggregation," *Phys. Rev. B*, vol. 27, no. 9, pp. 5686–5697, May 1983, doi: 10.1103/PhysRevB.27.5686.
- [154] P. Meakin and F. Family, "Structure and dynamics of reaction-limited aggregation," *Phys. Rev. A*, vol. 36, no. 11, pp. 5498–5501, Dec. 1987, doi: 10.1103/PhysRevA.36.5498.
- [155] I. Karatzas and S. E. Shreve, "Brownian Motion," Springer, New York, NY, 1998, pp. 47–127.
- [156] P. Meakin, "Diffusion-limited aggregation in three dimensions: Results from a new cluster-cluster aggregation model," *J. Colloid Interface Sci.*, vol. 102, no. 2, pp. 491–504, Dec. 1984, doi: 10.1016/0021-9797(84)90252-2.
- [157] G. Pranami, "Understanding nanoparticle aggregation." Accessed: May 14, 2020. [Online]. Available: <https://lib.dr.iastate.edu/etd/10859>.
- [158] V. Picciarelli and R. Stella, "An introduction to percolation," *Artic. Eur. J. Phys.*, vol. 15, no. 11994, pp. 97–101, 1999, doi: 10.1088/0143-0807/15/3/001.
- [159] R. C. Ball, D. A. Weitz, T. A. Witten, and F. Leyvraz, "Universal kinetics in reaction-

- limited aggregation," *Phys. Rev. Lett.*, vol. 58, no. 3, pp. 274–277, Jan. 1987, doi: 10.1103/PhysRevLett.58.274.
- [160] M. Kolb, R. Botet, and R. Jullien, "Scaling of kinetically growing clusters," *Phys. Rev. Lett.*, vol. 51, no. 13, pp. 1123–1126, Sep. 1983, doi: 10.1103/PhysRevLett.51.1123.
- [161] R. Jullien and M. Kolb, "Hierarchical model for chemically limited cluster-cluster aggregation," *J. Phys. A. Math. Gen.*, vol. 17, no. 12, p. L639, 1984, doi: 10.1088/0305-4470/17/12/003.
- [162] H. T. Nguyen, L. D. Gelb, A. L. Graham, and P. H. Koenig, "Computer simulations of colloidal gels: how hindered particle rotation affects structure and rheology."
- [163] M.-L. Bang *et al.*, "The Complete Gene Sequence of Titin, Expression of an Unusual  $\approx$ 700-kDa Titin Isoform, and Its Interaction With Obscurin Identify a Novel Z-Line to I-Band Linking System," *Circ. Res.*, vol. 89, no. 11, pp. 1065–1072, Nov. 2001, doi: 10.1161/hh2301.100981.
- [164] L. Tskhovrebova and J. Trinick, "Role of titin in vertebrate striated muscle," 2002, doi: 10.1098/rstb.2001.1028.
- [165] S. Labeit, B. Kolmerer, and W. A. Linke, "The Giant Protein Titin," *Circ. Res.*, vol. 80, no. 2, pp. 290–294, Feb. 1997, doi: 10.1161/01.RES.80.2.290.
- [166] E. Di Cola *et al.*, "Persistence length of titin from rabbit skeletal muscles measured with scattering and microrheology techniques," *Biophys. J.*, vol. 88, no. 6, pp. 4095–4106, Jun. 2005, doi: 10.1529/biophysj.104.054908.
- [167] L. Tskhovrebova and J. Trinick, "Roles of Titin in the Structure and Elasticity of the Sarcomere," *J. Biomed. Biotechnol.*, vol. 2010, 2010, doi: 10.1155/2010/612482.

- [168] C. Muhle Goll, A. Pastore, and M. Nilges, "The three-dimensional structure of a type I module from titin: A prototype of intracellular fibronectin type III domains," *Structure*, vol. 6, no. 10, pp. 1291–1302, Oct. 1998, doi: 10.1016/s0969-2126(98)00129-4.
- [169] S. Improta *et al.*, "The assembly of immunoglobulin-like modules in titin: Implications for muscle elasticity," *J. Mol. Biol.*, vol. 284, no. 3, pp. 761–777, Dec. 1998, doi: 10.1006/jmbi.1998.2028.
- [170] K. Trombitás *et al.*, "Titin extensibility in situ: Entropic elasticity of permanently folded and permanently unfolded molecular segments," *J. Cell Biol.*, vol. 140, no. 4, pp. 853–859, Feb. 1998, doi: 10.1083/jcb.140.4.853.
- [171] A. Soteriou, M. Gamage, and J. Trinick, "A survey of interactions made by the giant protein titin," *J. Cell Sci.*, vol. 104, no. 1, pp. 119–123, Nov. 1993.
- [172] S. Improta, A. S. Politou, and A. Pastore, "Immunoglobulin-like modules from titin I-band: Extensible components of muscle elasticity," *Structure*, vol. 4, no. 3, pp. 323–337, Mar. 1996, doi: 10.1016/S0969-2126(96)00036-6.
- [173] A. S. Politou, D. J. Thomas, and A. Pastore, "The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity," *Biophys. J.*, vol. 69, no. 6, pp. 2601–2610, Dec. 1995, doi: 10.1016/S0006-3495(95)80131-1.
- [174] W. A. Linke and A. Grützner, "Pulling single molecules of titin by AFM - Recent advances and physiological implications," *Pflügers Archiv European Journal of Physiology*, vol. 456, no. 1. Springer, pp. 101–115, Apr. 06, 2008, doi: 10.1007/s00424-007-0389-x.



- [175] J. Hsin, J. Strümpfer, E. H. Lee, and K. Schulten, "Molecular origin of the hierarchical elasticity of titin: Simulation, experiment, and theory," *Annu. Rev. Biophys.*, vol. 40, no. 1, pp. 187–203, Jun. 2011, doi: 10.1146/annurev-biophys-072110-125325.
- [176] S. Keten, Z. Xu, B. Ihle, and M. J. Buehler, "Nanoconfinement controls stiffness, strength and mechanical toughness of B-sheet crystals in silk," *Nat. Mater.*, vol. 9, no. 4, pp. 359–367, Mar. 2010, doi: 10.1038/nmat2704.
- [177] N. Crampton and D. J. Brockwell, "Unravelling the design principles for single protein mechanical strength," *Current Opinion in Structural Biology*, vol. 20, no. 4. Elsevier Current Trends, pp. 508–517, Aug. 01, 2010, doi: 10.1016/j.sbi.2010.05.005.
- [178] N. Duff, N. H. Duong, and D. J. Lacks, "Stretching the immunoglobulin 27 domain of the titin protein: The dynamic energy landscape," *Biophys. J.*, vol. 91, no. 9, pp. 3446–3455, Nov. 2006, doi: 10.1529/biophysj.105.074278.
- [179] D. J. Brockwell *et al.*, "The effect of core destabilization on the mechanical resistance of 127," *Biophys. J.*, vol. 83, no. 1, pp. 458–472, Jul. 2002, doi: 10.1016/S0006-3495(02)75182-5.
- [180] H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek, and J. M. Fernandez, "Point mutations alter the mechanical stability of immunoglobulin modules," *Nat. Struct. Biol.*, vol. 7, no. 12, pp. 1117–1120, 2000, doi: 10.1038/81964.
- [181] W. Stacklies, M. C. Vega, M. Wilmanns, and F. Gräter, "Mechanical network in Titin immunoglobulin from force distribution analysis," *PLoS Comput. Biol.*, vol. 5, no. 3, 2009, doi: 10.1371/journal.pcbi.1000306.
- [182] T. Hoffmann, K. M. Tych, T. Crosskey, B. Schiffrin, D. J. Brockwell, and L. Dougan,

- “Rapid and Robust Polyprotein Production Facilitates Single-Molecule Mechanical Characterization of  $\beta$ -Barrel Assembly Machinery Polypeptide Transport Associated Domains,” *ACS Nano*, vol. 9, no. 9, pp. 8811–8821, Sep. 2015, doi: 10.1021/acsnano.5b01962.
- [183] T. Hoffmann, K. M. Tych, T. Crosskey, B. Schiffrin, D. J. Brockwell, and L. Dougan, “Rapid and Robust Polyprotein Production Facilitates Single-Molecule Mechanical Characterization of  $\beta$ -Barrel Assembly Machinery Polypeptide Transport Associated Domains,” *ACS Nano*, vol. 9, no. 9, pp. 8811–8821, 2015, doi: 10.1021/acsnano.5b01962.
- [184] J. Oroz *et al.*, “The Y9P Variant of the Titin I27 Module: Structural Determinants of Its Revisited Nanomechanics,” *Structure*, vol. 24, no. 4, pp. 606–616, Apr. 2016, doi: 10.1016/j.str.2016.02.016.
- [185] S. B. Fowler *et al.*, “Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering,” *J. Mol. Biol.*, vol. 322, no. 4, pp. 841–849, Sep. 2002, doi: 10.1016/S0022-2836(02)00805-7.
- [186] “An Introduction to Rheology - Howard A. Barnes, John Fletcher Hutton, Kenneth Walters - Google Books.”  
[https://books.google.co.uk/books?hl=en&lr=&id=VKj8BAAAQBAJ&oi=fnd&pg=PP1&dq=intro+to+rheology&ots=E-Ly9FnI\\_B&sig=RcA8cCEPeT92oqbusO3ETwIjC5c&redir\\_esc=y#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=VKj8BAAAQBAJ&oi=fnd&pg=PP1&dq=intro+to+rheology&ots=E-Ly9FnI_B&sig=RcA8cCEPeT92oqbusO3ETwIjC5c&redir_esc=y#v=onepage&q&f=false)  
 (accessed May 21, 2020).
- [187] A. F.-B. of T. instruments and undefined 2004, “Understanding rheology of structured

fluids.”

[188] R. Landel and L. Nielsen, *Mechanical properties of polymers and composites*. 1993.

[189] A. K. Doolittle, “Studies in newtonian flow. II. the dependence of the viscosity of liquids on free-space,” *J. Appl. Phys.*, vol. 22, no. 12, pp. 1471–1475, Dec. 1951, doi: 10.1063/1.1699894.

[190] “Malvern Instruments White Paper - a basic introduction to rheology - Google Search.”

[https://www.google.com/search?ei=91HGXRhNC8PGgQa5y4e4Aw&q=Malvern+Instruments+White+Paper+-+a+basic+introduction+to+rheology&oq=Malvern+Instruments+White+Paper+-+a+basic+introduction+to+rheology&gs\\_lcp=CgZwc3ktYWIQAzoFCCEQoAE6BwghEAoQoAE6BAghEApQhHpY-aMBYIWmAWgAcAB4AIABuAGIAcMVkgEEMzAuMpgBAKABAaoBB2d3cy13aXo&scient=psy-ab&ved=0ahUKEwjxvsiB2sTpAhVDY8AKHbnlATcQ4dUDCAw&uact=5](https://www.google.com/search?ei=91HGXRhNC8PGgQa5y4e4Aw&q=Malvern+Instruments+White+Paper+-+a+basic+introduction+to+rheology&oq=Malvern+Instruments+White+Paper+-+a+basic+introduction+to+rheology&gs_lcp=CgZwc3ktYWIQAzoFCCEQoAE6BwghEAoQoAE6BAghEApQhHpY-aMBYIWmAWgAcAB4AIABuAGIAcMVkgEEMzAuMpgBAKABAaoBB2d3cy13aXo&scient=psy-ab&ved=0ahUKEwjxvsiB2sTpAhVDY8AKHbnlATcQ4dUDCAw&uact=5) (accessed May 21, 2020).

[191] R. Larson, “The structure and rheology of complex fluids,” 1999.

[192] G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler, “Multi-Particle Collision Dynamics: A Particle-Based Mesoscale Simulation Approach to the Hydrodynamics of Complex Fluids,” in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, Springer Berlin Heidelberg, 2009, pp. 1–87.

[193] R. H. Ewell and H. Eyring, “Theory of the Viscosity of Liquids as a Function of Temperature and Pressure,” *J. Chem. Phys.*, vol. 5, no. 9, pp. 726–736, Sep. 1937, doi: 10.1063/1.1750108.

- [194] M. S. Swanson and M. S. Swanson, “Newtonian fluid dynamics,” in *Classical Field Theory and the Stress-Energy Tensor*, IOP Publishing, 2015.
- [195] R. P. Chhabra, “Non-Newtonian fluids: An introduction,” in *Rheology of Complex Fluids*, Springer New York, 2010, pp. 3–34.
- [196] H. Qian and J. J. Hopfield, “Entropy-enthalpy compensation: Perturbation and relaxation in thermodynamic systems,” *J. Chem. Phys.*, vol. 105, no. 20, pp. 9292–9298, Nov. 1996, doi: 10.1063/1.472728.
- [197] P. Bergmann, *Introduction to the Theory of Relativity*. 1976.
- [198] J. N. Kapur and H. K. Kesavan, “Entropy Optimization Principles and Their Applications,” Springer, Dordrecht, 1992, pp. 3–20.
- [199] I. N. Berezovsky, W. W. Chen, P. J. Choi, and E. I. Shakhnovich, “Entropic stabilization of proteins and its proteomic consequences,” *PLoS Comput. Biol.*, vol. 1, no. 4, pp. 0322–0332, 2005, doi: 10.1371/journal.pcbi.0010047.
- [200] F. D. Murnaghan, “The Compressibility of Media under Extreme Pressures,” *Proc. Natl. Acad. Sci.*, vol. 30, no. 9, pp. 244–247, Sep. 1944, doi: 10.1073/pnas.30.9.244.
- [201] F. Birch, “The effect of pressure upon the elastic parameters of isotropic solids, according to Murnaghan’s theory of finite strain,” *J. Appl. Phys.*, vol. 9, no. 4, pp. 279–288, Apr. 1938, doi: 10.1063/1.1710417.
- [202] D. Roylance, “ENGINEERING VISCOELASTICITY,” 2001.
- [203] N. Makris and M. C. Constantinou, “Models of Viscoelasticity with Complex-Order Derivatives,” *J. Eng. Mech.*, vol. 119, no. 7, pp. 1453–1464, Jul. 1993, doi: 10.1061/(ASCE)0733-9399(1993)119:7(1453).

- [204] K. R. Rajagopal, "A note on a reappraisal and generalization of the Kelvin-Voigt model," *Mech. Res. Commun.*, vol. 36, no. 2, pp. 232–235, Mar. 2009, doi: 10.1016/j.mechrescom.2008.09.005.
- [205] M. Khan, S. Hyder Ali, and H. Qi, "On accelerated flows of a viscoelastic fluid with the fractional Burgers' model," *Nonlinear Anal. Real World Appl.*, vol. 10, no. 4, pp. 2286–2296, Aug. 2009, doi: 10.1016/j.nonrwa.2008.04.015.
- [206] M. Madaghiele, A. Sannino, L. Ambrosio, and C. Demitri, "Polymeric hydrogels for burn wound care: Advanced skin wound dressings and regenerative templates," *Burn. Trauma*, vol. 2, no. 4, p. 153, Oct. 2014, doi: 10.4103/2321-3868.143616.
- [207] G. C. Antunes, C. S. Dias, M. M. Telo Da Gama, and N. A. M. Araújo, "Optimal number of linkers per monomer in linker-mediated aggregation," *Soft Matter*, vol. 15, no. 18, pp. 3712–3718, 2019, doi: 10.1039/c9sm00483a.
- [208] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nature Reviews Molecular Cell Biology*, vol. 16, no. 1. Nature Publishing Group, pp. 18–29, Dec. 22, 2015, doi: 10.1038/nrm3920.
- [209] H. Neurath, J. P. Greenstein, F. W. Putnam, and J. O. Erickson, "The chemistry of protein denaturation," *Chem. Rev.*, vol. 34, no. 2, pp. 157–265, Apr. 1944, doi: 10.1021/cr60108a003.
- [210] M. M. Babu, R. van der Lee, N. S. de Groot, and J. Gsponer, "Intrinsically disordered proteins: Regulation and disease," *Current Opinion in Structural Biology*, vol. 21, no. 3. Elsevier Current Trends, pp. 432–440, Jun. 01, 2011, doi: 10.1016/j.sbi.2011.03.011.
- [211] O. B. Ptitsyn, V. E. Bychkova, and V. N. Uversky, "Kinetic and equilibrium folding

- intermediates,” *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, vol. 348, no. 1323, pp. 35–41, Apr. 1995, doi: 10.1098/rstb.1995.0043.
- [212] Y. Zhou and D. Yang, “Equilibrium folding dynamics of meACP in water, heavy water, and low concentration of urea,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, Dec. 2017, doi: 10.1038/s41598-017-16449-4.
- [213] S. J. White, S. D. Johnson, M. A. Sellick, A. Bronowska, P. G. Stockley, and C. Wälti, “The Influence of Two-Dimensional Organization on Peptide Conformation,” *Angew. Chemie Int. Ed.*, vol. 54, no. 3, pp. 974–978, Jan. 2015, doi: 10.1002/anie.201408971.
- [214] L. E. Silbert, “Jamming of frictional spheres and random loose packing,” *Soft Matter*, vol. 6, no. 13, pp. 2918–2924, Jul. 2010, doi: 10.1039/c001973a.
- [215] P. Vladimir *et al.*, “Optimization of Golden Gate assembly through application of ligation sequence-dependent fidelity and bias profiling,” *bioRxiv*, p. 322297, 2018, doi: 10.1101/322297.
- [216] J. R. Warren and J. A. Gordon, “On the refractive indices of aqueous solutions of urea,” *Journal of Physical Chemistry*, vol. 70, no. 1. American Chemical Society, pp. 297–300, 1966, doi: 10.1021/j100873a507.
- [217] L. Whitmore and B. A. Wallace, “Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases,” *Biopolymers*, vol. 89, no. 5, pp. 392–400, May 2008, doi: 10.1002/bip.20853.
- [218] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, “Enzymatic assembly of DNA molecules up to several hundred kilobases,” *Nat. Methods*, vol. 6, no. 5, pp. 343–345, Apr. 2009, doi: 10.1038/nmeth.1318.

- [219] R. H. Ewoldt, M. T. Johnston, and L. M. Caretta, "Experimental Challenges of Shear Rheology: How to Avoid Bad Data," pp. 207–241, 2015, doi: 10.1007/978-1-4939-2065-5\_6.
- [220] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nat. Methods*, vol. 6, no. 5, pp. 343–345, Apr. 2009, doi: 10.1038/nmeth.1318.
- [221] N. Sreerama, S. Y. Venyaminov, and R. W. Woody, "Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Inclusion of Denatured Proteins with Native Proteins in the Analysis," *Anal. Biochem.*, vol. 287, pp. 243–251, 2000, doi: 10.1006/abio.2000.4879.
- [222] N. Sreerama and R. W. Woody, "Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set," 2000, doi: 10.1006/abio.2000.4880.
- [223] L. Whitmore and B. A. Wallace, "DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data," doi: 10.1093/nar/gkh371.
- [224] L. A. Compton and W. C. Johnson, "Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication," *Anal. Biochem.*, vol. 155, no. 1, pp. 155–167, May 1986, doi: 10.1016/0003-2697(86)90241-1.
- [225] "The Design of Experiments: Statistical Principles for Practical Applications - R. Mead - Google Books."  
<https://books.google.co.uk/books?hl=en&lr=&id=CaFZPbCllrMC&oi=fnd&pg=IA4&dq=principles+of+design+of+experiments&ots=tiC2PYtAaK&sig=mpmBtEJfc1a5xBEyXBr>

H-eOgv5k&redir\_esc=y#v=onepage&q=principles of design of experiments&f=false  
(accessed Mar. 07, 2021).

- [226] R. A. Fisher, "The design of experiments.," *Des. Exp.*, no. 7th Ed, 1960.
- [227] F. W. Studier and B. A. Moffatt, "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes," *J. Mol. Biol.*, vol. 189, no. 1, pp. 113–130, May 1986, doi: 10.1016/0022-2836(86)90385-2.
- [228] F. W. Studier, "Protein production by auto-induction in high density shaking cultures.," *Protein Expr. Purif.*, vol. 41, no. 1, pp. 207–234, May 2005, doi: 10.1016/j.pep.2005.01.016.
- [229] A. E. Hirsh and H. B. Fraser, "Protein dispensability and rate of evolution," *Nature*, vol. 411, no. 6841. Nature Publishing Group, pp. 1046–1049, Jun. 28, 2001, doi: 10.1038/35082561.
- [230] C. B. Anfinsen, "Principles that Govern the Folding of Protein Chains," 1973.
- [231] C. W. Chen, M. H. Lin, C. C. Liao, H. P. Chang, and Y. W. Chu, "iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 622–630, Jan. 2020, doi: 10.1016/j.csbj.2020.02.021.
- [232] P. L. Privalov, "Intermediate states in protein folding," *Journal of Molecular Biology*, vol. 258, no. 5. Academic Press, pp. 707–725, May 24, 1996, doi: 10.1006/jmbi.1996.0280.
- [233] Z. Cheng, A. Al Zaki, J. Z. Hui, V. R. Muzykantov, and A. Tsourkas, "Multifunctional nanoparticles: Cost versus benefit of adding targeting and imaging capabilities,"



- Science*, vol. 338, no. 6109. American Association for the Advancement of Science, pp. 903–910, Nov. 16, 2012, doi: 10.1126/science.1226338.
- [234] I. Prigogine and S. Rice, *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics*. 2009.
- [235] Y. Yu, J. Wang, Q. Shao, J. Shi, and W. Zhu, “The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: A microscopic view,” *Sci. Rep.*, vol. 6, no. 1, pp. 1–12, Jan. 2016, doi: 10.1038/srep19500.
- [236] C. F. Wright, K. Lindorff-Larsen, L. G. Randles, and J. Clarke, “Parallel protein-unfolding pathways revealed and mapped,” *Nat. Struct. Biol.*, vol. 10, no. 8, pp. 658–662, Aug. 2003, doi: 10.1038/nsb947.
- [237] L. Sun, J. K. Noel, J. I. Sulkowska, H. Levine, and J. N. Onuchic, “Connecting Thermal and Mechanical Protein (Un)folding Landscapes,” *Biophys. J.*, vol. 107, no. 12, pp. 2950–2961, Dec. 2014, doi: 10.1016/j.bpj.2014.10.021.
- [238] M. Carrion-Vazquez *et al.*, “Mechanical and chemical unfolding of a single protein: A comparison,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 7, pp. 3694–3699, Mar. 1999, doi: 10.1073/pnas.96.7.3694.
- [239] G. I. Makhatadze, “Thermodynamics of protein interactions with urea and guanidinium hydrochloride,” *J. Phys. Chem. B*, vol. 103, no. 23, pp. 4784–4785, Jun. 1999, doi: 10.1021/jp990413q.
- [240] J. Gómez, V. J. Hilser, D. Xie, and E. Freire, “The heat capacity of proteins,” *Proteins Struct. Funct. Bioinforma.*, vol. 22, no. 4, pp. 404–412, Aug. 1995, doi: 10.1002/PROT.340220410.

- [241] K. P. Murphy and E. Freire, "Thermodynamics of structural stability and cooperative folding behavior in proteins," *Advances in Protein Chemistry*, vol. 43. Academic Press Inc., pp. 313–361, Jan. 01, 1992, doi: 10.1016/s0065-3233(08)60556-2.
- [242] "BLAST - Ian Korf, Mark Yandell, Joseph Bedell - Google Books."  
[https://books.google.co.uk/books?hl=en&lr=&id=zDqksXUaoBAC&oi=fnd&pg=PT7&dq=mutation+blast+matrix&ots=L2xBmfGi\\_P&sig=Kw3x9Onj5PZUeFKtAwXk4pcFp7l&redir\\_esc=y#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=zDqksXUaoBAC&oi=fnd&pg=PT7&dq=mutation+blast+matrix&ots=L2xBmfGi_P&sig=Kw3x9Onj5PZUeFKtAwXk4pcFp7l&redir_esc=y#v=onepage&q&f=false) (accessed Feb. 10, 2021).
- [243] D. T. Jones, W. R. Taylor, and J. M. Thornton, "The rapid generation of mutation data matrices from protein sequences," 1992. Accessed: Feb. 10, 2021. [Online]. Available: <https://academic.oup.com/bioinformatics/article/8/3/275/193076>.
- [244] P. C. Ng and S. Henikoff, "Predicting Deleterious Amino Acid Substitutions," doi: 10.1101/gr.176601.
- [245] L. Tjernberg, W. Hosiá, N. Bark, J. Thyberg, and J. Johansson, "Charge attraction and  $\beta$  propensity are necessary for amyloid fibril formation from tetrapeptides," *J. Biol. Chem.*, vol. 277, no. 45, pp. 43243–43246, Nov. 2002, doi: 10.1074/jbc.M205570200.
- [246] J. M. Khan, A. Malik, A. Ahmed, O. H. A. Alghamdi, and M. Ahmed, "SDS induces cross beta-sheet amyloid as well as alpha-helical structure in conconavalin A," *J. Mol. Liq.*, vol. 319, p. 114154, Dec. 2020, doi: 10.1016/j.molliq.2020.114154.
- [247] T. Casalini and G. Perale, "From Microscale to Macroscale: Nine Orders of Magnitude for a Comprehensive Modeling of Hydrogels for Controlled Drug Delivery," *Gels*, vol. 5, no. 2, p. 28, May 2019, doi: 10.3390/gels5020028.
- [248] S. Lin and L. Gu, "Influence of Crosslink Density and Stiffness on Mechanical

- Properties of Type I Collagen Gel,” *Materials (Basel)*., vol. 8, no. 2, pp. 551–560, Feb. 2015, doi: 10.3390/ma8020551.
- [249] M. D. G. Hughes, S. Cussons, N. Mahmoudi, D. J. Brockwell, and L. Dougan, “Single molecule protein stabilisation translates to macromolecular mechanics of a protein network,” *Soft Matter*, vol. 16, no. 27, pp. 6389–6399, Jul. 2020, doi: 10.1039/c9sm02484k.
- [250] L. D. Matt D G Hughes, Benjamin S Hanson, Sophie Cussons, Najet Mahmoudi, David J Brockwell, “Control of intra-protein nano-staples defines network architecture and mechanics of protein hydrogels due to in situ protein unfolding,” *ACS Nano (under Rev.*
- [251] D. A. Fancy and T. Kodadek, “Chemistry for the analysis of protein-protein interactions: Rapid and efficient cross-linking triggered by long wavelength light,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 11, pp. 6020–6024, 1999, doi: 10.1073/pnas.96.11.6020.
- [252] S. SAKURA and D. FUJIMOTO, “Absorption and Fluorescence Study of Tyrosine-Derived Crosslinking Amino Acids From Collagen,” *Photochem. Photobiol.*, vol. 40, no. 6, pp. 731–734, 1984, doi: 10.1111/j.1751-1097.1984.tb04644.x.
- [253] J. Li *et al.*, “Interfacial Stress in the Development of Biologics: Fundamental Understanding, Current Practice, and Future Perspective,” *AAPS J.*, vol. 21, no. 3, pp. 1–17, May 2019, doi: 10.1208/s12248-019-0312-3.
- [254] L. Cacopardo, N. Guazzelli, R. Nossa, G. Mattei, and A. Ahluwalia, “Engineering hydrogel viscoelasticity,” *J. Mech. Behav. Biomed. Mater.*, vol. 89, pp. 162–167, Jan. 2019, doi: 10.1016/j.jmbbm.2018.09.031.

- [255] N. Phan-Thien and N. Mai-Duy, *Understanding Viscoelasticity*. Cham: Springer International Publishing, 2017.
- [256] C. K. Lee, Y. M. Wang, L. S. Huang, and S. Lin, "Atomic force microscopy: Determination of unbinding force, off rate and energy barrier for protein-ligand interaction," *Micron*, vol. 38, no. 5. Pergamon, pp. 446–461, Jul. 01, 2007, doi: 10.1016/j.micron.2006.06.014.
- [257] J. R. Forman and J. Clarke, "Mechanical unfolding of proteins: insights into biology, structure and folding," *Current Opinion in Structural Biology*, vol. 17, no. 1. Elsevier Current Trends, pp. 58–66, Feb. 01, 2007, doi: 10.1016/j.sbi.2007.01.006.
- [258] E. M. Puchner and H. E. Gaub, "Force and function: probing proteins with AFM-based force spectroscopy," *Current Opinion in Structural Biology*, vol. 19, no. 5. Elsevier Current Trends, pp. 605–614, Oct. 01, 2009, doi: 10.1016/j.sbi.2009.09.005.
- [259] B. U. Felderhof, "Loss of momentum in a viscous compressible fluid due to no-slip boundary condition at one or two planar walls," *J. Chem. Phys.*, vol. 133, no. 7, p. 074707, Aug. 2010, doi: 10.1063/1.3473932.
- [260] K. C. Labropoulos, S. Rangarajan, D. E. Niesz, and S. C. Danforth, "Dynamic Rheology of Agar Gel Based Aqueous Binders," *J. Am. Ceram. Soc.*, vol. 84, no. 6, pp. 1217–1224, Dec. 2004, doi: 10.1111/j.1151-2916.2001.tb00819.x.
- [261] J. Kim, R. Roque, and B. Birgisson, "Interpreting Dissipated Energy from Complex Modulus Data," *Road Mater. Pavement Des.*, vol. 7, no. 2, pp. 223–245, Jan. 2006, doi: 10.1080/14680629.2006.9690034.
- [262] A. Puisto, M. Mohtaschemi, M. J. Alava, and X. Illa, "Dynamic hysteresis in the

- rheology of complex fluids,” *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 91, no. 4, p. 042314, Apr. 2015, doi: 10.1103/PhysRevE.91.042314.
- [263] A. M. Grillet, N. B. Wyatt, A. M. Grillet, and L. M. Gloc, “Polymer Gel Rheology and Adhesion 3 Polymer Gel Rheology and Adhesion,” 2014, doi: 10.5772/36975.
- [264] J. Ching-Wei Lee, K. M. Weigandt, E. G. Kelley, and S. A. Rogers, “Structure-Property Relationships via Recovery Rheology in Viscoelastic Materials,” 2019, doi: 10.1103/PhysRevLett.122.248003.
- [265] J. Liu and H. Qi, “Dissipated energy function, hysteresis and precondition of a viscoelastic solid model,” *Nonlinear Anal. Real World Appl.*, vol. 11, no. 2, pp. 907–912, Apr. 2010, doi: 10.1016/j.nonrwa.2009.01.033.
- [266] T. E. Brown *et al.*, “Photopolymerized Dynamic Hydrogels with Tunable Viscoelastic Properties through Thioester Exchange,” doi: 10.1016/j.biomaterials.2018.03.060.
- [267] A. Aufderhorst-Roberts *et al.*, “Reaction Rate Governs the Viscoelasticity and Nanostructure of Folded Protein Hydrogels,” *Biomacromolecules*, vol. 21, no. 10, pp. 4253–4260, Oct. 2020, doi: 10.1021/acs.biomac.0c01044.
- [268] B. S. Hanson and L. Dougan, “Network Growth and Structural Characteristics of Globular Protein Hydrogels,” *Macromolecules*, vol. 53, no. 17, pp. 7335–7345, Sep. 2020, doi: 10.1021/acs.macromol.0c00890.
- [269] J. M. Zuidema, C. J. Rivet, R. J. Gilbert, and F. A. Morrison, “A protocol for rheological characterization of hydrogels for tissue engineering strategies,” *J. Biomed. Mater. Res. Part B Appl. Biomater.*, vol. 102, no. 5, pp. 1063–1073, Jul. 2014, doi: 10.1002/jbm.b.33088.

- [270] J. Yang, C. Gong, F. K. Shi, and X. M. Xie, "High strength of physical hydrogels based on poly(acrylic acid)-g-poly(ethylene glycol) methyl ether: Role of chain architecture on hydrogel properties," *J. Phys. Chem. B*, vol. 116, no. 39, pp. 12038–12047, Oct. 2012, doi: 10.1021/jp303710d.
- [271] R. P. Chhabra, "Non-Newtonian fluids: An introduction," in *Rheology of Complex Fluids*, Springer New York, 2010, pp. 3–34.
- [272] D. Vigetti, E. Karousou, M. Viola, S. Deleonibus, G. De Luca, and A. Passi, "Hyaluronan: Biosynthesis and signaling ☆," 2014, doi: 10.1016/j.bbagen.2014.02.001.
- [273] J. McDonald, V. C. Hascall, K. Meyer, and J. W. Palmer, "Hyaluronan minireview series," *Journal of Biological Chemistry*, vol. 277, no. 7. American Society for Biochemistry and Molecular Biology, pp. 4575–4579, Feb. 15, 2002, doi: 10.1074/jbc.R100064200.
- [274] "On Growth and Form: Fractal and Non-Fractal Patterns in Physics - Google Books." [https://books.google.co.uk/books?hl=en&lr=&id=aHL8CAAAQBAJ&oi=fnd&pg=PA3&dq=meakin+on+growth+and+form&ots=TRiJmpsLT-&sig=LclQn2VlwcGP9rAqwvhqS83LQ9Q&redir\\_esc=y#v=onepage&q=meakin on growth and form&f=false](https://books.google.co.uk/books?hl=en&lr=&id=aHL8CAAAQBAJ&oi=fnd&pg=PA3&dq=meakin+on+growth+and+form&ots=TRiJmpsLT-&sig=LclQn2VlwcGP9rAqwvhqS83LQ9Q&redir_esc=y#v=onepage&q=meakin on growth and form&f=false) (accessed May 20, 2020).